NBER WORKING PAPER SERIES

WHAT DO CONSUMERS CONSIDER BEFORE THEY CHOOSE? IDENTIFICATION FROM ASYMMETRIC DEMAND RESPONSES

Jason Abaluck
Abi Adams

Working Paper 23566
http://www.nber.org/papers/w23566

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2017

Thanks to Leila Bengali and Mauricio Caceres for excellent research assistance and to Dan Ackerberg, Joe Altonji, Dan Benjamin, Steve Berry, Judy Chevalier, Jonathan Feinstein, Jeremy Fox, Xavier Gabaix, Jonathan Gruber, Phil Haile, Erzo Luttmer, Paola Manzini, Marco Mariotti, Costas Meghir, Olivia Mitchell, Fiona Scott Morton, Barry Nalebuff, Joe Shapiro, K. Sudhir and participants in the Heterogeneity in Supply and Demand Conference, the Roybal Annual Meeting, and the Yale IO, labor economics and econometrics workshops for helpful discussions. Also special thanks to Arthur Lewbel for retrieving Jason's keys when he left them in the seminar computer, and thanks to Raluca Ursu for help in replicating her Expedia analysis. We acknowledge financial support from NIA grant number R01 AG031270 and the Economic and Social Research Council, Grant ES/N017099/1. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses

Jason Abaluck and Abi Adams

NBER Working Paper No. 23566
June 2017
JEL No. D0,D8

ABSTRACT

Consideration set models relax the assumption that consumers are aware of all available options. Thus far, identification arguments for these models have relied either on auxiliary data on what options were considered or on instruments excluded from consideration or utility. In a discrete choice framework subsuming logit, probit and random coefficients models, we prove that utility and consideration set probabilities can be separately identified without these data intensive methods. In full-consideration models, choice probabilities satisfy a symmetry property analogous to Slutsky symmetry in continuous choice models. This symmetry breaks down in consideration set models when changes in characteristics perturb consideration, and we show that consideration probabilities are constructively identified from the resulting asymmetries. In a lab experiment, we recover preferences and consideration probabilities using only data on which items were ultimately chosen, and we apply the model to study hotel choices on Expedia.com and insurance choices in Medicare Part D.

Jason Abaluck
Yale School of Management
Box 208200
New Haven, CT 06520-8200
and NBER
jason.abaluck@yale.edu

Abi Adams
University of Oxford
Department of Economics
Manor Rd, Oxford OX1 3UQ
United Kingdom
abi.adams@economics.ox.ac.uk

1 Introduction

Discrete choice models generally assume that consumers are aware of all available options. This prevents researchers from asking many questions of interest. What factors lead consumers to consider more options? Will inertial consumers 'wake up' in response to a price increase but remain unresponsive if rivals lower prices? Which products will respond well to advertising because they have high market shares conditional on being noticed? Normatively, whether people eat the same foods and go to the same stores year after year because they like those options or because they do not know what else exists has first-order consequences for welfare. If one can measure preferences conditional on consideration, we can assess the benefits of policies that help consumers make more considered choices.

Consideration set models are a generalization of discrete choice models that relax the assumption that individuals consider all goods. These models instead specify a probability that each subset of options is considered. The framework has a long tradition in psychology and marketing (Hauser and Wernerfelt 1990; Shocker, Ben-Akiva, Boccara, and Nedungadi 1991) and has become increasingly popular in both theoretical and applied literatures in economics. Consideration sets might arise due to inattention or bounded rationality (Treisman and Gelade 1980), from search costs (Caplin, Dean, and Leahy 2016), or because consumers face (unobserved) constraints on what options can be chosen (Gaynor, Propper, and Seiler 2016).[1]

Identification is an immediate concern in consideration set models – if changes in prices or other characteristics perturb demand, can we tell whether this impact comes via consideration or utility? Empirical models in this literature have previously relied either on auxiliary data or exclusion restrictions to answer this question. For example, Conlon and Mortimer (2013) assume that consideration sets are known in some periods, Honka (2014) and Honka, Hortaçsu, and Vitorino (2015) assume consideration sets are partially observed from survey data,[2] while Caplin and Dean (2015) assume choice probabilities conditional on consideration can be measured. Numerous papers assume that observables impact either attention or utility but not both.[3,4] Manzini and Mariotti (2014) derive revealed preference conditions under which utility and consideration probabilities can be identified separately from choice data, but their approach requires a degree of choice set variation

[1] In this paper, we use "attentive" as synonymous with "a good is in the consumer's consideration set".

[2] Honka, Hortaçsu, and Vitorino (2015) further distinguish between "awareness" and consideration. The framework used in this paper can be thought of as a reduced form version of their model in which "awareness" and "consideration" jointly determine the set of goods from which consumers choose.

[3] Among others, Goeree (2008), Gaynor, Propper, and Seiler (2016), Heiss, McFadden, Winter, Wupperman, and Zhou (2016) and Hortaçsu, Madanizadeh, and Puller (2015) take this approach.

[4] Crawford, Griffith, and Iaria (2016) show that identification is possible without excluding any variables from utility or consideration with panel data, but only if one restricts how consideration sets change over time.

that is not observed outside of experimental settings, such as observing choices from every pairwise set of options.[5]

In this paper, we prove that the restrictions on choice probabilities imposed by economic theory are sufficient to separately identify preferences and consideration probabilities in many applied settings of interest. Our method does not require panel data or auxiliary information on consideration sets, and it allows all observables to impact both consideration and utility. We provide simple closed form expressions for consideration set probabilities in terms of differences in cross-derivatives (the discrete choice analogue of 'Slutsky asymmetries'). Our framework subsumes many of the consideration set models in the applied literature and does not rely on assuming a particular functional form for random utility errors. We also show that ad hoc attempts to model consideration sets such as fixed effects in utility for products on different shelves or interactions between prices and such fixed effects still lead to misspecified models; the consideration set models we consider are equivalent to full-consideration models where the utility of good j depends directly on characteristics of rival goods $j' \neq j$, a dependence which is typically ruled out.

Our identification result builds on the insight that imperfect consideration breaks the symmetry between cross-price responses (or more generally, cross-characteristic responses). For example, in a model with a default, symmetry would ordinarily require that switching decisions be equally responsive to an increase in the price of the default good by \$100 or a decrease in the price of all rival goods by \$100. Suppose instead that consumers will be inattentive and choose the default option unless the default good becomes sufficiently unsuitable. Now, switching decisions will be unresponsive to changes in the price of rival goods but more responsive to changes in the price of the default to the degree that these changes perturb attention. While the link between imperfect attention and Slutsky asymmetry has been discussed in the theoretical literature, notably in Gabaix (2014), this link has not previously been utilized in applied work.[6]

Our results imply that, in many applications, one could estimate consideration set models rather than the conventional discrete choice models that they nest.[7] In cross-sectional data, our results can be used to identify whether goods are demanded because they are high-utility or because they are more likely to be considered. In panel data, one can evaluate whether inertia reflects switching costs

[5] Masatlioglu, Nakajima, and Ozbay (2012) derive revealed preference tests for imperfect consideration in a deterministic model. However, their conditions are not sufficient for point identification of preferences nor attention.

[6] Chen, Levy, Ray, and Bergen (2008) also note a connection between inattention and asymmetries in theoretical models. Davis and Schiraldi (2014) provide generalizations of multinomial logit models that permit asymmetries, but they explicitly note that these models cannot be rationalized by an underlying random utility interpretation and do not attempt to use these asymmetries to identify inattention.

[7] A Stata command which implements several special cases of our model is available for download as "alogit"; a User's Guide as well as sample datasets can be downloaded at https://sites.google.com/view/alogit/home.

or inattention.[8] More generally, one can perform behavioral welfare analyses with no additional data beyond what is needed to estimate conventional logit, probit or random coefficients models. Our identification proofs are constructive and so consistent nonparametric estimators can be readily based on them. However, in most applications of interest, we advocate estimating parametric generalizations of conventional models. Our model is generally over-identified, and if instruments are available, one can use our results to test the validity of additional exclusion restrictions.

We illustrate the value of our identification results in several applications, showing empirically that consideration set models imply different substitution patterns and normative conclusions from the full consideration models that they nest. We consider two special cases of our model, although our general identification proof subsumes hybrid models combining features of both alternatives. We focus on these cases because they are well-motivated by theoretical microfoundations for inattention. However, these special cases do not *require* a particular microfoundation and thus permit more general forms of inattention as well.

First, we develop the "Alternative-Specific Consideration" (ASC) model, which assumes that the probability that a good is considered depends on characteristics of that good only. This is a natural framework in an online search setting, for example, where each item's ranking in search depends to first-order on its own attributes (Goeree 2008). We validate the ASC model in a lab experiment in which participants made a series of choices from proper subsets of 10 possible goods. Using only data on choices and ignoring information on what items were available, we use the ASC model to accurately recover the probabilities that each good was available as well as recovering the preference parameters that we would estimate conditional on knowing which items were available. Conventional models with a comparable number of parameters misspecify own- and cross-price elasticities relative to the "true" elasticities computed using data on which items were actually available. The average absolute error in cross-elasticities in conventional models is 2.5 times larger than in the ASC model or 45.5 percentage points larger as a fraction of the average absolute cross-elasticity.

We also apply the ASC model to hotel choice data from Expedia in which the order of hotels in online search results was randomized. We show that the ASC model implies the randomized ordering impacts attention but not utility. Further, when the model is estimated on hotels shown in the 3rd-10th search positions, we can predict out-of-sample which hotels will experience the largest increase in demand when they are put in the 1st and 2nd search positions. In so doing, we can decompose whether current demand is due to high utility, and thus whether a hotel would be more

[8] In this case, we use the term "inertia" to refer to the empirical tendency to make the same choices today as in the previous period. "Switching costs" refers to inertia which arises due to adjustment costs and persistent unobserved heterogeneity (both of which would make you worse-off if forced to switch) as opposed to inattention.

popular if more people noticed it, or high attention, suggesting that additional advertising is unlikely to be effective.

Our second special case is the "Default-Specific Consideration" (DSC) model, in which consideration probabilities depend only on the characteristics of a default good. This framework is often applied in settings with a clear default that is chosen unless consumers believe it to be so unsatisfactory that they actively seek out alternatives (Hortaçsu, Madanizadeh, and Puller 2015; Ho, Hogan, and Scott Morton 2015). We apply the DSC model to health plan choice data from Medicare Part D. We replicate the finding in Ho, Hogan, and Scott Morton (2015) that switching decisions are far more sensitive to characteristics of the default plan than characteristics of rival plans, and we show that this implies that the observed degree of inertia is largely due to inattention given the cross-derivative asymmetries this generates. While we find that most inertia is due to inattention, adjustment costs are sufficiently large that they offset the cost savings from assigning beneficiaries to the lowest cost plans. In conventional models, we estimate switching costs of \$1,000-\$1,400; after accounting for inattention, the DSC model implies adjustment costs of \$0-\$300. We also conduct overidentification tests to demonstrate that the specific patterns of asymmetries we observe in the data are consistent with our underlying model of inattention. The degree of inattention we estimate is consistent with that in Heiss, McFadden, Winter, Wupperman, and Zhou (2016). We also directly test their identifying assumption that changes in plan characteristics over time impact attention but not utility conditional on the level of those characteristics, finding that it is vindicated.

The rest of this paper proceeds as follows. In Section 2, we work through a simple example to illustrate our identification argument. Section 3 lays out our general model and identification proof. Section 4 outlines the ASC model, validates it with a lab experiment, and applies it to Expedia data. Section 5 develops the DSC model, describes overidentification tests that can be used for model validation, and applies these to data from Medicare Part D. Section 6 concludes.

2 Motivating Example

To illustrate our identification argument, we first outline a stylized example to highlight the main features of our approach. In this simple model, consumers pick a default option unless the default becomes so unsuitable that they are shocked into paying attention to other products. Note that many of the assumptions we make here are for expository purposes and will be relaxed in Section 3, where we consider a more general consideration set model.[9]

[9] In the existing literature, this model resembles those in Ho, Hogan, and Scott Morton (2015) and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) which we consider more generally below.

Consider a consumer selecting from two possible products, $j = \{0, 1\}$, for example insurance plans. Each plan has a price, x_j. One product, plan 0, is a default good that is always considered. The consumer may or may not pay attention to the other product depending on how expensive the default good is. If the consumer does not pay attention, they pick the default. However, if the consumer pays attention to the non-default good, then they pick the good that maximizes their quasilinear utility function from the set of plans considered.[10]

Let $\mu(x_0)$ give the probability that a consumer pays attention to both products as a function of the price of plan 0. The probability that a consumer picks plan j, s_j, in this model can then be expressed as:

$$\begin{aligned} s_0(x_0, x_1) &= (1 - \mu) + \mu s_0^\star(x_0, x_1) \\ s_1(x_0, x_1) &= \mu s_1^\star(x_0, x_1) \end{aligned} \tag{2.1}$$

where $s_j^\star$ gives choice probabilities conditional on paying attention.

We will show that μ, $s_0^\star$ and $s_1^\star$ can be separately identified in this model using data on how the observed shares s_j vary with product attributes. The key to our identification argument is that maximizing behavior implies symmetry given full consideration. In the two good case, this symmetry is very intuitive – with full attention, no income effects, and no outside option, consumers should only care about price *differences*. Symmetry of demand responses is violated if changes in product characteristics also impact consideration probabilities. Differentiating Equation 2.1 and using the fact that the market shares conditional on paying attention satisfy symmetry, we obtain:

$$\frac{\partial s_1}{\partial x_0} - \frac{\partial s_0}{\partial x_1} = \frac{\partial \mu}{\partial x_0} s_1^\star = \frac{\partial \log(\mu)}{\partial x_0} s_1 \tag{2.2}$$

where the second equality follows from the fact that $s_1 = \mu s_1^*$. Thus, changes in the probability of considering both goods are directly identified from data on choice probabilities:

$$\boxed{\frac{\partial \log(\mu)}{\partial x_0} = \frac{1}{s_1}\left[\frac{\partial s_1}{\partial x_0} - \frac{\partial s_0}{\partial x_1}\right]} \tag{2.3}$$

If the price of the default plan perturbs attention by causing consumers to "wake up" (the left-hand side), then the non-default plan will be more sensitive to the price of the default plan than is the default plan to the price of the non-default plan. This is a behavioral pattern noted in the health insurance literature (Ho, Hogan, and Scott Morton 2015).

Recovering the derivative of the attention probability identifies the level of attention up to a constant. This constant is determined by the fact that cross-derivatives are symmetric at $\mu(x_0) = 1$,

[10] In Section 3.3 we also relax quasilinearity.

a point expanded upon in Section 3. Integrating Equation 2.3 over the support of x_0 we obtain:

$$\boxed{\mu(\bar{x}_0) = \exp\left(-\int_{\bar{x}_0}^{\infty} \frac{1}{s_1}\left[\frac{\partial s_1}{\partial x_0} - \frac{\partial s_0}{\partial x_1}\right] dx_0\right)} \tag{2.4}$$

With a large number of consumers and exogenous variation in x_0 and x_1, we could in principle estimate consideration probabilities directly by estimating the functions $\partial s_1/\partial x_0$ and $\partial s_0/\partial x_1$. However, if the number of plans becomes large further parametric assumptions will be required, as in any discrete choice setting. In the next section, we show that the techniques used in this example generalize to a broader class of models before turning to estimation.

3 Model & Identification

In this section, we outline formally our analytic framework and nonparametric identification results. We begin by defining a consideration set model and stating the assumptions that suffice for cross-derivative symmetry given full consideration. We then give additional assumptions on consideration probabilities that suffice for point identification of consideration probabilities from asymmetries in demand.

We consider an individual i who makes a discrete choice among $J+1$ products, $\mathcal{J} = \{0, 1, ..., J\}$, with $J \geq 1$. Each product j is characterized as a bundle of $K \geq 1$ characteristics, x_{ij}, with support $\chi \subseteq \mathbb{R}^K$. Let $x_i = [x_{i0}, ..., x_{iJ}]$. We allow for individuals to consider an (unobserved) subset of available goods when making their choice. The set of goods that a consumer considers is called the consideration set. At this point, we place no restrictions on consideration set formation except that there exists a default option, good-0, that is always considered. The default may be the 'outside' good (without observed characteristics) or an 'inside' good (with observed characteristics). Let $\mathcal{P}(\mathcal{J})$ represent the power set of goods, with any given element of $\mathcal{P}(\mathcal{J})$ indexed by c. The set of consideration sets containing good j is then given as:

$$\mathbb{P}(j) = \{c : c \in \mathcal{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad 0 \in c\} \tag{3.1}$$

We will develop identification results for a set of choice models that imply choice probabilities of the following form:

$$s_{ij}(x_i) = \sum_{c \in \mathbb{P}(j)} \pi_{ic}(x_i) s^{\star}_{ij}(x_i|c) \tag{3.2}$$

where s_{ij} is the observed probability of individual i selecting j (the market share of good j), π_{ic} gives the probability that the set of goods c is considered, and $s^{\star}_{ij}(x_i|c)$ gives the probability that i

selects good j from the set c. For the most part, we suppress the dependence of these quantities on x. As π_{ic} and $s^{\star}_{ij}(x_i|c)$ represent proper probabilities, we have:

$$\sum_{c \in \mathcal{P}(\mathcal{J})} \pi_{ic} = 1 \quad , \quad \sum_{j \in c} s^{\star}_{ij}(c) = 1 \tag{3.3}$$

The structural objects of interest are the consideration set probabilities, π_{ic}, and the unobserved latent choice probabilities, $s^{\star}_{ij}(x_i|c)$. We do not directly address the identification of preference parameters given knowledge of $s^{\star}_{ij}(x_i|c)$ nor the identification of, for example, search costs given consideration probabilities. The parameters of any utility model that are identified from choice behavior with full consideration, and the parameters of models that provide microfoundations for consideration sets given consideration probabilities will follow from our identification results. Our aim is to provide general identification results that can be tailored by applied researchers to special cases of the framework considered here.

3.1 Key Assumptions

In this subsection, we state assumptions that suffice for cross-derivative symmetry with full consideration and, thus, assumptions under which asymmetries imply imperfect consideration. Individuals make choices from any given consideration set to maximize their utility. We take a random utility approach, decomposing individual i's utility from good j, u_{ij}, into a deterministic component that depends on the characteristics of good j and a random error term:

$$u_{ij} = v_{ij}(x_{ij}) + \epsilon_{ij} \tag{3.4}$$

Assumption 1. *Additive Separability* There exists a characteristic x^1_{ij} that is additively separable in the indirect utility function:

$$\begin{aligned} u_{ij} &= v_{ij}(x_{ij}) + \epsilon_{ij} && (3.5)\\ &= g_i(x^1_{ij}) + w_{ij}(x^2_{ij}) + \epsilon_{ij} && (3.6)\\ &= \beta_i x^1_{ij} + w_{ij}(x^2_{ij}) + \epsilon_{ij} && (3.7) \end{aligned}$$

where $x^2_{ij} \in \mathbb{R}^{K-1}$, $\beta_i \sim F(\beta_i)$ where β_i is independent of x_j for all $j = 0, ..., J$.

The restriction that consumers value the separable characteristic equally across choices can be substantive, although it is natural in many settings. While point identification of consideration

probabilities requires only additive separability (Equation 3.6), in the proof in the main text we will make the stronger assumption of quasilinearity (Equation 3.7) for expositional simplicity. We discuss the restrictiveness of the separability assumption further in Section 3.3.

Assumption 2. *Exogenous Characteristics:* $\epsilon_{ij} \perp\!\!\!\perp \mathbf{x}_{ij'}$ for $\forall i$. We focus on the question of identification without the additional complications arising from endogeneity in this paper. This assumption will be relaxed in future work.

Assumption 3. *One Continuous Characteristic:* x^1_{ij} is continuously distributed and the distribution of $x^1_{ij}|x^2_{ij}$ has a positive density everywhere on χ.

Assumption 4. $F(\epsilon_{i0}, ..., \epsilon_{iJ})$ is absolutely continuous with respect to the Lebesgue measure and gives rise to a density function that is everywhere positive on $\mathbb{R}$.

With [] denoting exclusion, the probability that individual i chooses option j having considered the set of options c, with $j \in c$, is given by:

$$s^\star_{ij}(c) = Pr\left(v_{ij} + \epsilon_{ij} = \max_{j' \in c} v_{ij'} + \epsilon_{ij'}\right) \tag{3.8}$$

Given consideration of all alternatives, choice probabilities that result from maximizing a utility function satisfying our assumptions will satisfy cross-derivative symmetry and lack nominal illusion.

Corollary 1. *Symmetry of Cross Derivatives*: with respect to the quasi-linear characteristic:

$$\frac{\partial s^\star_{ij}}{\partial x^1_{ij'}} = \frac{\partial s^\star_{ij'}}{\partial x^1_{ij}} \tag{3.9}$$

Corollary 2. *Absence of Nominal Illusion*: level shifts in the separable characteristic do not alter choice probabilities:

$$s^\star_{ij}(x^1_i, x^2_i) = s^\star_{ij}(x^1_i + \delta, x^2_i) \tag{3.10}$$

where δ denotes a level shift which impacts all goods equally. Proof in Appendix A.

ASSUMPTION 5. π_{ic} is continuously differentiable for all $c \in \mathcal{P}(\mathcal{J})$ with, for $\pi_{ic} < 1$ and $j = 1, ..., J$:

$$\frac{\partial \pi_{ic}}{\partial x_{ij}^1} \neq 0 \tag{3.11}$$

This assumption is natural in most applied settings of interest unless there is some mechanical reason why consideration is truly random. In many settings, the question of what drives changes in attention is itself a question of interest.

Slutsky Asymmetries & Nominal Illusion In our baseline model, only one mechanism is available to generate cross-derivative asymmetries: imperfect consideration.[11] Later in this section we describe overidentification tests and routes to determine whether the pattern of asymmetries is consistent with our model of imperfect consideration as opposed to resulting from violations of our underlying assumptions or other behavioral anomalies.

LEMMA 1. ASYMMETRIES & NOMINAL ILLUSION IMPLY IMPERFECT CONSIDERATION. Given Assumptions 1-5, if

$$\frac{\partial s_{ij}}{\partial x_{ij'}^1} \neq \frac{\partial s_{ij'}}{\partial x_{ij}^1} \tag{3.12}$$

$$s_{ij}(x_i^1, x_i^2) \neq s_{ij}(x_i^1 + \delta, x_i^2) \tag{3.13}$$

for $\delta \neq 0$, then $\pi_i(\mathcal{J}) < 1$, where $\pi_i(\mathcal{J})$ is the probability that an individual considers all goods $\mathcal{J} = \{0, ..., J\}$. Proof in Appendix A.

3.2 Identification

To make progress towards point identification of the structural functions of interest, we must place some additional restrictions on consideration set probabilities. If the π_{ic} are allowed to vary arbitrarily, then identification of the underlying structural functions is hopeless (Manzini and Mariotti 2014). As a simple illustration, consider a case with two choice sets, $\{0, 1\}$ and $\{0, 1, 2\}$ where:

$$s_{i1} = \pi_i(\{0, 1\}) s_{i1}^\star(\{0, 1\}) + \pi_i(\{0, 1, 2\}) s_{i1}^\star(\{0, 1, 2\}) \tag{3.14}$$

[11] Lemma 1 states that cross-derivative asymmetries and nominal illusion are sufficient but not necessary for imperfect attention. Thus, some models of imperfect attention in certain scenarios do not predict these choice patterns. For example, Matejka and McKay (2014) show that when actions are homogeneous a priori and exchangable in the decision maker's prior, a rational inattention model provides a foundation for the mulitnomial logit (which yields symmetric cross-derivatives).

If we substitute for $\pi_i(\{0,1\})$ and $\pi_i(\{0,1,2\})$ with $\hat{\pi}_i(\{0,1\}) = \pi_i(\{0,1,2\})s^\star_{i0}(\{0,1,2\})/s^\star_{i0}(\{0,1\})$ and $\hat{\pi}_i(\{0,1,2\}) = \pi_i(\{0,1\})s^\star_{i0}(\{0,1\})/s^\star_{i0}(\{0,1,2\})$, then one obtains a model with the same observed market shares and conditional choice probabilities, but different consideration set probabilities. For point identification, restrictions must be imposed on how consideration probabilities can vary with the underlying characteristics of available goods.

Consideration set models in the applied literature typically take one of two forms. The 'Default-Specific Consideration' (DSC) model, assumes the existence of an inside default good and allows the probability of considering all alternative options to vary only as a function of the characteristics of that default. Under this approach, the market shares of the default (good 0) and non-default goods take the form:

$$\begin{aligned} s_{i0} &= (1 - \mu_{i0}) + \mu_{i0} s^\star_{i0}(\mathcal{J}) \\ s_{ij} &= \mu_{i0} s^\star_{ij}(\mathcal{J}) \quad \text{for } j > 0 \end{aligned} \tag{3.15}$$

where $\mu_{i0}(x_{i0})$ gives the probability of considering all available products. The models developed by Ho, Hogan, and Scott Morton (2015) and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) provide microfoundations for this framework: consumers either consider just the default or the full set of available products; only if the characteristics of the default get sufficiently bad do consumers pay a cost to search among all available products.[12]

An alternative strand of the literature assumes that each good has an independent probability of being considered that depends on characteristics of that good. This includes the models in Goeree (2008), Gaynor, Propper, and Seiler (2016) and Manzini and Mariotti (2014). Under this 'Alternative-Specific Consideration' (ASC) approach, consideration set probabilities take the form:

$$\pi_{ic} = \prod_{j \in c} \phi_{ij} \prod_{j' \notin c} \left(1 - \phi_{ij'}\right) \tag{3.16}$$

where the probability of good j being considered, $\phi_{ij}(x_{ij})$, is a function of own characteristics only.

Many rational-inattention microfoundations lead to models where a good is considered if a good-specific function exceeds a threshold value (Caplin, Dean, and Leahy 2016). The ASC model would then be appropriate in cases when this threshold is insensitive to the characteristics of any single good. In Appendix A.5 we give a wider discussion of when the characteristics of rival goods will only have a second order impact on consideration thresholds, so that rank models are well-approximated

[12] This model is also related to a more general class of Downwards Recursive sequential search models (Weitzman 1979). We develop this connection further in Appendix A.5, proving identification of the ex ante ranking of options and the probability of product availability in a simple model of this class.

by the ASC model.

In the remainder of this section we show that consideration probabilities are identified from cross-derivative asymmetries in a hybrid model subsuming the ASC and DSC models. However, if interest lies in scenarios that cannot be nested within our hybrid framework, in Appendix A we show in a more general environment that features of consideration probabilities are identified up to a monotonic transformation using the same methods.[13] Let the market share of the inside default, good-0, and non-default goods take the form:

$$\begin{array}{rcl} s_{i0} & = (1-\mu_{i0}) \quad + \quad & \mu_{i0} \sum_{c \in \mathbb{P}(0)} \prod_{l \in c} \phi_{il} \prod_{l' \notin c} (1-\phi_{il'})\, s^{\star}_{i0}(c) \\ s_{ij} & = & \mu_{i0} \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il} \prod_{l' \notin c} (1-\phi_{il'})\, s^{\star}_{ij}(c) \quad \text{for } j>0 \end{array} \tag{3.17}$$

where $\phi_{i0}(x_{i0}) = 1$ for all $x_{i0} \in \chi$.

Restricting $\phi_{ij} = 1$ for all $j > 0$ gives the DSC model. Restricting $\mu_{i0} = 1$ gives the ASC model. While discussion of our identification results will proceed assuming an inside default good, our results hold with minimal changes if interest lies in the ASC model with an outside default good (i.e. a default with unobserved characteristics, e.g. buy none of the options) or in the ASC model where the probability of considering good j also depends directly on the characteristics of the default. These variants are discussed in Appendix A.

Identifying Changes in Consideration Probabilities The central insight of our proof is that changes in consideration probabilities can be expressed as a function of observable differences in cross-derivatives and market shares. First imagine that consumer choice is observed in a market where it is known that good $j' > 0$ is not available. For example, in the beer market, a local craft beer might not be available in all locations or, for health insurance, a plan introduced at time t is not available at $t-1$. With a slight abuse of notation, let the set of consideration sets containing good $j > 0$ and not containing $j' > 0$ be given as:

$$\mathbb{P}(j/j') = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad j' \notin c \quad \& \quad 0 \in c\} \tag{3.18}$$

[13] More precisely, when consideration probabilities can be written as a function of good-specific indices, so $\pi_c = \pi(v_{i1}, ..., v_{iJ})$ with $\frac{\partial v}{\partial x^1_{ij}}$ constant across goods, we can recover the v_{ij} up to a monotonic transformation. This identifies the relative impact of different characteristics on good-specific indices.

The probability that good j is chosen in a market in which j' is not available is:

$$s_{ij}(\mathcal{J}/j') = \mu_{i0} \sum_{c \in \mathbb{P}(j/j')} \prod_{l \in c} \phi_{il} \prod_{l' \notin \{c, j'\}} (1 - \phi_{il'})\, s^{\star}_{ij}(c) \tag{3.19}$$

The change in the probability of choosing good j when j' is removed from the choice set can be decomposed into two terms: the probability that the consumer was paying attention to j' times the impact of j' on purchasing good j within each consideration set including j:

$$s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j') = \overbrace{\mu_{i0}\phi_{ij'}}^{\text{Probability consider } j'} \underbrace{\sum_{c \in \mathbb{P}(j/j')} \prod_{l \in c} \phi_{il} \prod_{l' \notin \{c, j'\}} (1 - \phi_{il'}) \left(s^{\star}_{ij}(c \cup j') - s^{\star}_{ij}(c)\right)}_{\text{Impact of adding } j' \text{ to all consideration sets including } j} \tag{3.20}$$

The impact of removing j' from the choice set is informative for the magnitude of cross derivative differences. As the impact on choice probabilities within consideration sets cancel out (due to symmetry conditional on consideration), the size of the cross derivative differences depends on the degree to which consideration probabilities are altered times the difference that adding that good to the choice set has on choice probabilities. Expressing cross derivative differences as a function of 'leave-one-out' market share differences for $j, j' \neq 0$ gives:

$$\frac{\partial s_{ij}}{\partial x^1_{ij'}} - \frac{\partial s_{ij'}}{\partial x^1_{ij}} = \frac{\partial \log(\phi_{ij'})}{\partial x^1_{ij'}} \left(s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')\right) - \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} \left(s_{ij'}(\mathcal{J}) - s_{ij'}(\mathcal{J}/j)\right) \tag{3.21}$$

Cross derivatives with respect to characteristics of the default good take a slightly different form as the default is present in all choice sets. Cross derivative differences with $j' = 0$ are given by the linear system:

$$\frac{\partial s_{ij}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{ij}} = \frac{\partial \log(\mu_{i0})}{\partial x^1_{i0}} s_{ij}(\mathcal{J}) - \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} \left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)\right) \tag{3.22}$$

Equations 3.21 and 3.22 give closed form expressions for cross-derivative differences as a linear function of $\partial \log(\phi_{ij})/\partial x^1_{ij}$. Let this system be expressed as:

$$c_i = D_i \theta_i \tag{3.23}$$

where c_i is the vector of cross derivative differences, θ_i is the $J+1$-vector of log consideration probability derivatives, and D_i is the coefficient matrix of leave-one-out differences.[14].

As there are typically more than $J+1$ cross-derivative differences, it is convenient to work with

[14] See Appendix A for illustrations of the structure of these matrices.

the system:[15]

$$D_i' c_i = D_i' D_i \theta_i \tag{3.24}$$

If $D_i' D_i$ is full rank, there is a unique solution to this system and changes in consideration probabilities are uniquely identified from choice data.

Assumption 6. (Rank Condition) The matrix $D_i' D_i$ is full rank.

Appendix A discusses the restrictions on structural functions required for Assumption 6 to hold. A strength of our approach is that the rank condition is testable given market share data. If the rank condition holds, then the derivatives of log consideration probabilities are given as:

$$\theta_i = \left(D_i' D_i\right)^{-1} D_i' c_i \tag{3.25}$$

While the above proof relies on "leave-one-out" variation, the amount of choice set variation required is substantially less than earlier work (Manzini and Mariotti 2014). Further, this variation can be replaced by a large support assumption on x_{ij}^1 for nonparametric identification of changes in choice probabilities.

Assumption 7A. As $x_{ij}^1 \to -\infty, \quad s_{ij} \to 0$.

Assumption 7A imposes that at low values of x_{ij}^1, either good j is not paid attention to (as $x_{ij}^1 \to -\infty, \phi_{ij} \to 0$) or it is not chosen because it generates low utility (as $x_{ij}^1 \to -\infty, s_{ij}^\star(c) \to 0$ for all $c \in \mathbb{P}(j)$) even if it is available. Then, $s_{ij}(\mathcal{J}) \to s_{ij}(\mathcal{J}/j') \quad$ as $x_{ij}^1 \to -\infty$.

In practice, parametric assumptions can also replace this type of variation in the data as we will discuss in Section 4.

Identifying the Level of Consideration Probabilities Given identification of the derivatives of log consideration probabilities by the argument above, ϕ_{ij} is identified up to a scale factor C by integrating over the support of x_{ij}^1:

$$\log(\phi_{ij}) = \int \frac{\partial \log(\phi_{ij})}{\partial x_{ij}^1} \, \mathrm{d}x_{ij}^1 + C \tag{3.26}$$

[15] Alternative weighting matrix, W_i, can be used: $D_i' W_i D_i$.

Identifying the level of attention requires an additional assumption to pin down the constant of integration, C. Assuming that consumers are prompted to pay attention to good j when x_{ij}^1 reaches an extreme value enables the level of attention to be identified. This assumption is analogous to those made in the literature on nonparametric identification of multinomial discrete choice models (Berry and Haile (2009), Lewbel (2000)), treatment effects (Heckman and Vytlacil 2005; Lewbel 2007)), the identification of binary games and entry models (Tamer 2003), and the use of special regressors more generally. Further, this assumption is testable in our setting by checking that cross derivative differences are symmetric at that value of the covariate.[16]

ASSUMPTION 7B. As $x_{ij}^1 \to \infty$, $\phi_{ij} \to 1$.

Identifying Full Consideration Market Shares Nominal illusion facilitates the identification of the 2^J independent latent choice probabilities, $s_{ij}^\star$.[17] We will treat ϕ_{ij} as known in this subsection given the argument above. Imagine that $N = 2^J$ level shifts in the separable characteristic are observed. These shifts alter consideration probabilities but do not alter latent choice probabilities conditional on consideration. Let $k = 1, ..., C$ index the consideration sets of which j is a member. The probabilities of these consideration sets containing j are given as $\pi_{j1}, ..., \pi_{jC}$. For each good $j > 0$,[18] define the matrices:

$$\Pi_j = \begin{bmatrix} \pi_{j1}(\delta_1) & \cdots & \pi_{jC}(\delta_1) \\ \vdots & \ddots & \vdots \\ \pi_{j1}(\delta_N) & \cdots & \pi_{jC}(\delta_N) \end{bmatrix} \tag{3.27}$$

$$s_{ij}^\star = \left[s_{ij}^\star(c_{j1}), ..., s_{ij}^\star(c_{jC}) \right] \tag{3.28}$$

$$s_{ij}^\delta = \left[s_{ij}(x_i^1 + \delta_1, x_i^2), ..., s_{ij}(x_i^1 + \delta_N, x_i^2) \right] \tag{3.29}$$

where

$$\pi_c(\delta) = \mu_{i0}(x_{i0}^1 + \delta) \prod_{l \in c} \phi_{il}(x_{il}^1 + \delta) \prod_{l' \notin c} \left(1 - \phi_{il'}(x_{il'}^1 + \delta)\right) \tag{3.30}$$

[16] 'Thin set identification' will not be a problem in our intended applications so long as x_{ij}^1 has a strictly positive probability of attaining the value at which attention is paid with probability one, this problem is overcome.

[17] This identification problem is analogous to the problem of identifying the 'long' regression. While the functions of interest are typically only partially identified without instruments (Henry, Kitamura, and Salanié 2014), we show that optimizing behavior here results in point identification of the objects of interest.

[18] The latent market shares of the default good are given by adding up within each consideration set.

with the dependence of ϕ_{ij} on x_{ij}^2 suppressed for notational simplicity and $\phi_{i0} = 1$ for all x_{i0}. Unobserved latent choice probabilities are defined as the solution to the following linear system:

$$\Pi_j s_{ij}^\star = s_{ij}(\delta) \tag{3.31}$$

$$s_{ij}^\star = \Pi_j^{-1} s_{ij}^\delta \tag{3.32}$$

There is a unique solution to this system, and thus all $s_{ij}^\star$ are identified, when all Π_j are full rank.

Assumption 8. (Rank Condition) Π_j is full rank for $j = 1, ..., J$.

Appendix A discusses the restrictions on structural functions required for Assumption 8 to hold. Again, given identification of ϕ_{ij}, these assumptions are testable and thus their validity can be assessed for the particular application on hand.

Theorem 1. (Identification of Hybrid Consideration Set Model) Given Assumptions 1-8, consideration probabilities, $\mu_{i0}(x_{i0})$ and $\phi_{ij}(x_{ij})$, and latent market shares conditional on consideration set c, $s_{ij}^\star(c)$ are identified for all $c \in \mathbb{P}(\mathcal{J})$ and $j \in \mathcal{J}$.

Overidentification With $J > 2$, the derivative of the log of consideration probabilities (and thus consideration set probabilities) are over-identified. With $N > 2^J$, latent market shares are over-identified. This provides the potential to test the validity of the consideration set model outlined in this paper. From Equation 3.23, changes in consideration set probabilities, θ_i, are defined by the linear system:

$$D_i \theta_i - c_i = 0 \tag{3.33}$$

where c_i is the vector of cross derivative differences, θ_i is the $J+1$-vector of log consideration probability derivatives, and D_i is the coefficient matrix of leave-one-out differences. There are

$$\underbrace{\frac{1}{2}J(J+1)}_{\#\text{ Independent Cross Deriv. Diffs}} - \overbrace{(J+1)}^{\#\ \phi_{ij}\text{ Derivatives}} \tag{3.34}$$

overidentifying restrictions. Similar reasoning shows that there are $N - 2^J$ overidentifying restrictions for latent market shares. In Section 5, we conduct overidentification tests based on this reasoning

in a parametric model.

Estimation In principle, nonparametric estimation of consideration probabilities and latent market shares is possible given market share data and application of the analogy principle. Our identification proofs are constructive and so consistent nonparametric estimators can be readily based on them. However, in practice, as with most discrete choice models, parametric assumptions will be necessary to deal with the dimensionality problem that arises with many goods and many characteristics. Specifically, suppose that consideration probabilities and choice probabilities conditional on consideration can be written as a function of parameters θ. Then, choice probabilities are given by:

$$\begin{aligned} s_{i0} &= (1-\mu_{i0}(\theta)) \quad + \quad \mu_{i0}(\theta) \textstyle\sum_{c\in\mathbb{P}(0)} \prod_{l\in c} \phi_{il}(\theta) \prod_{l'\notin c} (1-\phi_{il'}(\theta))\, s^{\star}_{i0}(c,(\theta)) \\ s_{ij} &= \qquad\qquad\qquad\quad \mu_{i0}(\theta) \textstyle\sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_{il}(\theta) \prod_{l'\notin c} (1-\phi_{il'}(\theta))\, s^{\star}_{ij}(c,(\theta)) \qquad \text{for } j>0 \end{aligned} \tag{3.35}$$

And we can construct the resulting likelihood function given our assumptions about how choices correlate across individuals and time. In the simplest case where these choices are uncorrelated, we have:

$$\log L = \sum_{ij} d_{ij} \log(s_{ij}) \tag{3.36}$$

In sections 4 and 5, we provide additional details in various parametric models.

3.3 Other Sources of Asymmetry

Given our assumptions, imperfect consideration is the only mechanism giving rise to an asymmetric cross-derivative matrix. Relaxing our background assumptions might, however, give rise to alternative sources of asymmetry that our framework could incorrectly attribute to inattention. The main substantive assumption required for symmetry in Section 3.1 is that there is at least one characteristic for which utility is a common separable function of that characteristic across goods. We first discuss neoclassical reasons why this assumption might fail. Next, we consider alternative behavioral explanations for asymmetries and show that our identification result still holds in models that accommodate these alternative behavioral anomalies.

Income Effects and Nonlinear Price Responses Our identification proof can allow for income effects as usually estimated in empirical models provided we observe sufficient characteristic variation

conditional on income. For example, if utility is given by:

$$u_{ij} = g_i(y_i, p_{ij}) + w_{ij}(x_{ij}) + \epsilon_{ij} \tag{3.37}$$

where $g_i(\cdot)$ is a smooth function, then cross-price effects are equal conditional on derivatives evaluated at the same values of price and income:

$$\frac{\partial s^\star_{ij}}{\partial p_{ij'}}\Big|_{y_i=y, p_{ij}=p_{ij'}=p} = \frac{\partial s^\star_{ij'}}{\partial p_{ij}}\Big|_{y_i=y, p_{ij}=p_{ij'}=p} \tag{3.38}$$

Identification of consideration probabilities then follows analogously to Section 3.2, where we can recover the derivative of consideration probabilities at a given point provided we observe choice probabilities for different goods at the same income and prices. If income effects are large and non-separable however, they might confound identification of asymmetries.

The restriction that $g_i(\cdot)$ (or β_i in our proof) is the same across goods is another substantive restriction, although one that is theoretically well-motivated in many cases. This assumption may fail in cases where consumers value the same amenity differently across goods, or when characteristics that appear observationally equivalent to the econometrician may in fact differ (e.g. in-room dining may generate different utility at different hotels). Fortunately, our model only requires a single characteristic that is separable with a common coefficient – and this follows if there is at least one characteristic which is separable in the direct utility function. For example, if income is separable in direct utility, the coefficient on price will be common across goods and give the marginal utility of wealth.

Behavioral Explanations for Asymmetries One might also ask whether there are alternative 'behavioral' stories other than imperfect consideration that might lead to cross-derivative asymmetries. We do not attempt the impossible task of enumerating every possible psychological anomaly that can occur – instead, we focus on well-documented phenomenon that might lead to asymmetries.[19]

One robustly documented pattern is that consumers respond more to larger proportional changes in prices. This is sometimes referred to as the "Weber-Fechner law of psychophysics". This effect could be captured by allowing indirect utility to be a nonlinear function of price and can be analyzed in an analogous manner to our discussion of income effects; utility will be symmetric conditional

[19] We note a tendency among some economists to argue that if any behavioral phenomenon is permitted, *all* behavioral phenomena must be allowed for. We believe that this critique is sometimes applied in an arbitrary and unfair way. *All* models should consider behavioral factors when they are well-documented and empirically or normatively first-order for the question asked, and this is what we attempt to do here.

on two goods having the same price level. Additionally, this model predicts a different pattern of asymmetries than our consideration set model. In our consideration set model, asymmetries scale with latent utilities. In the Weber-Fechner model, cross-price asymmetries scale with the difference in prices between two goods. In Section 5, we show in one application that cross-derivatives scale in the manner predicted by our consideration set model.

Loss aversion is also associated with asymmetries. However, the asymmetry involved is somewhat different to that which arises from consideration sets. Suppose that consumers respond asymmetrically to price changes relative to a reference point, so that in Equation 3.5, we replace $\beta_i x_{ij}^1$ with $\beta_i^+(x_{ij}^1 - \bar{x}_{ij})$ for $x_{ij}^1 - \bar{x}_{ij} > 0$ and $\beta_i^-(x_{ij}^1 - \bar{x}_{ij})$ for $x_{ij}^1 - \bar{x}_{ij} < 0$. Loss aversion does not aways break the fundamental symmetry in cross-derivatives. As long as β_i^+ and β_i^- are the same across goods, price increases for good A will have the same impact on the demand for B as price increases for B have on the demand for A. Thus, the fundamental insight underlying our identification result still holds, and consideration probabilities can be separately identified in a model that allows for loss aversion.

More generally, asymmetries might also arise from different forms of inattention from those modeled here. We have assumed that attention occurs at the level of goods. An alternative possibility, developed in Gabaix (2014), is that inattention occurs at the level of *characteristics*. While a comprehensive treatment of inattention to characteristics is beyond the scope of this paper, we here show that the patterns of asymmetries implied by Gabaix (2014) are distinguishable from those in our model of good-specific attention. Adapting Gabaix (2014) to a discrete choice setting gives an indirect utility function of the form:

$$u_{ij} = \beta(p_{id} + m_j^p(p_{ij} - p_{id})) + w_{ij}(x_{id} + m_j^x(x_{ij} - x_{id})) + \epsilon_{ij} \tag{3.39}$$

where m^p represents the attention paid to the price of j relative to the price of a default good, d. m^x is a $(K-1)$-vector of analogous attention parameters for the remaining characteristics.

Discussing consumer choice, Gabaix (2014) treats m_j^p as structural parameters that are fixed independently of the realized characteristics of each good. In this model, the ratio of cross-derivatives is constant but not generally equal to 1 (implying asymmetric cross-derivatives):[20]

$$\frac{\frac{\partial s_{ij}^C}{\partial p_{ij'}}}{\frac{\partial s_{ij'}^C}{\partial p_{ij}}} = \frac{\theta_{j'}^p}{\theta_j^p} \quad \neq 1 \quad \text{when } \theta_{j'} \neq \theta_j \tag{3.40}$$

[20]We here focus on cross-derivative ratios rather than differences as they take a particularly simple form in the Gabaix (2014) model.

where the C superscript denotes that we are considering cross-derivatives with inattention to characteristics rather than goods. In our consideration set model, however,the ratio of cross-derivatives is not constant but instead scales with 'leave-one-out' market shares.

A generalization of the Gabaix (2014) model, closer in spirit to our framework, would permit m_j to vary with x_{ij}. However, this model imposes few testable restrictions on the data since consumers potentially attend differently to every characteristic of every good. There are a variety of ways to discipline it to become more empirically relevant. One possibility is to assume that consumers differentially attend to some characteristics, but that inattention does not vary across goods. Perhaps attention to a particular characteristic varies with the variance of that characteristic across goods, for example. However, cross-price effects are fully symmetric in such a model. Unless m_j^p varies across goods, inattention to characteristics will not generate asymmetries. Alternatively, one could consider a model in which m_j^x varied across goods but not characteristics (so that $m_j^x = m_j^p$ for all $x \neq p$). This pattern would be more naturally captured by our model of good-specific consideration.

Thus, while alternative behavioral stories can generate asymmetries, a response to proportional changes in prices, loss aversion, or other forms of inattention are distinguishable from the models of good-specific consideration we consider here.

4 Alternative-Specific Consideration

In the following sections we consider the two special cases that motivate our general model in more detail. The proof in Section 3 shows that one can, if desired, combine the ASC and DSC models.[21]

In the Alternative-Specific Consideration (ASC) model, consideration probabilities for each good are independent and depend on characteristics of that good. The ASC model is appropriate in many settings where one observes cross-sectional data with no clear default good. In online applications, the ranking of a product in search will depend on attributes of that product. In bricks and mortar retail, the shelf or location in the store is likewise a function of observable attributes of that product. The ASC model has been widely applied in the literature but has thus far relied on additional exclusion restrictions for identification (Goeree 2008; Gaynor, Propper, and Seiler 2016).

In this section we give parametric assumptions that facilitate estimation, describe the misspecification that arises in full-consideration models that ignore consideration sets, and discuss two applications designed to validate the empirical usefulness of the model. We show in a lab experiment that we can use the ASC model to recover consideration probabilities and preferences, and we

[21] We also have a Stata command that estimates the special cases laid out in Sections 4 and 5. Type "ssc install alogit". A User's Guide and example datasets are available at: https://sites.google.com/view/alogit/home. Contact us if you run into any problems.

use the randomly assigned ordering of hotels in Expedia choice data to further validate our model.

4.1 Parametric Assumptions

The ASC model defines market shares as:

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il}(x_{il}) \prod_{l' \notin c} (1 - \phi_{il'}(x_{il'})) \, s^{\star}_{ij}(c) \tag{4.1}$$

where $\phi_{i0} = 1$ and $\mathbb{P}(j) = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad 0 \in c\}$ and as in Section 3, ϕ_{ij} denotes the probability that consumer i considers option j and $s^{\star}_{ij}(c)$ denotes choice probabilities conditional on consideration set c. This model is a special case of our general framework with the restriction $\mu_{i0} = 1$. We assume that a default, good-0, is present in every consideration set.[22]

We restrict our empirical parametric analysis to a linear random utility model with errors, ϵ_{ij}, distributed Type 1 Extreme Value:

$$u_{ij} = x_{ij}\beta_i + \epsilon_{ij} \tag{4.2}$$

Following Goeree (2008), let good j be considered if $x_{ij}\gamma > \eta_{ij}$, where η_{ij} is distributed logistic. The probability that good j is considered is:

$$\phi_{ij} = Pr\left(x_{ij}\gamma - \eta_{ij} > 0\right) \tag{4.3}$$

$$= \frac{\exp(x_{ij}\gamma)}{1 + \exp(x_{ij}\gamma)} \tag{4.4}$$

This model can be estimated by maximizing the likelihood given by:

$$P(Y_{ij} = 1|\theta) = P(c = \emptyset|\theta) \cdot \mathbb{1}_{j=d} + \sum_{c \in C} \prod_{l \in c} \phi_{il}(\theta) \prod_{k \notin c} (1 - \phi_{ik}(\theta)) P(Y_{ij} = 1|c, \theta) \tag{4.5}$$

Goeree (2008) shows how computational issues in summing over 2^J choice sets can be avoided using a simulated likelihood approach. We provide details in Appendix C.

[22] We prove in Appendix A that the model is identified if one chooses as the default an outside good with utility normalized to 0 or an 'inside' good with attention probability less than 1. We also show that a generalization of this model in which consideration probabilities depend both on own characteristics and default characteristics ($\phi_{ij}(x_{i0}, x_{ij})$ for $j > 0$) is identified.

4.2 Relationship to Full-Consideration Models

These assumptions allow us to express this consideration set model as a full consideration model where the utility of each good depends directly on the characteristics of rival goods. This allows us to show precisely how conventional models which attempt to deal with consideration sets in ad hoc ways are misspecified. We show in Appendix B that the ASC model can be rewritten as a full-consideration model where the utility of each option j is given by:

$$u_{ij} = x_{ij}\beta_i + \psi_{ij} + \epsilon_{ij} \tag{4.6}$$

where:

$$\begin{aligned} \psi_{ij} &= \ln\left(\frac{s_{ij}}{1-s_{ij}}\right) - \ln\left(\frac{s^\star_{ij}(j)}{1-s^\star_{ij}(j)}\right) \\ &= \ln\left(\frac{\phi_{ij}\sum_{k\neq j}\exp(x_{ik}\beta_i+\psi_{ik})}{(1-\phi_{ij})\exp(x_{ij}\beta_i)+\sum_{k\neq j}\exp(x_{ik}\beta_i+\psi_{ik})}\right) \end{aligned} \tag{4.7}$$

s_{ij} is the probability that option j is chosen and $s^\star_{ij}(j)$ denotes the probability of choosing option j conditional on always considering that option.[23]

In this representation, the utility of good j depends directly on the characteristics of rival goods j'. With full consideration, if the price of good j increasing raises demand for rival goods, then the price of rival goods increasing should raise the demand for good j. If this does not happen because consumers are inattentive to j, we can model this as the prices of rival goods decreasing the utility of good j. With full attention, $\phi_{ij} = 1$ and $\psi_{ij} = 0$ and this additional effect is not present.

This representation shows that ad hoc specifications of random utility models designed to adjust for consideration sets are generally misspecified. Suppose one observes a variety of products on different shelves and wants to model consumers as more attentive to products on certain shelves than on others. One might try adding shelf fixed effects to a conventional logit model, but this would fail to capture the fact that, for example, demand will be less responsive to all characteristics of products on out of the way shelves. Is it sufficient to interact shelf fixed effects with all other observed characteristics? The above representation shows that this flexible model is still misspecified. To permit the relevant asymmetries, one would need to allow characteristics of all goods $j' \neq j$ to directly enter the utility of each good j.

[23] This representation links with the approach of Crawford, Griffith, and Iaria (2016), who show that the impact of time invariant consideration sets can be 'differenced out', by treating them as fixed effect.

4.3 Lab Experiment

Our proof gives conditions under which utility and consideration probabilities can in principle be identified given observed choices. In practice, one might worry that this is placing 'too much structure' on the observed data. Perhaps consideration set models are highly sensitive to a small amount of misspecification or require an unreasonable amount of data to accurately recover the structural functions of interest.

To investigate these issues, we conduct a lab experiment in which consumers make choices from known subsets of a superset of 10 goods. We then ask, given these choices, can we recover the consideration probabilities as well as preferences conditional on consideration using only information on observed choices from the superset of 10 goods? We show that the attentive logit model *does* recover preferences and attention probabilities, and additionally, it yields consistent estimates of own and cross-price elasticities while flexible full-information models with an equal or greater number of price parameters fail to do so. Some goods seem low utility in conventional models because they are rarely considered – the attentive logit model correctly recovers that they have high utility conditional on being considered.

Set-up We selected 10 goods sold at the Yale Bookstore with list prices ranging from \$19.98-\$24.98. These goods and their list prices are shown in Table 7 in Appendix D. Each participant was endowed with \$25 and made 50 choices from randomly chosen subsets of the 10 goods with randomized prices (one third of the list price plus a uniformly distributed amount between \$0 and \$16). After making all 50 choices, one of these choices was randomly selected and they were given that item as well as \$25 minus the price of the item in cash. Prior to the experiment, participants were given several examples to illustrate the incentive scheme and were quizzed on their understanding. 70% correctly answered our test of understanding (and all participants were told why their answer was correct or incorrect). Appendix Table 10 reports results using only this subset of users who passed this test and shows that results are qualitatively unchanged.

The probability that each good appeared in a given choice set was fixed in advance – this probability varied across goods and with prices so that goods were more likely to be considered if they had a higher price. The probabilities were chosen so that most choice sets would range from 2-7 products. See Table 2 for the precise coefficients.[24] To increase the likelihood that participants considered all of the products, 2-7 products they were presented with, we required consumers to spend at least 10 seconds before finalizing their choices. This allows us to take choices from the

[24]Consideration probabilities for each good are given by the function $\phi_j = \Lambda(\xi_j + 0.15p_{ij})$ where ξ_j is the coefficient reported in Table 2 (either -2.5 or 0 for all goods) and $\Lambda(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$.

generated choice set as representative of consumers' true preferences. A sample product selection screen is shown in Figure D.1 in Appendix D. Consumers are shown images of all the products in their choice set along with the (randomly chosen) prices. They click the radio button for the product they want, and can click "Next" after 10 seconds.

Estimation of the ASC model requires specifying a default good which is chosen if the consideration set is empty. We specify this as good 10. At the estimated parameter values, an empty consideration set has a 0.2% chance of occurring so the choice of default does not impact estimation. To recover the preference parameters and consideration probabilities, we estimate the model by maximum likelihood. We also compare our results to a variety of flexible full-consideration specifications with a similar or larger number of parameters. In total, we ran the experiment with 149 participants, resulting in 7,450 choices.[25]

Results Table 1 compares the average estimated consideration probabilities by good to the true consideration probabilities which are known to the experimenter. Like in any real-world setting, the underlying utility model could be a nested logit, a random coefficients model, a multinomial probit or anything else – but our experiment shows that the attentive logit model without any random coefficients nonetheless correctly recovers the underlying attention probabilities. The 2nd panel of Table 2 reports the coefficients which give rise to the consideration probabilities and Figure 1 shows this information graphically. Across products, the confidence intervals on the fixed effects in the attention equation include the true values with the exception of product 1, which lies close to the boundary of the confidence interval. We also correctly recover the impact of price on attention. Price has a (known) coefficient of 0.15 – by construction, consumers are more likely to see a product if the price is higher, as might arise in the real world if sellers advertise their premium products. In the attentive logit model, we estimate 0.137 (.017).

Table 2 compares the estimated parameters from a conditional logit model (estimated as if all 10 goods are considered), our 'attentive logit' model, and the 'true' values. The 'true' preference parameters are estimated given the actual choice sets that consumers faced. In contrast, the attentive logit parameters are estimated using only information about the product consumers actually chose and not information about the specific subset of 10 goods they could choose from in each instance. Consider the preference parameters shown in the top panel of Table 2 (we consider the implied

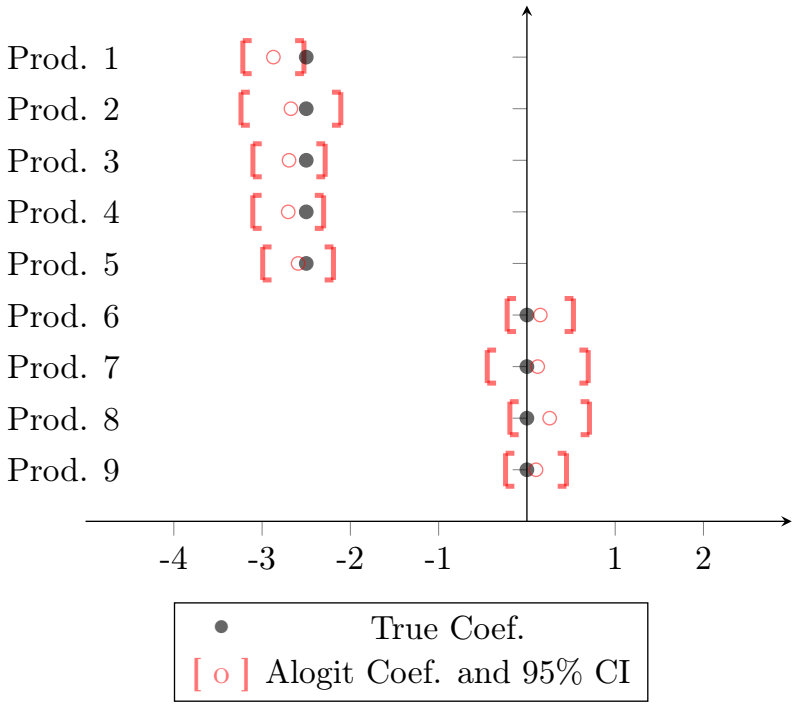
[25]There were 150 in total, but one participant's data was not recorded properly because they refreshed the browser several times during the experiment – this participant is dropped from the final analysis. When a participant refreshed the browser, the choice recorded in our data was whatever choice they made from the previous choice set. In 12 of 7,450 remaining choices, we observe the recorded choice was not available in the choice set likely because of refreshing. We would not be able to observe cases where the browser was refreshed and last period's choice was still available this period, but since that occurs about half the time, the total number of affected choices was likely around 25, or less than 0.35% of all choices. Dropping the cases we can identify has no impact on the results.

Table 1: Consideration Probabilities: Actual vs. Estimated

	Truth	Attentive Logit
Consideration Probability		
1	18.3%	10.7%
2	19.7%	13.5%
3	18.0%	12.3%
4	17.7%	12.2%
5	19.3%	14.4%
6	70.3%	68.1%
7	69.7%	67.4%
8	64.9%	65.2%
9	69.5%	67.1%
10	70.0%	64.9%

Notes: table compares the true average consideration probability for each good to the probability estimated in the attentive logit model.

Figure 1: Product Fixed Effects in Attention: Truth vs. ASC Model



elasticities below). The claim that the model recovers the true preference parameters requires assuming that the 'truth' is a conditional logit model estimated using realized choice sets. The conditional logit model estimated based on choices from all 10 goods gives a price effect of -0.05, less than a third of the value recovered from a logit model given actual choice sets. This is because the conditional logit model wrongly infers from the fact that high priced products are more likely to be considered (and thus chosen) that consumers do not really dislike high prices. The attentive logit

model gives a value of -0.20 (0.03) – the confidence interval includes the true value of -0.17. The conditional logit fixed effects are systematically biased because they conflate attention and utility. Products that are rarely in the choice set are assumed to be low utility. In contrast, attentive logit recovers the true fixed effects – the confidence interval on the attentive logit estimates includes the true values estimated as if choice sets are observed for products 2-9 with product 1 lying on the edge of the confidence interval. These intervals are relatively wide, but that is a feature, not a bug relative to the conditional logit model: the attentive logit model correctly recognizes that rare products are rare and that only limited information is available about how much consumers value them. The attentive logit confidence intervals on the less rare products (products 6-9 in the table) are reasonably precise.

An additional question of interest is whether the implied price elasticities differ from conventional models. We can compute the 'true' price elasticities (using the known consideration set probabilities as well as the preference parameters estimated given realized choice sets) and compare them to the price elasticities implied by a variety of models. The comparison with the conditional logit model is unfair in the sense that the attentive logit model is a strict generalization that includes more parameters. Thus we also compare the attentive logit model to a logit model with quadratic price parameters and a model with alternative-specific price parameters. The conditional logit model has 1 price parameter, the attentive logit model has 2 (β and γ), the quadratic model has 2, and the alternative-specific model has 10 (one for each good).

Figure 2 shows the own-price elasticities by good in each model. For goods 1-4, true own-price elasticities are positive because a higher price makes a good more likely to be considered. As noted above, this is an intentional feature of the model designed to mimic the fact that in some real world settings, consumers might be more likely to see higher priced items. Conditional on consideration, Table 2 shows that price responses are negative as expected. The logit and quadratic model both badly fail to characterize how elasticities vary across goods. With a separate price coefficient for each good, the product-specific model is able to capture these patterns as is the ASC model. But the product-specific model still performs badly in capturing cross-elasticities. The average magnitude of the 90 "true" cross-elasticities in the data is 0.090. The logit model has an average absolute deviation of 0.083, the quadratic model has an average deviation of 0.068, the product-specific model has an average deviation of 0.080, and the ASC model has an average deviation of 0.027, less than half of any of the alternative models. As a function of the original elasticities, the bias is 45.5 percentage points smaller.

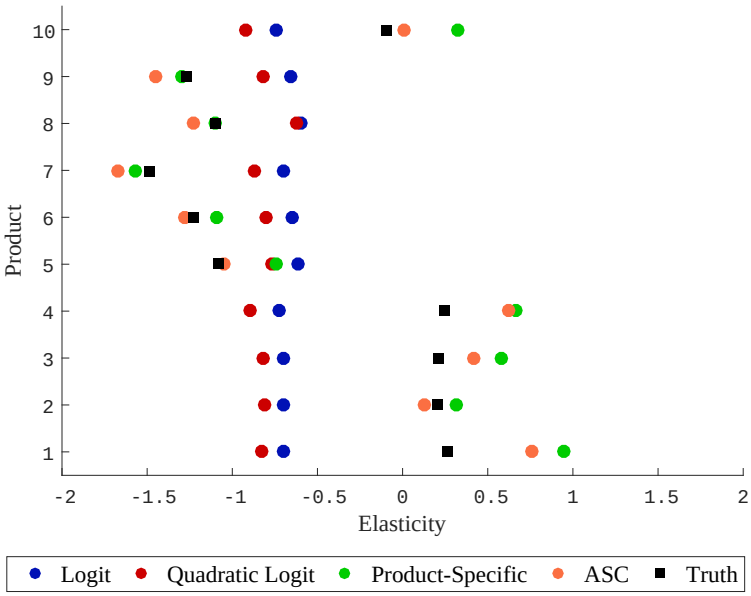
Of course, the game is rigged in that this setting is designed to perfectly match the underlying

Table 2: Experimental Data Estimation Results

	Conditional Logit	Attentive Logit	Truth
Utility:			
Price (dollars)	-0.054***	-0.196***	-0.173***
	(0.003)	(0.028)	(0.004)
Product 1	-1.411***	1.465***	0.368***
	(0.054)	(0.539)	(0.069)
Product 2	-1.955***	-0.065	-0.497***
	(0.069)	(0.478)	(0.080)
Product 3	-1.627***	0.625	0.093
	(0.059)	(0.476)	(0.073)
Product 4	-1.640***	0.629	0.088
	(0.060)	(0.466)	(0.073)
Product 5	-1.447***	0.707	0.306***
	(0.056)	(0.478)	(0.070)
Product 6	-0.435***	-0.737***	-0.581***
	(0.039)	(0.121)	(0.045)
Product 7	-0.855***	-1.280***	-1.075***
	(0.045)	(0.141)	(0.051)
Product 8	-0.662***	-1.185***	-0.909***
	(0.041)	(0.137)	(0.048)
Product 9	-0.316***	-0.561***	-0.405***
	(0.038)	(0.118)	(0.044)
Attention:			
Price (dollars)		0.137***	0.15
		(0.017)	
Product 1		-2.872***	-2.5
		(0.177)	
Product 2		-2.674***	-2.5
		(0.288)	
Product 3		-2.695***	-2.5
		(0.209)	
Product 4		-2.704***	-2.5
		(0.205)	
Product 5		-2.592***	-2.5
		(0.204)	
Product 6		0.152	0
		(0.192)	
Product 7		0.123	0
		(0.292)	
Product 8		0.258	0
		(0.230)	
Product 9		0.103	0
		(0.176)	

Notes: Table reports coefficient estimates from conditional logit and attentive logit models. Estimates are the coefficients in the utility and attention equations (not marginal effects). The conditional logit coefficients are recovered from estimating a model assuming all 10 possible goods are considered. The "true" utility parameters are estimated using a conditional logit model given the actual choice set consumers faced. The true attention parameters are known in advance. The attentive model also includes a constant. *** Denotes significance at the 1% level, ** significance at the 5% level and * significance at the 10% level.

Figure 2: Experimental Data: Own-Price Elasticities by Good



structural assumptions about consideration set probabilities. The subsequent exercises apply these models to real-world data and conduct various validation and specification checks which show that attentive logit models also perform well when consideration probabilities arise more organically.

4.4 Informative Advertising and Hotel Choice

We now apply the ASC model to data from online hotel choices made via Expedia.com. A subset of the data randomizes the search position in which hotels are displayed to consumers. We use this variation to test whether we are able to verify that search position impacts attention but not utility using our model, as well as to validate our framework as a tool for generating out of sample predictions of the efficacy of promotions that increase consumer awareness of products.

The ASC model fits well here because a product's ranking in search results depends on the observable attributes of that product. Before estimating the model, we demean the attributes at the individual level. This means that the attention probabilities depend on the *relative* value of the observed attributes compared to other options in each consumer's choice set, as they should if these probabilities arise from each product's placement in search results.[26]

[26]This demeaning would make no difference in a conventional logit model, but it can make a difference in an attentive logit model. Formally, the model with demeaned characteristics in attention violates the Goeree (2008) assumption that the attention probability for a given good does not depend directly on the characteristics of rival goods. These violations will disappear asymptotically as the number of goods becomes large.

Data The full dataset contains results from 166,036 consumer queries, including the hotels consumers were shown, attributes of those hotels, as well as whether they ultimately purchased the hotel. The main attributes we consider are price, star rating, review score, a location desirability score, whether there is an on-going promotion and the position of the hotel in the search results. The data span 54,877 hotels in 788 destinations. Ursu (2015) contains a detailed discussion and describes several sample selection restrictions designed primarily to clean the data (e.g. dropping all hotels with prices of less than $10 per night or more than $1,000 per night). We impose the exact same sample selection restrictions as Ursu (2015) with two exceptions: we restrict to the top 10 choices but we do not restrict to the 4 largest hotel destinations, which results in a much larger sample.[27]

After restricting to the sample with a randomized hotel ordering in search results, we end up with 2,441 total queries spanning 9,851 hotels for which we observe a final transaction.[28] Summary statistics from our data after all sample selection restrictions are imposed are reported in Table 8 in Appendix D. The average hotel costs about $160 a night, is rated 3.2 out of 5 stars, receives an average review score of 3.9 out of 5 from Expedia users, has a 74% chance of being from a popular brand and has a 20% chance of being currently undergoing a promotion (meaning that the sale price was noted as being lower than is typical). We can see that hotels which were actually chosen tend to be lower priced, more likely to be undergoing a promotion, and ranked higher in search.

Results Given that the order of the hotels was randomized, we might expect the position of the hotels in the search results to impact only attention and not utility. This need not be the case – Expedia did not inform consumers that the order was randomized so individuals may have believed that higher ranked hotels were better in some unobservable respect.

The estimation results for a conditional logit model and the attentive logit model are shown in Table 3. First, in both models all the coefficients have the expected sign – consumers dislike high prices and like hotels with more stars, higher review scores, better locations and higher positions in

[27] Among consumers who record a choice, 78% of consumers choose one of the top 10 hotels. In Appendix F, we estimate the model using the top 20 goods, which includes 94% of observed choices; the restriction to the top 10 goods is made for computational reasons.

[28] Many users search for a hotel and do not make an active choice. These consumers may purchase no hotel, they may search on another site, or they may search again on Expedia. Since the data is indexed by search impression and not by user, we cannot tell how many times each user searched. This raises two issues for our analysis. First, the sample of users who purchased a hotel may be selected. Formally, the logit IIA assumption implies that restricting only to beneficiaries who chose one of the observed goods should not alter our results, but this assumption may fail – in Appendix F, we rerun our main specification including search impressions for which we record no choice (recorded as choosing the outside option). Second, we cannot characterize optimal behavior on the part of firms without knowing whether consumers ultimately return and book a room. However, our goal is a more narrow one: we seek to show that the ASC model's decomposition of demand into utility and attention is consistent with our intuitive understanding of attention and has predictive power in describing what will happen when attention is (randomly) perturbed.

search. The conditional logit model implies that their responsiveness to a hotel moving from search position 10 to search position 1 is about the same as an \$80 – or 50% – decrease in the price per day.

The attentive logit model shows that the impact of search position on choices comes entirely through the impact on attention rather than utility.[29] The model also implies that consumers are much more likely to consider hotels that have a desirable location score. This makes intuitive sense and is consistent with a world in which consumers make a query, find the hotels located nearby their destination, and then compare prices and other attributes to come to their final choice.

Table 4 shows how choice probabilities and attentive probabilities vary with the ranking. The model suggests that the attentive probability ranges from 21.7% for a hotel in the 10th position to 52.9% for the highest ranked hotel (the choice probability increases by a factor of 3 – which differs from the ratio of average attention probabilities due to Jensen's inequality). We also compare the price elasticities estimated in the conditional logit model with the attentive logit model. The logit model modestly attenuates own-price elasticities, with an average error of about 10%. This arises because consumers are insensitive to price variation for goods to which they are inattentive. More generally, the direction of the bias in own-price elasticities is ambiguous and depends on the correlation between prices and attention probabilities (which is empirically close to zero in this case).

As an increase in the position of one product necessarily entails a decline in the position of alternative products, it is not possible to perturb position for one product while holding everything else fixed. Identification comes from the fact that we observe position changing for hotels of varying popularity – if position increases from 4 to 3 for a given hotel at the expense of an extremely unpopular hotel that was unlikely to be chosen in either case, this reveals the impact of perturbing position holding everything else fixed.

Given the estimated coefficients, we compute the estimated cross-derivatives with respect to the position variable. These can be thought of as the impact of increasing position for a given hotel and then using the model to "undo" the impact on demand of the resulting position changes for rival hotels. For each individual and each pair of hotels, we can compute the magnitude of the asymmetry as a percentage of the average absolute cross-effect in the data. Figure D.2 in Appendix D graphs the resulting asymmetries. The mean asymmetry is 68.7% of the average absolute cross-effect (and in the model, it is never identically zero). A model which ignores these asymmetries would badly misspecify substitution patterns.

[29] In our robustness checks in Appendix F with 20 goods per beneficiary or an outside option, we do find that search position significantly enters utility, but the implied willingness to pay in all cases is at least 75% smaller in magnitude than that implied by the conditional logit model.

Table 3: Expedia Data: β and γ

	Conditional Logit	Attentive Logit
Utility:		
Price (dollars)	-0.015***	-0.025***
	(0.001)	(0.003)
Hotel Stars (1-5)	0.566***	0.805***
	(0.044)	(0.138)
Hotel Review Score (1-5)	0.410***	0.768***
	(0.049)	(0.183)
Popular Brand Indicator	0.075	0.344**
	(0.058)	(0.161)
Location Score (normalized)	0.695***	0.249**
	(0.047)	(0.109)
Ongoing Promotion Indicator	0.191***	0.065
	(0.057)	(0.156)
Position in Search	-0.104***	-0.002
	(0.008)	(0.027)
Attention:		
Price (dollars)		-0.001
		(0.001)
Hotel Stars (1-5)		0.092
		(0.106)
Hotel Review Score (1-5)		-0.004
		(0.115)
Popular Brand Indicator		-0.180
		(0.179)
Location Score (normalized)		0.813***
		(0.129)
Ongoing Promotion Indicator		0.195
		(0.170)
Position in Search		-0.154***
		(0.022)
Constant		0.358
		(0.532)

Notes: Table reports coefficient estimates from the Goeree (2008) model. Estimates are the coefficients in the utility and attention equations (not marginal effects). Standard errors are in parentheses. ***Denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. The model is also includes a default which is a randomly chosen alternative for each consumer. Given the estimated attention probabilities, this default is chosen less than 1% of the time.

Out of Sample Validation Because we recover the probability of attention for each good, we can ask – for which goods is it the case that demand is substantially higher if the probability goes to 1? This exercise provides a bound on the potential effectiveness of informative advertising. The random assignment in the Expedia data provides a natural experiment that we can use to test that bound.

To do so, we estimate the model using only the hotels in search positions 3 through 10. We then compute demand for each hotel if the constant in the attention equation becomes arbitrarily large

Table 4: Expedia Data: Choice Probabilities and Elasticities

			Own-Price Elasticity	
	Market Share	Attention Probability	Conditional Logit	Attentive Logit
Search Position				
1	0.178	52.9%	-2.1	-2.32
2	0.154	49.9%	-2.11	-2.35
3	0.122	45.5%	-2.18	-2.4
4	0.101	41.8%	-2.17	-2.42
5	0.089	37.7%	-2.24	-2.47
6	0.078	34.4%	-2.23	-2.5
7	0.074	30.5%	-2.23	-2.52
8	0.06	27.2%	-2.3	-2.56
9	0.08	24.2%	-2.28	-2.55
10	0.062	21.7%	-2.33	-2.54

Notes: the market share variable gives the average market share of hotels in each position in the randomized Expedia data. Attentive Probability gives the average attentive probability for hotels in that position implied by estimation of the Goeree (2008) model. Conditional Logit Elasticity is the average own-price elasticity of hotels in the first position in the conditional logit model and Attentive Logit Elasticity is the corresponding elasticity in the attentive logit model (taking into account the impact of characteristics on both utility and attention).

so that the attentive probability goes to 1 – this is our bound for that hotel. We then ask how well the bound does in accounting for the observed behavior in positions 1 and 2. While we cannot know ex ante how the attention probability will change if a hotel is placed in positions 1 or 2, we know that demand in those positions should be less than the bound given by perfect attention. Thus, we ask first whether the bound implied by the ASC model is indeed a bound on choice probabilities for hotels in positions 1 and 2 and second, whether this bound has predictive power in accounting for the choice probabilities conditional on observed demand.

In practice, we compute this bound separately for each hotel in the data, but we collapse down to categories of hotels for expository purposes. The upper bound on the effectiveness of informative advertising is given by transaction probabilities when the the probability of paying attention is one. Figure D.3 in Appendix D shows how this bound compares to the observed demand for a variety of different types of hotels in each search position. The thick horizontal line shows the bound, the 10 colored dots show demand in each search position (with higher dots corresponding to lower search positions). The main takeaways from this figure are first that demand is always less than the bound implied by perfect attention and second that the bound is non-trivial. For example, average demand for hotels in positions 1-3 exceeds the bound placed on the demand for the maximum price hotels.

Finally, we ask whether the bound has predictive power – if we see two hotels with the same level of demand in positions 3 - 10, will the hotel with the larger bound experience a larger increase in demand if it is randomly assigned to search position 1 or 2? Table 5 shows that the answer is yes. Specification (1) shows that across hotels, the bound constructed from the model estimated

on positions (3)-(10) predicts demand in positions (1) and (2). Specifications (2)-(4) show that it continues to have predictive power even after we condition on the observed choice probability for that hotel in positions 3-10 as well as the choice probability implied by a logit model given the choice set and the characteristics of the hotel in question. Thus, the attentive logit model can be used to forecast which products will benefit from informative advertising given their current level of utility according to a conditional logit model.

Table 5: Expedia Data: Regression of Choice Probability in Positions 1-2 on Hotel Characteristics

	(1)	(2)	(3)	(4)
Bound	0.576***	0.588***	0.225***	0.261***
	(0.035)	(0.040)	(0.074)	(0.086)
Hotel Prob (pos $<$ 2)		0.057		0.044
		(0.030)		(0.030)
Logit (pos $<$ 2)			0.712***	0.660***
			(0.143)	(0.166)
Number of Hotels	4882	3722	4882	3772

Notes: Table reports coefficients from a regression at the hotel level of the transaction probability of that hotel in positions 1 and 2 on hotel-level covariates. 4,882 hotels appeared in the data in positions 1 and 2 and 3,722 of these also appeared in positions 3-10 (the bound can be constructed for hotels based on their characteristics and the estimated model coefficients even if they did not appear in positions 3-10). "Bound" indicates the alogit forecast of demand for a hotel with those characteristics with attention probability 1. Hotel Prob (pos $<$ 2) is the empirical choice probability in positions 3-10 if available. Logit (pos $<$ 2) is the logit choice probability given the observed characteristics of the hotel and the coefficients estimated on hotels in positions 3-10..

5 Default-Specific Consideration

Another popular specification in the applied literature assumes the existence of a default good among 'inside' goods (i.e. goods for which observable characteristics vary) and allows the probability of considering all other options to vary only as a function of the characteristics of that default good. Only if the default good becomes sufficiently unattractive will consumers incur the cost required to observe all other options in the market. This model can be used to identify whether inertia, and choice of defaults more generally, arises because consumers do not consider other options (and thus might be better off if they switched) or because consumers are actively choosing not to switch due to adjustment costs or persistent unobserved heterogeneity. The DSC model has been used to study inertia in health insurance and residential electricity markets (Ho, Hogan, and Scott Morton 2015; Heiss, McFadden, Winter, Wupperman, and Zhou 2016; Hortaçsu, Madanizadeh, and Puller 2015). These earlier studies either assume that only one of adjustment costs or inattention are operative or rely on additional exclusion restrictions, such as that attentive consumers respond only to the level

of prices and not to changes over time.

We start by laying out parametric assumptions which facilitate estimation of DSC models. Next, we discuss how DSC models relate to full-consideration models and use this representation to develop an overidentification test. Finally, we estimate the DSC model in Medicare Part D data and conduct two separate overidentification tests, testing the internal consistency of our model and the validity of instruments used in the prior literature.

5.1 Parametric Assumptions

To nest the DSC model in our general framework, let $\phi_{ij} = 1$ for all $j > 0$. Consideration set probabilities then take the following form:

$$\pi_i(\mathcal{J}) = \mu_i(x_{id}) \tag{5.1}$$

$$\pi_i(d) = 1 - \mu_i(x_{id}) \tag{5.2}$$

$$\pi_i(c) = 0 \quad \text{for c} \notin \{\mathcal{J}, d\} \tag{5.3}$$

The probability of selecting option j is expressed as:

$$s_{ij} = (1 - \mu_i)\, 1(j = d) + \mu_i s^\star_{ij} \tag{5.4}$$

where $s^\star_{ij}$ denotes the probability of choosing j conditional on considering all available goods. In other words, consumers pay attention to all goods in the choice set with probability μ, which is a function of default characteristics only. Conditional on paying attention, consumers make an active choice which might include choosing the default if its characteristics compare favorably to alternatives and/or switching costs are high.

Following the argument given in Section 3, the probability of considering all goods, μ_i, is constructively identified from cross derivative differences:

$$\frac{\partial \log(\mu_i)}{\partial x^1_{id}} = \frac{1}{s_{ij}}\left[\frac{\partial s_{ij}}{\partial x^1_{id}} - \frac{\partial s_{id}}{\partial x^1_{ij}}\right] \tag{5.5}$$

$$\mu_i = \exp\left(-\int \frac{1}{s_{ij}}\left[\frac{\partial s_{ij}}{\partial x^1_{id}} - \frac{\partial s_{id}}{\partial x^1_{ij}}\right] dx^1_{id}\right) \tag{5.6}$$

These expressions are similar to those in the ASC model but do not involve leave-one-out market share differences. While nonparametric identification in the ASC models requires that we observe choice sets where some goods are unavailable or unlikely to be chosen, no such variation is required

to identify changes in attention probabilities in the DSC model.

As in the ASC model, we make standard functional form assumptions to bring the framework to data in a convenient form. Suppose, as in Ho, Hogan, and Scott Morton (2015), that consumers are inattentive whenever:

$$x_{id}\beta + \epsilon_{id} > f(z_i) + v_i \tag{5.7}$$

where x_{id} are characteristics of the default good and z_i is a vector of other individual characteristics and ϵ_{id} and v_i are both type 1 extreme value. The probability of being inattentive is then:

$$1 - \mu_i = \frac{\exp(x_{id}\beta)}{\exp(f(z_i)) + \exp(x_{id}\beta)} \tag{5.8}$$

Note first that we do not need to observe any additional individual characteristics in order to estimate this model. We can assume that $f(z_i) = 0$ and the model is still identified. Including individual characteristics just produces a more flexible model of inattention and thus reduces the likelihood that the error term is misspecified. With these functional form assumptions, it is straightforward to estimate the model by maximum likelihood.

With the logit functional form assumption, the expression for cross-derivative differences in Equation 5.5 takes the form:

$$\frac{\partial s_{ij}}{\partial x_{id}} - \frac{\partial s_{id}}{\partial x_{ij}} = \gamma(1 - \mu_i)s_{ij} \tag{5.9}$$

Conditional on the probability of paying attention, cross-derivatives are therefore predicted to scale in a particular way with market shares. Whether cross-derivatives follow this pattern in the data can thus be used as a test of the model.

As it is infeasible to nonparametrically estimate cross-derivatives for choice sets with a large number of goods (in our application, consumers face a choice of up to 50 insurance plans), we compare the predictions of the DSC model to an alternative discrete choice model that nests the DSC model and which includes flexible interaction terms between the characteristics of every good and the default good allowing us to flexibly model cross-elasticities. To estimate this generalization, it will be convenient to first derive a random utility representation of the DSC model.

5.2 Relationship to Full-Consideration Models

Like the ASC model, the DSC model is equivalent to a standard logit model with an additional inertial term through which the utility of good j depends directly on the characteristics of rival

goods:

$$u_{ij} = x_{ij}\beta + \xi_{i,j=d} + \psi_{i,j=d} + \epsilon_{ij} \tag{5.10}$$

where $\psi_{i,j=d}$ takes the value ψ_i for plan d and is 0 otherwise. We show in Appendix B that ψ_i is given by:

$$\begin{aligned} \psi_{j=d} &= \ln\left(\frac{s_{id}}{1-s_{id}}\right) - \ln\left(\frac{s^\star_{id}}{1-s^\star_{id}}\right) \\ &= \ln\left(\frac{1+(1-\mu_i)\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta)}{\mu_i}\right) \end{aligned} \tag{5.11}$$

where $s^\star_{id}$ is the probability of choosing the default conditional on paying attention. The term $\xi_{i,j=d}$ in this model represents all of the reasons why an attentive consumer might nonetheless prefer to choose the same plan – for example, because there are switching costs or persistent unobserved heterogeneity.[30] This is the normatively relevant component of inertia. The $\psi_{i,j=d}$ term by contrast captures the possibility that the consumer chose the default plan not because it had higher utility, but simply because they were inattentive to the available options. The first line of Equation 5.11 shows that $\psi_{i,j=d}$ can be written explicitly as a monotonic transformation of observed choice probabilities minus what choice probabilities would be were consumers fully attentive.[31]

We can utilize this random utility representation of the DSC model to develop a parametric overidentification test. The random utility representation shows that we can equivalently write the DSC model as a model where $u_{id} = A_{id} + \epsilon_{id}$, $u_{ij} = A_{ij} + \epsilon_{ij}$, with $A_{ij} = x_{ij}\beta$ and $A_{id} = x_{ij}\beta + \xi_i + \psi_i$. To allow for a more flexible model, we first estimate the DSC model to recover $\hat{\beta}_i$, $\hat{\xi}_{i,j=d}$ and $\hat{\psi}_{ij}$ (and thus $\hat{A}_{id}$ and $\hat{A}_{ij}$), then estimate:

$$\begin{aligned} u_{id} &= \hat{A}_{id} + \epsilon_{id} \\ u_{ij} &= \hat{A}_{ij} + \sum_k \sum_{k'} x_{idk} x_{ijk'} \alpha_{k,k'} + \epsilon_{ij} \text{ for } j \neq d \end{aligned} \tag{5.12}$$

where $\hat{A}_{id} = x_{id}\hat{\beta} + \hat{\xi}_{i,j=d} + \hat{\psi}_{i,j=d}$ is the predicted component of utility from the DSC model which

[30] The distinction between switching costs and persistent unobserved heterogeneity is relevant for some counterfactuals but not others. Separately identifying these factors would be important to predicting how many consumers would switch back to their original plan were they defaulted into an alternative plan, but these need not be separated if we only want to predict whether consumers would switch were they fully informed about possible alternatives.

[31] Note that if we observed some subset of consumers that we knew were paying attention *and* we knew had exactly the same preferences and choice set as inattentive consumers, then we could estimate $s^\star_{ij}$ and directly compute ψ_{ij}. In practice however, this condition is unlikely to be met. Consider the context of health insurance plan choice. One might consider using the choices of new enrollees making a de novo choice to estimate $s^\star_{id}$. This method would incorrectly assume returning enrollees have no true adjustment costs or persistent unobserved preferences. The proof in section 3 shows that this model is identified without these restrictive assumptions.

is now regarded as a known constant. If the model fit perfectly, we would observe $\alpha_{k,k'} = 0$ for all k, k'. We can test whether the cross derivative differences implied by the more flexible model are statistically different from those implied by the ASC framework as a test of the model. In Appendix E, we derive an analytical expression for $\frac{\partial s_{ij}}{\partial x_{id}} - \frac{\partial s_{id}}{\partial x_{ij}}$ which we use to evaluate how the cross-derivatives implied by the DSC model compare to the cross-derivatives implied by the more flexible model in Section 5.3.

5.3 Adjustment Costs and Inattention in Medicare Part D

We apply the model to evaluate whether the observed inertia in Medicare Part D plans is due to inattention, adjustment costs or both. Medicare Part D plans provide prescription drug insurance to elderly beneficiaries in the United States. Beneficiaries in the median choice set have 48 Medicare plans they can choose among, including both plans which provide only prescription drug coverage and plans which provide broader medical insurance ("Medicare Advantage"). Our analysis focuses on the stand-alone prescription drug insurance plans (PDP plans). 90% of beneficiaries choose to remain enrolled in the same plan as last year (Abaluck and Gruber 2016). An important question is whether this is because those beneficiaries would be worse off if they switched plans (because they like the plan they chose or have high adjustment costs) or because they are not paying attention and would switch if they understood that they could save money with rival plans. We estimate the DSC model described in Section 5.1 to separately identify an inertial term $\xi_{j=d}$ and an adjustment cost term. In addition, we perform the overidentification test described in Section 5.2 to evaluate whether the patterns of asymmetry in our data are consistent with the underlying model of inattention.

Heiss, McFadden, Winter, Wupperman, and Zhou (2016) perform a similar exercise but additionally rely on assumptions that some variables impact attention and not utility. We instead rely only on the asymmetry between how the market share of the default plan responds to prices of alternative plans relative to how the market shares of alternative plans respond to prices of the default plans. We also test one of the exclusion restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016) – that characteristic levels impact utility while characteristic changes conditional on levels impact only attention – and we find that this restriction is roughly satisfied in our data.

Data We use administrative data from a 20% sample of Part D beneficiaries. The full dataset contains 7.2 million Medicare eligible beneficiaries (a 20% sample of all Part D beneficiaries from 2006-2009). We use the sample selection approach described in Abaluck and Gruber (2016) and consider choices from 2007-2009. We impose a number of restrictions to isolate beneficiaries who get no Part D coverage from their employer and no low income subsidies; we take a further random

2% sample of the remaining beneficiaries for computational reasons. In the end, we are left with 30,937 beneficiaries choosing from an average of 40 prescription drug insurance plans.

Some of the variables we include in our choice model, such as premiums or plan quality ratings, are directly observable. Plans also differ on a variety of dimensions related to the amount of coverage they provide – they have different lists of covered drugs (formularies) and different copays and coinsurance rates for the drugs that are covered. Abaluck and Gruber (2016) summarize these features by constructing a "calculator" that can be used to determine given the totality of each plan's coverage characteristics what out of pocket costs would be for that plan for a given set of claims. Given this calculator, several alternative measures of expected out of pocket cost and the variance of out of pocket costs are constructed. We use the "rational expectations" measure based on a forecast of what costs will be in the coming year given other individuals who look similar at the start of the year. Summary statistics from our data after all sample selection restrictions are imposed are reported in Table 9 in Appendix D. We report the mean and standard deviation of a variety of characteristics for all plans and also for chosen plans.

To address concerns about endogeneity, we observe and include in our model much of the publicly available information that might be used by individuals to make their choices – including premiums, deductibles, donut hole coverage, as well as various measures of formulary completeness and cost sharing. This approach is standard in the recent literature on health plan choices (Handel and Kolstad 2015; Heiss, Leive, McFadden, and Winter 2013; Abaluck and Gruber 2011; Abaluck and Gruber 2016). In our baseline specification, we do not include brand fixed effects for computational reasons. In Appendix F, we replicate our main specification restricting only to brands chosen by at least 400 beneficiaries in our data and including brand fixed effects – we show that we estimate similar adjustment costs and inattention.

Estimation Results Table 6 shows the results of estimating a conditional logit model and the DSC model in the Part D data. The conditional logit results resemble those in Abaluck and Gruber (2016). Consumers dislike premiums and out of pocket costs, and even conditional on the out of pocket cost consequences they dislike deductibles. A few coefficients have unexpected signs relative to prior work – for example, in 2007 and 2009 consumers appear to favor plans with less favorable average cost-sharing features. Most notably for our purposes, they are willing to pay between $1,000 and $1,400 depending on the year to choose the same plan they chose in the previous year (obtained by dividing the coefficient on the prior year plan dummy by the coefficient on premiums to express the effect in dollar terms).

The attentive logit model coefficients have (mostly) the same sign as the conditional logit coef-

ficients with a few exceptions where unexpected signs in the conditional logit model become right-signed in the attentive logit model. We now see generally the characteristic pattern reported in Abaluck and Gruber (2016): even conditional on out of pocket cost consequences, consumers prefer plans with nominally desirable plan features like lower deductibles, donut hole coverage, and lower cost sharing.[32]

The attentive logit coefficients are also typically larger in magnitude, reflecting the fact that conditional on paying attention, observables in the attentive logit model explain a greater share of choices relative to unobservables. The impacts of default characteristics on attention probabilities have mostly the expected signs: consumers are more likely to pay attention if the default plan has higher premiums or out of pocket costs, has a higher variance of costs (less risk-protection), has a higher deductible or a lower quality rating. For a few other variables, the sign switches from year to year.

The attentive logit model implies that most of the observed degree of inertia is due to inattention. The average attentive probability in the data is 11.4%, which would imply an inertia rate of 88.6% just from inattention. The actual inertial rate is 90.74%. This implies that of the 11.4% of consumers making an active choice, almost 24% chose the default plan. Thus, the model continues to imply non-trivial adjustment costs, at least in some years. In 2007, we estimate adjustment costs of \$0 (the observed degree of inertia is almost fully explained by inattention),[33] while in 2008 and 2009 we estimate adjustment costs of around \$300 and \$200 respectively. These are in the range of the average cost savings estimated in Abaluck and Gruber (2016) from every beneficiary switching to the lowest cost plan. This implies that if all beneficiaries were assigned to the lowest cost plan, the adjustment costs would roughly offset the cost savings leaving consumers no better off.

The above results suggest that a model that permits asymmetries in demand responses fits the data better than one that does not. However, do we see the specific pattern of asymmetries which is characteristic of our particular model of inattention? To test this, we implement the overidentification test outlined in Section 5.2. We estimate a model which flexibly parametrizes the

[32] Abaluck and Gruber (2016) estimates a conditional logit specification that includes interactions between the prior year plan dummy and default plan characteristics – in that study, those interactions were included as an ad hoc way of controlling for the fact that the decision to switch might be driven by different factors than the choice of plans conditional on switching (the conditional logit results conflate the two). The attentive logit specification deals with this in a more principled way through an explicit model of inattention. For this reason, only the attentive logit results show the characteristic oversensitivity to premiums relative to out of pocket costs that emerges in Abaluck and Gruber (2016) when the coefficients are identified using the choices of active choosers. This difference also explains a handful of coefficients with unexpected signs, such as the coefficient on donut hole coverage in 2007, which accord with the pattern reported in Abaluck and Gruber (2016) in the attentive logit model.

[33] In the attentive logit model, this shows up as a large negative and imprecisely estimated adjustment costs term. This is because, conditional on inattention fully explaining the observed degree of inertia, the model cannot distinguish between adjustment costs of zero and adjustment costs of negative infinity, both of which would imply little or no *additional* choice of the default plan beyond that which arises from inattention. If we bound true adjustment costs from below at zero, then this estimate implies adjustment costs of zero.

Table 6: Part D Data: Conditional Logit and Attentive Logit Estimates

	2007		2008		2009	
	Clogit	Alogit	Clogit	Alogit	Clogit	Alogit
Utility:						
Annual Premium (hundreds)	-0.415***	-0.909***	-0.596***	-1.074***	-0.599***	-1.245***
	(0.012)	(0.029)	(0.013)	(0.026)	(0.015)	(0.027)
Annual Out of Pocket Costs (hundreds)	-0.418***	-0.661***	-0.691***	-0.923***	-0.433***	-0.484***
	(0.020)	(0.028)	(0.029)	(0.047)	(0.034)	(0.054)
Variance of Costs (millions)	-2.131***	-3.359***	-1.809***	-2.351***	-2.056***	-0.702
	(0.178)	(0.248)	(0.299)	(0.448)	(0.326)	(0.526)
Deductible (hundreds)	-0.208***	-0.355***	-0.737***	-0.792***	-0.231***	-0.590***
	(0.024)	(0.032)	(0.027)	(0.037)	(0.030)	(0.043)
Donut Hole Coverage	-0.178***	0.505***	-0.263***	-0.798***	1.335***	1.917***
	(0.055)	(0.074)	(0.065)	(0.120)	(0.083)	(0.142)
Average Consumer Cost Sharing %	0.704**	-0.071	-2.002***	-4.274***	0.798**	-1.898***
	(0.280)	(0.376)	(0.333)	(0.450)	(0.358)	(0.541)
# of Top 100 Drugs in Formulary	0.641***	1.078***	0.749***	0.826***	-0.060***	0.022*
	(0.040)	(0.071)	(0.046)	(0.057)	(0.008)	(0.013)
Normalized Quality Rating	0.087***	0.319***	0.299***	0.688***	0.564***	0.659***
	(0.017)	(0.025)	(0.018)	(0.028)	(0.017)	(0.026)
Prior Year Plan	5.930***	-15.619	6.380***	3.370***	6.525***	2.410***
	(0.025)	(846.880)	(0.034)	(0.122)	(0.038)	(0.208)
Attention:						
Annual Premium (dollars)		0.240***		0.364***		0.068**
		(0.016)		(0.023)		(0.027)
Annual Out of Pocket Costs (dollars)		0.141***		0.186***		-0.029
		(0.038)		(0.051)		(0.064)
Variance of Costs (millions)		2.037***		-0.113		1.777***
		(0.315)		(0.455)		(0.589)
Deductible (hundreds)		0.373***		0.182***		0.075
		(0.046)		(0.053)		(0.065)
Donut Hole Coverage		0.829***		-1.364***		-0.268*
		(0.082)		(0.128)		(0.142)
Average Consumer Cost Sharing %		1.321**		-5.493***		0.060
		(0.538)		(0.693)		(0.733)
# of Top 100 Drugs in Formulary		-0.211***		0.429***		0.099***
		(0.065)		(0.102)		(0.021)
Normalized Quality Rating		0.002		0.034		-0.600***
		(0.024)		(0.036)		(0.032)

Notes: "Clogit" refers to the conditional logit model; "alogit" refers to the attentive logit model. The table reports coefficient estimates from the DSC model. Estimates are the coefficients in the utility and attention equations (not marginal effects). The coeffic1ents in the attention equation are the coefficients on the listed characteristics of the default good (demeaned). Standard errors are in parentheses. The attentive model also includes a constant. *** denotes significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level. Standard errors in parentheses.

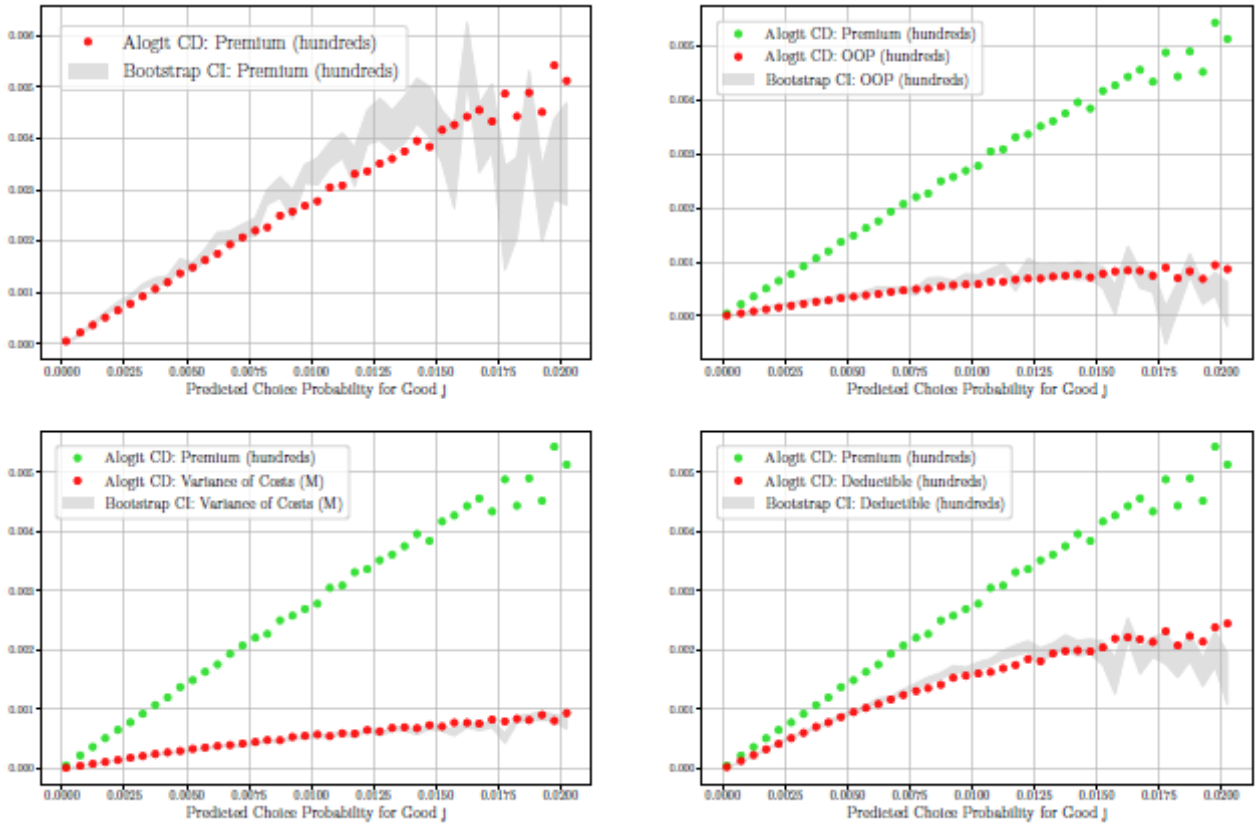
asymmetries in the data and compare the results to the more parsimonious attentive logit model.

Figure D.4 gives the predicted cross derivative difference between default and non-default goods for included plan characteristics for four variables; the charts for all variables are in Appendix D. We graph both the estimated cross derivatives from Equation 5.12 and the cross-derivatives implied by the DSC model ($\gamma_k(1-\mu_{i0})\hat{s}_{ij}$) against the predicted market share of plan j, $\hat{s}_{ij}$. To capture the uncertainty in the estimated cross-derivatives, we bootstrap estimation of Equation 5.12 and graph the resulting confidence interval.

In all graphs, the green dots indicate the empirical cross-derivatives with respect to premiums –

this is exactly the same data in all graphs, and is included for scale (the green dots are absent in the premium graph itself since they would overlap perfectly with the red dots). For each variable, the red dots indicate the predicted cross-derivative difference from the DSC model and the grey confidence region indicates the "empirical" cross-derivative difference from the more flexible specification in equation 5.12. We can see that in nearly all cases, the DSC model cross-derivatives match up well with empirical cross-derivatives. There are a few exceptions – for example, there are some nonlinearities in the cross-derivatives with respect to the quality rating which are not well-accounted for by the underlying model of inattention. But overall, the patterns in the cross-derivatives are extremely well-explained by the relatively parsimonious model of inattention.

Figure 3: Empirical vs. Model Predicted Cross-derivatives



6 Conclusion

Discrete choice models with consideration sets relax the strong assumption that beneficiaries consider all of the options available to them before making a choice. In the literature to date, such models have been identified either by bringing in auxiliary information on what options consumers consider

or assuming that some characteristics impact attention or utility but not both. This paper shows that these assumptions are unnecessary. We show that a broad class of such models are identified from variation already available in the data. Consideration set probabilities are constructively identified from asymmetries in the matrix of cross-derivatives of choice probabilities with respect to characteristics of rival goods.

We illustrate a number of practical applications of the model. In a lab experiment, using only data on observed choices, we recover consideration probabilities and obtain accurate estimates of the elasticities that we would estimate if we observed consideration sets. In data from Expedia, we show that the latent 'attention probability' corresponds to what we would intuitively call attention; perturbing the order of items in search rankings impacts attention but not utility. The model can also predict which items will experience larger demand increases if advertised conditional on current demand. In data from Medicare Part D, we validate the model further by demonstrating that the cross-derivative asymmetries follow the specific pattern predicted by our model of inattention. Our model implies that, while most inertia is driven by inattention in this market, there remain non-trivial adjustment costs.

Our results highlight that consideration set models are able to capture substitution patterns that are present in the data which conventional models rule out. There may be large asymmetries in cross-derivatives with respect to some characteristics. Failing to allow for these asymmetries may lead other parameters, such as own-price elasticities, to be misspecified. Extensions to traditional discrete choice models will still be unable to capture the substitution patterns permitted by consideration set models unless the characteristics of all rival goods are allowed to enter utility directly for each good. Consideration set models represent a more parsimonious extension to standard discrete choice models, and they make systematic and testable restrictions on how asymmetries in the cross-derivatives vary across characteristics for different goods.

It is important to note that consideration set models are not necessarily behavioral – search costs or unobserved constraints can lead to a lack of full consideration, and inattention of any kind can be rationalized by sufficiently large search or computational costs (Simon 1971). That being said, consideration set models relax 'full' rationality in the sense that consumers are not necessarily choosing the best option given their utility functions and the choices observable to the econometrician. Unobserved constraints aside, this allows for a general measure of the quality of consumers' choices, measured as the welfare loss relative to what consumers would choose given full consideration with the estimated utility function. While it is sometimes argued that relaxing full rationality leads to a lack of discipline, we show that the consideration set models we develop are

over-identified and their validity can thus be empirically determined.

Our results are sufficient but not necessary for identification of consideration probabilities – one direction for future work is to characterize more generally when consideration probabilities can be recovered from choice data. We hope that the sufficient conditions given here will make it possible to adapt consideration set models to a wider range of settings than they have previously been applied. With additional structure, such models can be used to identify parameters of interest such as search costs, and they enable us to construct counterfactuals and explore normative questions that would not be possible in conventional models. We can ask, for example, how might beneficiaries choose if they considered all available options? What is the potential value of information? When choices correlate with cognitive ability, is this because cognitive ability impacts preferences or because it impacts consumers' ability to consider all options? Do some demographic or choice set features (such as the number of plans) increase the likelihood that consumers are attentive? How much better off might consumers be if they were fully informed about the relevant choices? We hope that future work will explore these questions in more detail.

References

Abaluck, J. and J. Gruber (2011). Choice inconsistencies among the elderly: evidence from plan choice in the medicare part d program. *The American economic review 101*(4), 1180–1210.

Abaluck, J. and J. Gruber (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *The American Economic Review 106*(8), 2145–2184.

Berry, S. T. and P. A. Haile (2009). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research.

Caplin, A. and M. Dean (2015). Revealed preference, rational inattention, and costly information acquisition. *The American Economic Review 105*(7), 2183–2203.

Caplin, A., M. Dean, and J. Leahy (2016). Rational inattention, optimal consideration sets and stochastic choice. Technical report.

Chade, H. and L. Smith (2006). Simultaneous search. *Econometrica 74*(5), 1293–1307.

Chen, H. A., D. Levy, S. Ray, and M. Bergen (2008). Asymmetric price adjustment in the small. *Journal of Monetary Economics 55*(4), 728–737.

Conlon, C. T. and J. H. Mortimer (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics 5*(4), 1–30.

Crawford, G. S., R. Griffith, and A. Iaria (2016). Demand estimation with unobserved choice set heterogeneity.

Davis, P. and P. Schiraldi (2014). The flexible coefficient multinomial logit (fc-mnl) model of demand for differentiated products. *The RAND Journal of Economics 45*(1), 32–63.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics 129*(4), 1661–1710.

Gaynor, M., C. Propper, and S. Seiler (2016). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review*.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Handel, B. R. and J. T. Kolstad (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *The American Economic Review 105*(8), 2449–2500.

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

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica 73*(3), 669–738.

Heiss, F., A. Leive, D. McFadden, and J. Winter (2013). Plan selection in medicare part d: Evidence from administrative data. *Journal of Health Economics 32*(6), 1325–1344.

Heiss, F., D. McFadden, J. Winter, A. Wupperman, and B. Zhou (2016). Inattention and switching costs as sources of inertia in medicare part d. Working Paper.

Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics 5*(1), 123–144.

Ho, K., J. Hogan, and F. Scott Morton (2015). The impact of consumer inattention on insurer pricing in the medicare part d program. Technical report, National Bureau of Economic Research.

Honka, E. (2014). Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics 45*(4), 847–884.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2015). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *Consumer Awareness, and Choice: Evidence from the US Banking Industry (September 16, 2015)*.

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2015). Power to choose? an analysis of consumer inertia in the residential electricity market. Technical report, National Bureau of Economic Research.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(1), 145–177.

Lewbel, A. (2007). Endogenous selection or treatment model estimation. *Journal of Econometrics 141*(2), 777–806.

Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica 82*(3), 1153–1176.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012). Revealed attention. *The American Economic Review 102*(5), 2183–2205.

Matejka, F. and A. McKay (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review 105*(1), 272–298.

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

Simon, H. A. (1971). Designing organizations for an information-rich world.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics 50*(3), 665–690.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Treisman, A. M. and G. Gelade (1980). A feature-integration theory of attention. *Cognitive psychology 12*(1), 97–136.

Ursu, R. M. (2015). The power of rankings: Quantifying the effects of rankings on online consumer search and choice. *Available at SSRN 2729325*.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.

A Model & Identification Proof

A.1 Results to Complement Section 3

Corollary 1. *Symmetry of Cross Derivatives*:

$$\frac{\partial s^{\star}_{ij}}{\partial x^{1}_{ij'}} = \frac{\partial s^{\star}_{ij'}}{\partial x^{1}_{ij}} \tag{A.1}$$

Proof: Differentiating the market share of good j with respect to the quasilinear characteristic of good j' gives:

$$\frac{\partial s^{\star}_{ij}}{\partial x^{1}_{ij'}} = \int -\beta_i \int ... \left[\int_{-\infty}^{v_{ij}+e-v_{ij}}\right] ... \left[\int_{-\infty}^{v_{ij}+e-v_{ij'}}\right] ... \int_{-\infty}^{v_{ij}+e-v_{iJ}} f(z_0, ..., e, ..., v_{ij}+e-v_{ij'}, ..., z_J)\mathrm{d}z_J ...[\mathrm{d}z'_j]...[\mathrm{d}z_j]...\mathrm{d}z_0 \mathrm{d}e \ \mathrm{d}F(\beta_i) \tag{A.2}$$

Using the change of variables $t = v_{ij} + e - v_{ij'}$, one obtains:

$$\begin{aligned}\frac{\partial s^{\star}_{ij}}{\partial x^{1}_{ij'}} &= \int -\beta_i \int \int_{-\infty}^{v_{ij'}+t-v_{i0}} ... \left[\int_{-\infty}^{v_{ij'}+t-v_{ij'}}\right] ... \left[\int_{-\infty}^{v_{ij'}+t-v_{ij}}\right] ... \int_{-\infty}^{v_{ij'}+t-v_{iJ}} \\ &\quad f(z_0, ..., v_{ij'}+t-v_{ij}, ..., t, ..., z_J)\mathrm{d}z_J ...[\mathrm{d}z'_j]...[\mathrm{d}z_j]...\mathrm{d}z_0 \mathrm{d}t \ \mathrm{d}F(\beta_i) \\ &= \frac{\partial s^{\star}_{ij'}}{\partial x^{1}_{ij}}\end{aligned} \tag{A.3}$$

Corollary 2. *Absence of Nominal Illusion*:

$$s^{\star}_{ij}(x^1_{ij}, x^2_{ij}) = Pr\left(v_{ij} + \epsilon_{ij} = \max_{j' \in \{0,...,J\}} v_{ij'} + \epsilon_{ij'}\right) \tag{A.4}$$

$$= Pr\left(\beta_i x^1_{ij} + w_{ij}(x^2_{ij}) + \epsilon_{ij} = \max_{j' \in \{0,...,J\}} \beta_i x^1_{ij'} + w_{ij}(x^2_{ij'}) + \epsilon_{ij'}\right) \tag{A.5}$$

$$= Pr\left(\beta_i x^1_{ij} + \beta_i\delta + w_{ij}(x^2_{ij}) + \epsilon_{ij} = \max_{j' \in \{0,...,J\}} \beta_i x^1_{ij'} + \beta_i\delta + w_{ij}(x^2_{ij'}) + \epsilon_{ij'}\right) \tag{A.6}$$

$$= Pr\left(\beta_i \left(x^1_{ij} + \delta\right) + w_{ij}(x^2_{ij}) + \epsilon_{ij} = \max_{j' \in \{0,...,J\}} \beta_i \left(x^1_{ij'} + \delta\right) + w_{ij}(x^2_{ij'}) + \epsilon_{ij'}\right) \tag{A.7}$$

$$= s^{\star}_{ij}(x^1_{ij} + \delta, x^2_{ij}) \tag{A.8}$$

Proof of Lemma 1. With a slight abuse of notation, let the set of consideration sets containing good j and j' be given as:

$$\mathbb{P}(j, j') = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad j' \in c \quad \& \quad 0 \in c\}, \tag{A.9}$$

Given symmetry of choice probabilities conditional on goods belonging to the same consideration set, the magnitude of cross derivative asymmetries depends on how market shares change with the variation in consideration set probabilities generated by variation in characteristics.

$$\frac{\partial s_{ij}}{\partial x^1_{ij'}} - \frac{\partial s_{ij'}}{\partial x^1_{ij}} = \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}}{\partial x^1_{ij'}} s^\star_{ij}(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}}{\partial x^1_{ij}} s^\star_{ij'}(c') + \sum_{c'' \in \mathbb{P}(j,j')} \pi_{ic''} \left(\frac{\partial s^\star_{ij}(c'')}{\partial x^1_{ij'}} - \frac{\partial s^\star_{ij'}(c'')}{\partial x^1_{ij}} \right) \tag{A.10}$$

$$= \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}}{\partial x^1_{ij'}} s^\star_{ij}(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}}{\partial x^1_{ij}} s^\star_{ij'}(c') \tag{A.11}$$

Thus, non-zero cross-derivative asymmetries imply:

$$\sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}}{\partial x^1_{ij'}} s^\star_{ij}(c) \neq \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}}{\partial x^1_{ij}} s^\star_{ij'}(c') \tag{A.12}$$

$$\text{either} \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}}{\partial x^1_{ij'}} s^\star_{ij}(c) \neq 0 \quad \text{and/or} \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}}{\partial x^1_{ij}} s^\star_{ij'}(c') \neq 0 \tag{A.13}$$

Given Assumption 5, this is only possible when $\pi(\mathcal{J}) < 1$.

Similarly, while level shifts in the quasi-linear characteristic do not cause choice probabilities conditional on a given consideration set to change, they do alter consideration set probabilities. Thus, absence of nominal illusion is violated. For $\delta \neq 0$,

$$s_{ij}(x^1_i + \delta, x^2_i) = \sum_{c \in \mathbb{P}(j)} \pi_{ic}(x^1_i + \delta, x^2_i) Pr\left(v_{ij} + \epsilon_{ij} = \max_{j' \in c} v_{ij'} + \epsilon_{ij'} \right) \tag{A.14}$$

If $s_{ij}(x^1_i + \delta, x^2_i) \neq s_{ij}(x^1_i + \delta, x^2_i)$, this implies that for at least one consideration set c

$$\pi_{ic}(x^1_i + \delta, x^2_i) \neq \pi_{ic}(x^1_i, x^2_i), \tag{A.15}$$

which given Assumption 5, is only possible with $\pi_{ic}(x^1_i, x^2_i) < 1$ or $\pi_{ic}(x^1_i, x^2_i) \neq 0$ and thus with $\pi(\mathcal{J}) < 1$.

Assumption 6. (Rank Condition) The matrix $D_i' D_i$ is full rank.

For the rank condition to hold, we must have that the number of independent cross-derivative

differences is at least as large as the number of derivatives of the log of consideration probabilities:

$$\frac{1}{2}J(J+1) \geq J+1 \tag{A.16}$$

$$J \geq 2 \tag{A.17}$$

Thus there must be at least two non-default goods plus the default. Further, all columns of D_i must be linearly independent. Sufficient conditions for this are:

$$s_{ij}(\mathcal{J}) \neq s_{ij'}(\mathcal{J}) \tag{A.18}$$

$$\frac{s_{il}(\mathcal{J}) - s_{il}(\mathcal{J}/j)}{s_{ij'}(\mathcal{J}) - s_{ij'}(\mathcal{J}/j)} \neq \frac{s_{il}(\mathcal{J}) - s_{il}(\mathcal{J}/j')}{s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')} \tag{A.19}$$

$$s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j') \neq 0 \tag{A.20}$$

for all $j, j', l \in \mathcal{J}$ with $j, j' > 0$. Equation A.20 will be met when good j' is considered with strictly positive probability and good j' is purchased with strictly positive probability from some choice set that includes j. Equation A.19 will be satisfied whenever goods are imperfect substitutes and/or are considered to different degrees. A strength of our approach is that the rank condition is testable given market share data.

To see the logic of these conditions, consider the just identified case where $J = 2$. In this example, the linear system defining the derivative of log consideration probabilities takes the form:

$$\begin{bmatrix} -(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1)) & 0 & s_{i1}(\mathcal{J}) \\ 0 & -(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2)) & s_{i2}(\mathcal{J}) \\ -(s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)) & (s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \log(\phi_{i1})}{\partial x_{i1}} \\ \frac{\partial \log(\phi_{i2})}{\partial x_{i2}} \\ \frac{\partial \log(\mu_{i0})}{\partial x_{i0}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_{i1}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{i1}} \\ \frac{\partial s_{i2}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{i2}} \\ \frac{\partial s_{i1}}{\partial x^1_{i2}} - \frac{\partial s_{i2}}{\partial x^1_{i1}} \end{bmatrix} \tag{A.21}$$

The determinant of D_i is:

$$\det(D_i) = s_{i2}(\mathcal{J})\left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1)\right)\left(s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)\right) - s_{i1}(\mathcal{J})\left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2)\right)\left(s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)\right) \tag{A.22}$$

When D_i is singular:

$$\frac{1}{s_{i1}(\mathcal{J})}\left(\frac{s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1)}{s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)}\right) = \frac{1}{s_{i2}(\mathcal{J})}\left(\frac{s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2)}{s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)}\right) \tag{A.23}$$

Assumption 8. (Rank Condition) $\Pi(\delta)$ is full rank.

Sufficient conditions for $\Pi(\delta)$ to be full rank are:

$$\frac{\phi_{ij}(x_{ij}^1+\delta_i)}{1-\phi_{ij}(x_{ij}^1+\delta_i)} \neq \frac{\phi_{ij}(x_{ij}^1+\delta_{i'})}{1-\phi_{ij}(x_{ij}^1+\delta_{i'})} \tag{A.24}$$

$$\phi_{ij}(x_{ij}^1+\delta_i) \neq \phi_{ij'}(x_{ij'}^1+\delta_i) \quad \text{at, at least one } i=1,...,N \tag{A.25}$$

for $j, j' > 0$.

To see the logic of these conditions, consider the just identified case where $J=2$ and $N=2$. The coefficient matrix then takes the form:

$$\Pi(\delta) = \begin{bmatrix} \mu_{i0}(\delta_1)\phi_{i1}(\delta_1)(1-\phi_{i2}(\delta_1)) & \mu_{i0}(\delta_1)\phi_{i1}(\delta_1)\phi_{i2}(\delta_1) & 0 & 0 \\ \mu_{i0}(\delta_2)\phi_{i1}(\delta_2)(1-\phi_{i2}(\delta_2)) & \mu_{i0}(\delta_2)\phi_{i1}(\delta_2)\phi_{i2}(\delta_2) & 0 & 0 \\ 0 & 0 & \mu_{i0}(\delta_1)\phi_{i2}(\delta_1)(1-\phi_{i1}(\delta_1)) & \mu_{i0}(\delta_1)\phi_{i1}(\delta_1)\phi_{i2}(\delta_1) \\ 0 & 0 & \mu_{i0}(\delta_2)\phi_{i2}(\delta_2)(1-\phi_{i1}(\delta_2)) & \mu_{i0}(\delta_2)\phi_{i1}(\delta_2)\phi_{i2}(\delta_2) \end{bmatrix} \tag{A.26}$$

$$= \begin{bmatrix} \Pi_1(\delta) & 0 \\ 0 & \Pi_2(\delta) \end{bmatrix} \tag{A.27}$$

The determinant of $\Pi(\delta)$ takes the form:

$$\det(\Pi(\delta)) = \det(\Pi_1(\delta))\det(\Pi_2(\delta)) \tag{A.28}$$

Simple arithmetic shows that $\Pi_1(\delta)$ is singular when:

$$\frac{1-\phi_{i2}(\delta_1)}{\phi_{i2}(\delta_1)} = \frac{1-\phi_{i2}(\delta_2)}{\phi_{i2}(\delta_2)} \tag{A.29}$$

Similarly, $\Pi_2(\delta)$ is singular when:

$$\frac{1-\phi_{i1}(\delta_1)}{\phi_{i1}(\delta_1)} = \frac{1-\phi_{i1}(\delta_2)}{\phi_{i1}(\delta_2)} \tag{A.30}$$

When $J > 2$, we require that $\phi_{ij}(x_{ij}^1+\delta_i) \neq \phi_{ij'}(x_{ij'}^1+\delta_i)$ at, at least one shift of the quasilinear characteristic to prevent columns of $\Pi_j(\delta)$ being perfectly collinear.

A.2 ASC Identification with Dependence on Default Characteristics

A version of the ASC model in which the probability of considering non-default goods depends on both own and default characteristics is also identified given our background assumptions. Let the probability of considering the default be one, with market shares taking the form:

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il}(x_{i0}, x_{il}) \prod_{l' \notin c} (1 - \phi_{il'}(x_{i0}, x_{il})) \tag{A.31}$$

with $\phi_{i0} = 1$ and $\mathbb{P}(j) = \{c : c \in \mathcal{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad 0 \in c\}$.

Changes in the characteristics of the default good alter all consideration probabilities. Cross derivative differences involving $j = 0$ are given by the linear system:

$$\frac{\partial s_{ij}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{ij}} = \frac{\partial \log(\phi_{ij})}{\partial x^1_{i0}} s_{ij}(\mathcal{J}) + \sum_{j' \neq \{j,0\}} \frac{\partial \log(\phi_{ij'})}{\partial x^1_{i0}} \left(s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')\right) - \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} \left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)\right) \tag{A.32}$$

Thus there are now $2J$ derivatives of log consideration probabilities to identify: $\partial \log(\phi_{ij})/\partial x^1_{ij}$ and $\partial \log(\phi_{ij})/\partial x^1_{i0}$ for $j > 0$.

The conditions for the rank condition for identification of the derivatives of log consideration probabilities are now altered. We require a larger number of goods to attain sufficient cross derivatives for the order condition to hold (Assumption 6):

$$\frac{1}{2} J(J+1) \geq 2J \tag{A.33}$$

$$J \geq 3 \tag{A.34}$$

In this model, we cannot allow $\phi_{i0}(x_{i0}) \leq 1$ and the rank condition still hold. This is because we will only ever have J independent cross derivatives involving the default good but there will be $J+1$ changes in consideration probabilities with respect to the default good to identify. Other than this restriction, the rest of the proof in Section 3 goes through without modification.

A.3 ASC Identification with an 'Outside' Default Good

When interest is in the ASC model with an outside default that is always considered, one cannot make use of cross derivatives which rely on variation in characteristics of the default good. In this case, the order condition for the identification of the derivative of log consideration probabilities

changes (Assumption 6). We now require:

$$\frac{1}{2}J(J-1) \geq J \tag{A.35}$$

$$J \geq 3 \tag{A.36}$$

All cross derivative differences take the form given by Equation and the rest of the identification proof continues as in Section 3.

A general version of the ASC model defines market shares as:

$$s_{i0} = \prod_{j \in \mathcal{J}} (1 - \phi_{ij}(x_{ij})) + \sum_{c \in \mathbb{P}(0)} \prod_{l \in c} \phi_{il}(x_{il}) \prod_{l' \notin c} (1 - \phi_{il'}(x_{il'}))\, s^{\star}_{i0}(c) \tag{A.37}$$

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il}(x_{il}) \prod_{l' \notin c} (1 - \phi_{il'}(x_{il'}))\, s^{\star}_{ij}(c) \tag{A.38}$$

where $\phi_{i0} = 1$ and $\mathbb{P}(j) = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad 0 \in c\}$. This framework allows for a default good, good-0, that is imperfectly considered but purchased if no goods are considered. For example, if a consumer fails to consider any health insurance or pension plans, they may be auto-enrolled onto a default option.

A.4 ASC Identification with an Inside Default Good with $\phi_{i0} < 1$

In some scenarios, it might be natural to allow for an inside good that is not always considered but is defaulted to if the choice set is empty. For example, if a consumer doesn't consider any health insurance or pension plans, they may be auto-enrolled into some option.

In this case, choice probabilities take the following form:

$$s_{i0} = \prod_{j \in \mathcal{J}} (1 - \phi_{ij}) + \sum_{c \in \mathbb{P}(0)} \prod_{l \in c} \phi_{il} \prod_{l \notin c} (1 - \phi_{il'}) s^{\star}_{i0}(c) \tag{A.39}$$

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il} \prod_{l \notin c} (1 - \phi_{il'}) s^{\star}_{ij}(c) \tag{A.40}$$

The structure of cross derivative differences is as the standard case for $j, j' > 0$. However, for cross-derivative differences involving the default:

$$\frac{\partial s_{i0}}{\partial x^1_{ij}} - \frac{\partial s_{ij}}{\partial x^1_{i0}} = \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} \left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)\right) - \frac{\partial \log(\phi_{i0})}{\partial x^1_{i0}} \left(s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/0)\right) \tag{A.41}$$

This expression might seem somewhat odd given that 'leave-zero-out' variation is required. How natural this assumption is might vary across contexts. If default goods are randomly assigned in

the population, this variation (or permitting the market share of good-0 to go to zero) might be plausible. If Assumption 7a holds, then the proof of identification follows as in Section 3 with the above modification to cross derivative differences involving the default.

A.5 Relationship between ASC & DSC Models and Search Models

The DSC Model & Sequential Seach Models Ho, Hogan, and Scott Morton (2015) and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) provide microfoundations for the DSC model. However, we here note that our identification results can be easily adapted to a wider class of sequential search models. Rather than searching only if the default good is unattractive, a consumer searches the next best option if their top ranked option is unavailable. For example, the optimal policy of Weitzman (1979) involves searching options in order of their reservation value, stopping when an option is available. We here develop identification results for this model and show how these relate to those of the DSC model.

Using the notation of Chade and Smith (2006), order the $J+1$ options according to their (unobserved) expected utility or reservation value. The probability that an option j is available is independent across options and given by $\alpha_{ij}(x_{ij})$. Thus the probability of consumer i choosing options ranked first through to J is:

$$s_{i1} = \alpha_{i1} \tag{A.42}$$

$$s_{i2} = \alpha_{i2}(1-\alpha_{i1}) \tag{A.43}$$

$$\vdots \tag{A.44}$$

$$s_{iJ} = \alpha_{iJ}\prod_{j'=1}^{J-1}(1-\alpha_{ij'}) \tag{A.45}$$

Let the default be chosen only if all other options are unavailable, i.e. $s_{i0} = \prod_{j=1}^{J}(1-\alpha_{ij})$.

The ranking of options is unobserved by the econometrician. The aim is to recover this latent ranking and the probability of option availability. The cross-derivative of choice probabilities takes the form:

$$\frac{\partial s_{ij}}{\partial x_{ij'}} = -1\left(j' > j\right)\left[\frac{\partial \alpha_{ij'}}{\partial x_{ij'}}\alpha_{ij}\prod_{l\neq j'<j}(1-\alpha_{il})\right] \tag{A.46}$$

and 0 if $j' > j$. The pattern of zero cross-derivative effects thus identifies the underlying ranking of

options. For $j' < j$, we have:

$$\frac{\partial s_{ij}}{\partial x_{ij'}} = -\frac{\partial \alpha_{ij'}}{\partial x_{ij'}} \alpha_{ij} \prod_{l \neq j' < j} (1 - \alpha_{il}) \tag{A.47}$$

$$= \frac{\partial \log\left(1 - \alpha_{ij'}\right)}{\partial x_{ij'}} s_{ij} \tag{A.48}$$

Therefore, derivatives of the log of availability probabilities are identified in an analogous manner to the DSC model:

$$\frac{\partial \log\left(1 - \alpha_{ij'}\right)}{\partial x_{ij'}} = \frac{1}{s_{ij}} \left[\frac{\partial s_{ij}}{\partial x_{ij'}} - \frac{\partial s_{ij'}}{\partial x_{ij}} \right] \tag{A.49}$$

Indeed, if many cross-derivatives are zero and there exist asymmetries with respect to cross-derivative differences that do not involve the default good, this suggests that one should perhaps consider whether this wider class of models better represents the choice problem to hand than the DSC model.

The ASC Model & Search Models While the ASC model places strong restrictions on the functional form of consideration set probabilities, it can nonetheless be a good approximation to a class of optimising models of consumer search.

As described by Honka (2014), when there is first order stochastic dominance among the good-specific distributions over which a consumer is searching, consideration sets are defined by a simple cut-off rule (see Chade and Smith (2006)). Similarly, Caplin, Dean, and Leahy (2016) show that the rational inattention model of Sims (2003) implies the formation of consideration sets defined by cut-off strategies given restrictions on the correlation between the ex ante valuation of alternatives. The precise cut-off point between the set of considered and unconsidered goods is endogenous and will generally depend on the characteristics of all goods and search costs. For example, in Honka (2014), assuming homogeneous constant search costs across goods, the cut-off rule defining a consideration set c given a particular realisation of expected utility shocks is:

$$\min_{j \in c} \widetilde{u}_{ij}(x_{ij}, \widetilde{\epsilon}_{ij}) \geq t(x_i, \widetilde{\epsilon}_i) \geq \max_{j' \notin c} \widetilde{u}_{ij'}(x_{ij'}, \widetilde{\epsilon}_{ij'}) \tag{A.50}$$

where $\widetilde{u}_{ij}(x_{ij}, \widetilde{\epsilon}_{ij})$ gives the expected utility of good j given random shock $\widetilde{\epsilon}_{ij}$ and $t(x_i)$ gives the endogenous cut-off rule. The probability that a good is considered once independent expected utility

errors are taken into account can then be expressed as:

$$\phi_{ij}(x_i) = \Pr\left(\widetilde{u}_{ij}(x_{ij}, \widetilde{\epsilon}_{ij}) \geq t(x_i, \widetilde{\epsilon}_i)\right) \tag{A.51}$$

This model deviates from the ASC model as the probability of considering good j depends on the characteristics of all goods through the cut-off rule: $\partial\phi_{ij}/\partial x_{ij'} \neq 0$.[34] However, the ASC model will nonetheless be a good approximation to this model when:

$$\frac{\partial t}{\partial x_{ij'}} \approx 0 \quad \text{as this implies} \quad \frac{\partial \phi_{ij}}{\partial x_{ij'}} \approx 0 \tag{A.52}$$

for $j' \neq j$ and $j \neq \arg\min_{j \in c} \widetilde{u}_{ij}(x_{ij}, \widetilde{\epsilon}_{ij})$ and $j \neq \arg\max_{j \notin c} \widetilde{u}_{ij}(x_{ij}, \widetilde{\epsilon}_{ij})$.

However, even in scenarios where this assumption is suspect (e.g. the overidentification tests in Section 3 are violated), features of consideration set probabilities can still be identified from cross-derivative asymmetries. This remains the case even with correlation between the unobservables driving consideration probabilities, e.g. correlation in expected random utility errors, although, of course, we still require independence of unobservables driving attention and utility. To illustrate, let $g_{ij} = x_{ij}\gamma$ and assume that the impact of characteristics on attention probabilities comes via the indices g_{ij}. The general expression for cross-derivative differences in consideration set models then takes the form:

$$\frac{\partial s_{ij}}{\partial x_{ij'k}} - \frac{\partial s_{ij'}}{\partial x_{ijk}} = \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}(g_{i0}, ..., g_{iJ})}{\partial x_{ij'}} s^{\star}_{ij}(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}(g_{i0}, ..., g_{iJ})}{\partial x_{ij}} s^{\star}_{ij'}(c') \tag{A.53}$$

$$= \gamma_k \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}(g_{i0}, ..., g_{iJ})}{\partial g_{ij'}} s^{\star}_{ij}(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}(g_{i0}, ..., g_{iJ})}{\partial g_{ij}} s^{\star}_{ij'}(c') \tag{A.54}$$

Thus, γ is identified up to a scale by relative differences in cross-derivative asymmetries.

$$\frac{\frac{\partial s_{ij}}{\partial x_{ij'k}} - \frac{\partial s_{ij'}}{\partial x_{ijk}}}{\frac{\partial s_{ij}}{\partial x_{ij'k}} - \frac{\partial s_{ij'}}{\partial x_{ijk}}} = \frac{\gamma_k}{\gamma_{k'}} \tag{A.55}$$

While further structure is required to point identify all structural functions of interest, cross-derivative differences nonetheless remain a source of identifying power in much more complicated frameworks than those considered in the main text of this paper, for example, those that permit dependence between the probability of considering good j and of considering good j', or dependence between the probability of considering good j and the characteristics of good j'.

[34] This being said we point the reader to Section A.2 for a generalisation of the ASC model that allows for dependence of own-good consideration upon the default characteristics.

B Proof of Utility Representations for Consideration Set Models

Consider first the ASC model. We start by assuming there is a default plan to which you are always attentive (plan 0) and an alternative, plan 1, to which you might be inattentive. Let preferences be given by:

$$u_{ij} = \beta x_{ij}^1 + w_j(x_{ij}^2) + \epsilon_{ij} \tag{B.1}$$

$$= v_{ij} + \epsilon_{ij} \tag{B.2}$$

In this two-good ASC model, we can write the probability of choosing good 1 as:

$$s_{i1} \quad = \quad \phi_{i1} s_{i1}^\star \tag{B.3}$$

where $s_{i1}^\star$ is the probability of choosing good 1 conditional on paying attention and $x_{ij} = [x_{ij}^1, x_{ij}^2]$.

With i.i.d. extreme value errors, this model is equivalent to a full-consideration model with preferences specified as:

$$\widetilde{u}_{ij} = v_{ij} + \psi_{i,j=1} + \epsilon_{ij} \tag{B.4}$$

where $\psi_{i,j=1} = \psi_{i1}$ for plan 1 and is 0 otherwise, where ψ_{i1} is given by:

$$\psi_{i1} = \ln\left(\frac{\phi_{i1}(x_{i1})\exp(v_{i0})}{(1-\phi_{i1}(x_{i1}))\exp(v_{i1}) + \exp(v_{i0})}\right) \tag{B.5}$$

This follows since:

$$s_{i1} = \frac{\exp(v_{i1}+\psi_{i,j=1})}{\exp(v_{i1}+\psi_{i,j=1}) + \exp(v_{i0})} = \phi_{i1}(x_{i1})\frac{\exp(v_{i1})}{\exp(v_{i1}) + \exp(v_{i0})} \tag{B.6}$$

We prove that an analogous result holds in a J good model by the inductive hypothesis with $\psi_{i,j=d} = 0$ for the default plan and $\psi_{i,j\neq d}$ otherwise implicitly defined by the system of $J-1$ equations:

$$\psi_{i,j} = \ln\left(\frac{\phi_{ij}\sum_{k\neq j}\exp(v_{ik}+\psi_{ik})}{(1-\phi_{ij})\exp(v_{ij}) + \sum_{k\neq j}\exp(v_{ik}+\psi_{ik})}\right) \tag{B.7}$$

We showed above that this holds for the case where $J = 2$. Let s_{ij}^a denote the probability of choosing good j conditional on paying attention to good j, i.e. $s_{ij}^a = s_{ij}(x_i|\phi_{ij} = 1)$. In the two-good case,

$s^a_{i1} = s^\star_{i1}$ but more generally:

$$s^a_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c, l \neq j} \phi_{il} \prod_{l' \notin c} (1 - \phi_{il'})\, s^\star_{ij}(c) \tag{B.8}$$

Thus, consider adding a Jth plan to which you might be inattentive:

$$s_{iJ} \quad = \quad \phi_{iJ} s^a_{iJ} \tag{B.9}$$

By the inductive hypothesis, we have:

$$s^a_{iJ} = \frac{\exp(v_{iJ})}{\exp(v_{iJ}) + \sum_{k \neq J} \exp(v_{ik} + \psi_{ik})} \tag{B.10}$$

Therefore,

$$s_{iJ} \quad = \quad \phi_{iJ} \frac{\exp(v_{iJ})}{\exp(v_{iJ}) + \sum_{k \neq J} \exp(v_{ik} + \psi_{ik})} \tag{B.11}$$

It is straightforward to confirm that these choice probabilities result from full-consideration utility maximization given that the Jth good has utility given by:

$$u_{iJ} = v_{iJ} + \psi_{iJ} + \epsilon_{iJ} \tag{B.12}$$

where:

$$\psi_{ij} = \ln \left(\frac{\phi_{ij} \sum_{k \neq j} \exp(v_{ik} + \psi_{ik})}{(1 - \phi_{ij}) \exp(v_{ij}) + \sum_{k \neq j} \exp(v_{ik} + \psi_{ik})} \right) \tag{B.13}$$

Thus, if this representation holds for a choice set with $J - 1$ plans, it holds for a choice set with J plans, and the proof is complete for the ASC model.

Next, consider the DSC model.

$$\begin{aligned} s_{id} &= (1 - \mu_{i0}) + \mu_{i0} s^\star_{id} \\ s_{ij} &= \mu_{i0} s^\star_{ij} \text{ for } j \neq d \end{aligned} \tag{B.14}$$

where $s^\star_{ij} = 1$ are the choice probabilities which result from maximizing:

$$u_{ij} = v_{ij} + \xi_{i,j=d} + \epsilon_{ij} \tag{B.15}$$

We want to show that this is equivalent to a full-consideration model where choice probabilities are

given by:

$$u_{ij} = v_{ij} + \xi_{i,j=d} + \psi_{i,j=d} + \epsilon_{ij} \tag{B.16}$$

Let $\psi_{i,j=d} = \psi_i$ and zero when $j \neq d$ and $\xi_{i,j=d} = \xi_i$ and zero when $j \neq d$. The full-consideration model will be equivalent to the DSC model with:

$$\psi_i = \ln\left(\frac{1 + (1-\mu_{i0})\sum_{k\neq d}\exp(v_{ik} - v_{id} - \xi_i)}{\mu_{i0}}\right) \tag{B.17}$$

C Estimation

Goeree (2008) provides details of the estimation process for the ASC model. We sketch the main ideas here. With a small number of available alternatives, estimation is straightforward. The probability of choosing any specific alternative as a function of the parameters $\theta = (\beta, \gamma)$ is given by:

$$s_{ij}(\theta) = \sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_{il}(\theta) \prod_{l'\notin c} (1 - \phi_{il'}(\theta)) s^{\star}_{ij}(c, \theta) \tag{C.1}$$

We can use this to construct the likelihood function and then estimate the parameters β and γ by maximum likelihood.

In larger choice sets, a major computational issue arises - there are 2^J possible consideration sets to sum over. To deal with this problem, we follow Goeree (2008) in using a simulated likelihood approach. The basic idea is to estimate the term $\sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_{il}(\theta) \prod_{l'\notin c}(1 - \phi_{il'}(\theta)) s^{\star}_{ij}(c, \theta)$ by simulating R consideration sets per individual where, for each $r = 1, ..., R$, each option is added to the consideration set with probability ϕ_{ij} so that the probability a given consideration set is simulated is given by: $\prod_{l\in c} \phi_{il} \prod_{k\notin c}(1 - \phi_{ik})$. We then compute:

$$\hat{s}_{ij} = \frac{1}{R} \sum_{r} s^{\star}_{ij}(c_r, \theta) \tag{C.2}$$

Since each c_r is chosen with probability $\prod_{l\in c} \phi_{il} \prod_{l'\notin c}(1 - \phi_{il'})$, we have that:

$$\begin{aligned} \hat{s}_{ij} &= \frac{1}{R} \sum_{r=1}^{R} \prod_{l\in c_r} \phi_{il}(\theta) \prod_{li\notin c_r} (1 - \phi_{il'}(\theta)) s^{\star}_{ij}(c, \theta) \\ &\to_p \sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_{il}(\theta) \prod_{l'\notin c} (1 - \phi_{il'}(\theta)) s^{\star}_{ij}(c, \theta) \end{aligned} \tag{C.3}$$

However, this procedure is still computationally burdensome as it requires computing $s^{\star}_{ij}(c_r, \theta)$

for every simulation r for all individuals at each candidate set of parameter values (since as the underlying parameters shift, the ϕ, and thus the choice sets would shift).

Following Goeree (2008), two additional tricks are used so that the choice probabilities need to be evaluated only once per person for each simulation r. First, we use the same uniform draws to simulate choice sets at each set of parameter values. Second, we use an importance sampler so that the choice probabilities need only be evaluated at the consideration sets implied by the parameters at their initial values. Specifically, we can compute Equation C.2 using:

$$\hat{s}_{ij} = \frac{1}{R} \sum_{r} \prod_{l \in c_r} \phi_{il} \prod_{l \notin c_r} (1 - \phi_{il'}) \frac{s^{\star}_{ij}(c_0, \theta)}{\phi^0_{ir}(\theta_0)} \tag{C.4}$$

where $\phi^0_{ir}(\theta_0) = \prod_{l \in c_0} \phi_{il} \prod_{k \notin c_0} (1 - \phi_{ik})$ and each consideration set is sampled with probability $\phi^0_{ir}(\theta_0)$.[35]

D Additional Tables and Figures for Empirical Work

This Section provides additional tables and figures that illustrate the empirical results of Sections 4 and 5.

Lab Experiment Table 7 shows the products used in the experiment and their list prices. A sample product selection screen is shown in Figure D.1.

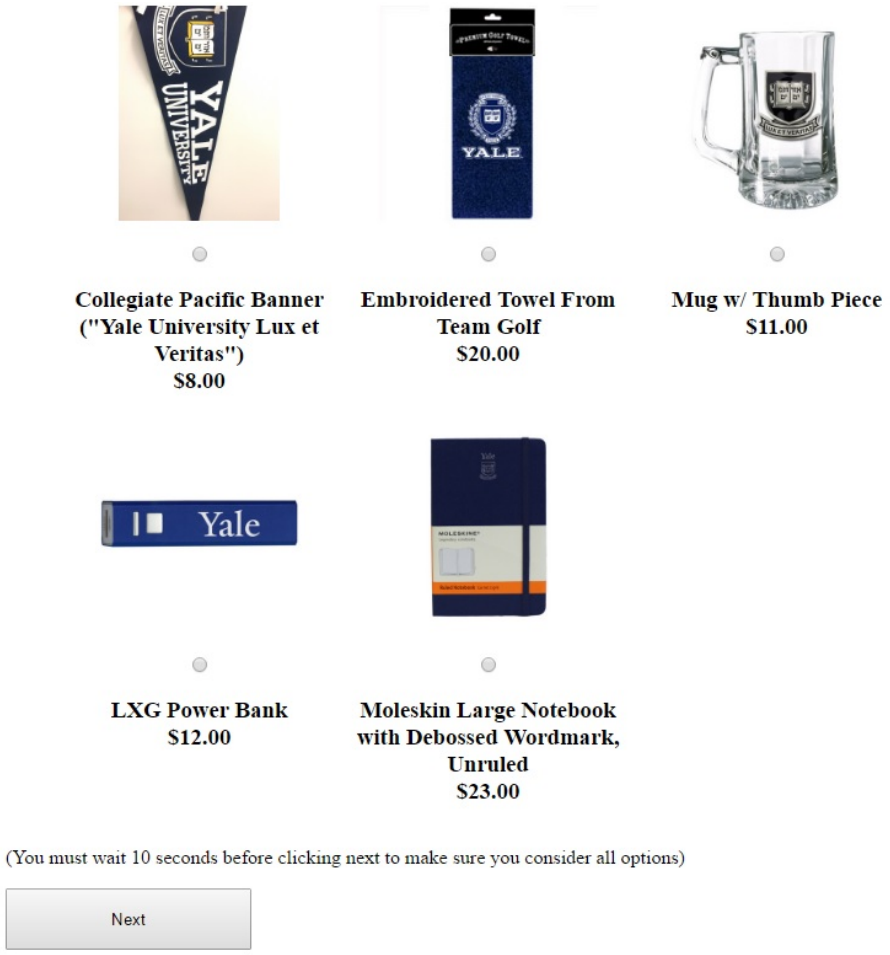
Table 7: Product Names and Prices

Product Name	**List Price** ($)
Yale Bulldogs Carolina Sewn Large Canvas Tote	22.98
10 Inch Custom Mascot	24.98
Alta Ceramic Tumbler	22.98
Yale Insulated Gemini Bottle	22.98
Yale Bulldogs Legacy Fitted Twill Hat	24.98
Moleskin Large Notebook with Debossed Wordmark, Unruled	25.00
Collegiate Pacific Banner ("Yale University Lux et Veritas")	24.98
Embroidered Towel From Team Golf	19.98
Mug w/ Thumb Piece	24.98
LXG Power Bank (USB Stick)	24.98

Notes: Table shows items used in experiment & their list prices.

[35] An importance sampler estimates a density $f(x)$ by drawing from a density $g(x)$, labeling the resulting value as x_1 and then weighting each draw by $f(x_1)/g(x_1)$. The resulting density is equivalent to drawing directly from $f(x)$.

Figure D.1: Lab Experiment: Sample Product Selection Screen



Expedia Analysis Table 8 gives summary statistics for the sample used in the Expedia analysis. As explained in the main text, given the estimated coefficients, we compute the estimated cross-derivatives with respect to the position variable. These can be thought of as the impact of increasing position for a given hotel and then using the model to "undo" the impact on demand of the resulting position changes for rival hotels. For each individual and each pair of hotels, we can compute the magnitude of the asymmetry as a percentage of the average absolute cross-effect in the data. Figure D.2 graphs the resulting asymmetries.

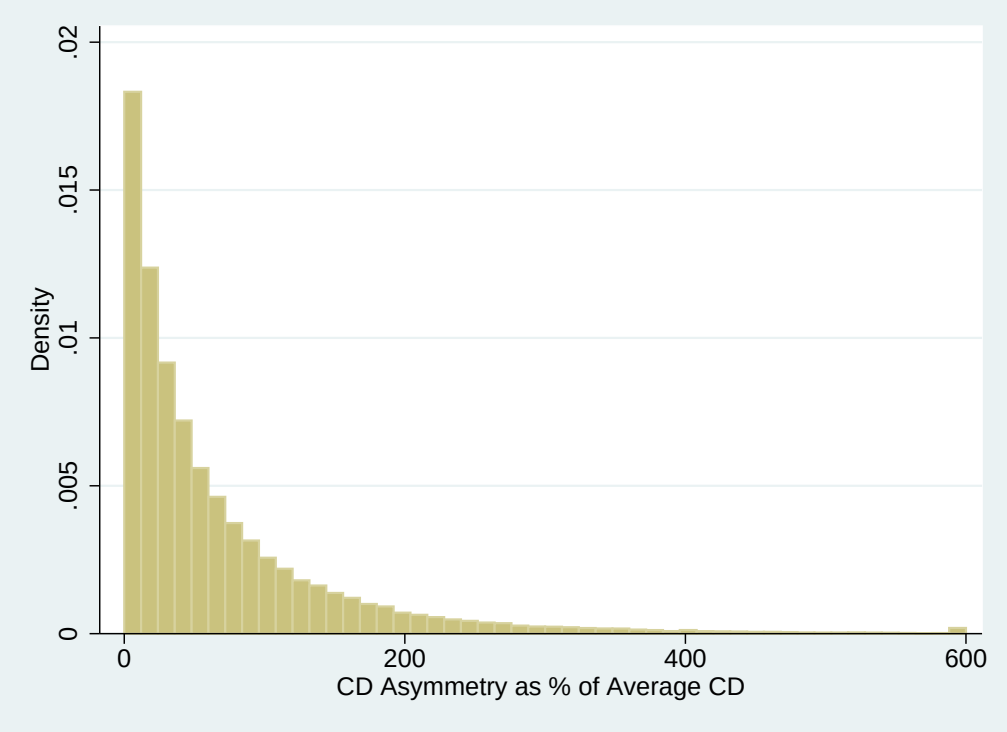
For our validation exercise we estimate the model using only the hotels in search positions 3 through 10. We then compute demand for each hotel as if the constant in the attention equation becomes arbitrarily large so that the attentive probability goes to 1 – this is our bound for that hotel. We then ask how well the bound does in accounting for the observed behavior in positions 1 and 2. While we cannot know ex ante how the attention probability will change if a hotel is placed in positions 1 or 2, we know that demand in those positions should be less than the bound given by perfect attention. Thus, we ask first whether the bound implied by the ASC model is indeed a

Table 8: Expedia Data: Summary Statistics

	All Hotels	Chosen Hotels
Price ($)	156	136
	(97.2)	(67.9)
Hotel Stars (1-5)	3.21	3.28
	(0.88)	(0.80)
Hotel Review Score (1-5)	3.93	3.99
	(0.72)	(0.61)
Popular Brand Indicator	0.74	0.74
	(0.44)	(0.44)
Location Score (normalized)	-0.12	-0.10
	(0.87)	(0.86)
Ongoing Promotion Indicator	0.20	0.27
	(0.40)	(0.45)
Position in Search	5.5	4.52
	(2.87)	(2.89)
Number of Observations	24,410	2,441

Notes: Table reports means and standard deviations (in parenthesis) for the sample of consumers who received a randomized hotel ordering in search and recorded a final transaction. Price is dollars per night, the popular brand indicates the hotel is part of a "major hotel chain" (as defined by Expedia), and the online promotion indicator indicates that the hotel is highlighted because the listed price is lower than is typical for that hotel.

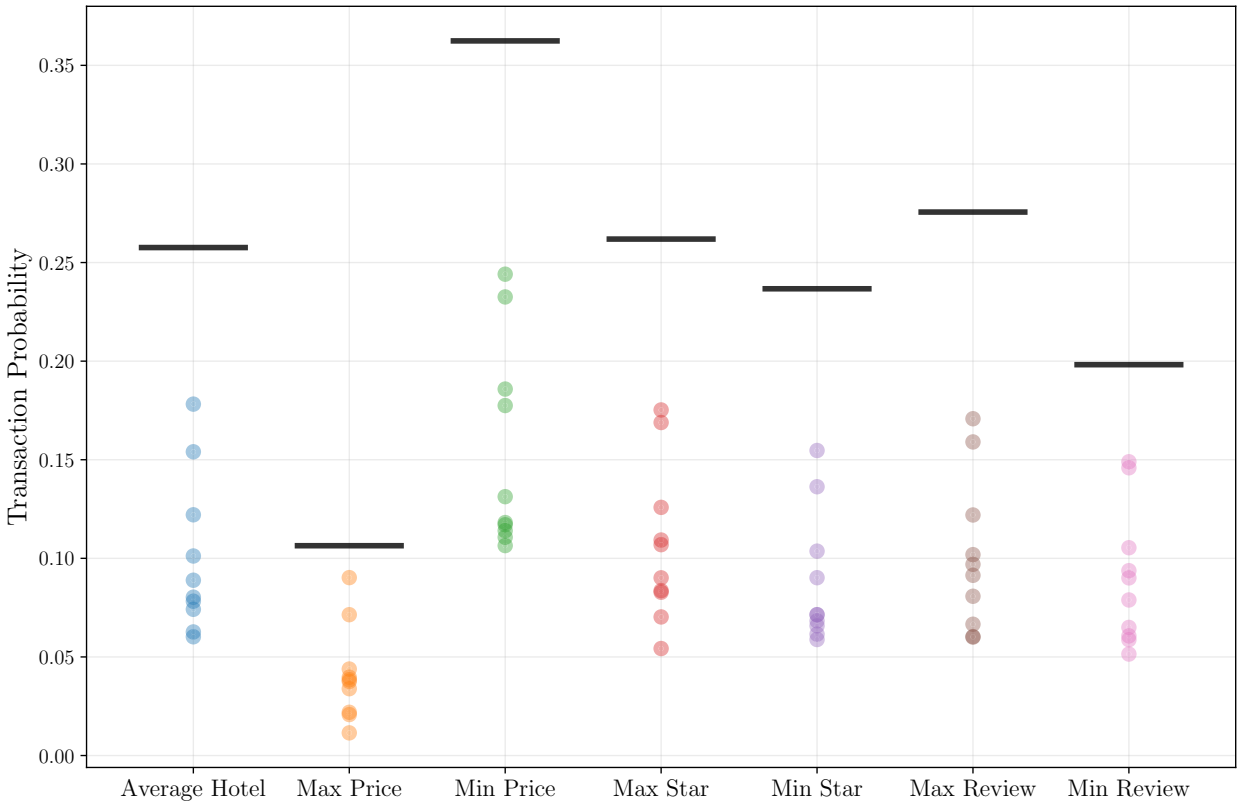
Figure D.2: Estimated Asymmetries in Position Responses



bound on choice probabilities for hotels in positions 1 and 2 and second, whether this bound has predictive power in accounting for the choice probabilities conditional on observed demand.

In practice, we compute this bound separately for each hotel in the data, but we collapse down to categories of hotels for expository purposes. The upper bound on the effectiveness of informative advertising is given by transaction probabilities when the the probability of paying attention is one. Figure D.3 shows how this bound compares to the observed demand for a variety of different types of hotels in each search position. The thick horizontal line shows the bound, the 10 colored dots show demand in each search position (with higher dots corresponding to lower search positions). The main takeaways from this figure are first that demand is always less than the bound implied by perfect attention and second that the bound is non-trivial. For example, average demand for hotels in positions 1-3 exceeds the bound placed on the demand for the maximum price hotels.

Figure D.3: Expedia: Bound vs. Demand by Search Position



Part D Analysis Summary statistics from our data after all sample selection restrictions are imposed are reported in Table 9 in Appendix D. We report the mean and standard deviation of a variety of characteristics for all plans and also for chosen plans.

Figure D.4 gives the predicted cross derivative difference between default and non-default goods for included plan characteristics for four variables; the charts for all variables are in Appendix D.

Table 9: Part D Data: Summary Statistics

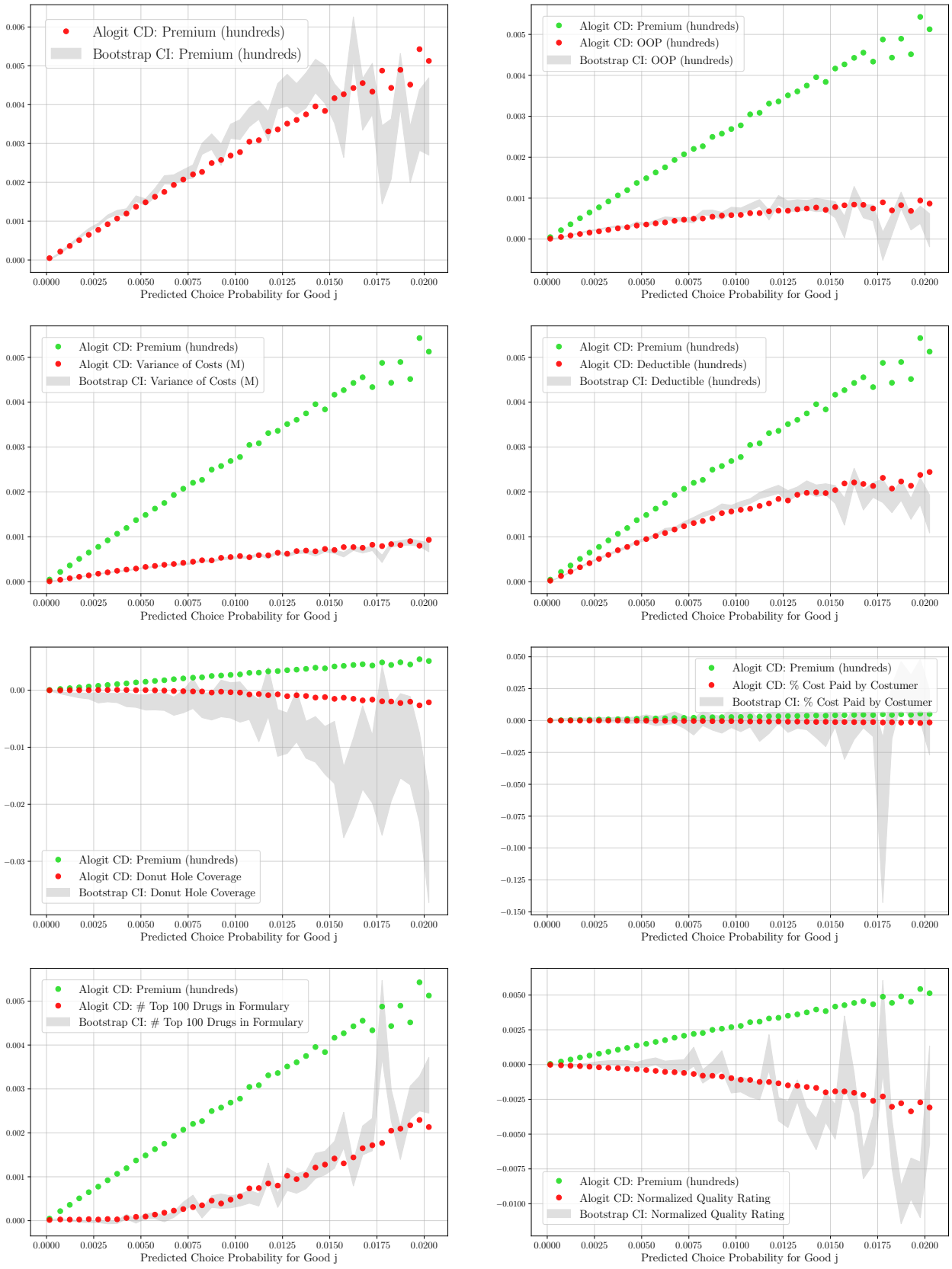
	All Plans	Chosen Plans
Annual Premium ($)	493	423
	(242)	(199)
Annual Out of Pocket Costs ($)	874	881
	(710)	(700)
Variance of Costs (millions)	0.618	0.615
	(0.525)	(0.519)
Deductible	65.3	62.3
	(113)	(114)
Full Donut Hole Coverage	0.003	0.005
	(0.055)	(0.067)
Generic Donut Hole Coverage	0.230	0.126
	(0.421)	(0.332)
% of Costs Paid by Consumer	0.377	0.391
	(0.101)	(0.099)
# of Top 100 Drugs in Formulary	99.4	99.7
	(1.61)	(0.964)
Normalized Quality Rating	0.081	0.434
	(0.952)	(1.216)
Number of (year, beneficiary, plans)	1,363,761	68,469
Number of Beneficiaries	30,937	30,937

Notes: Table reports means and standard deviations (in parenthesis) of each variable for the beneficiaries in our final sample. The sample consists of an observation for each (year, beneficiary, plan).

We graph both the estimated cross derivatives from Equation 5.12 and the cross-derivatives implied by the DSC model ($\gamma_k(1-\mu_{i0})\hat{s}_{ij}$) against the predicted market share of plan j, $\hat{s}_{ij}$. To capture the uncertainty in the estimated cross-derivatives, we bootstrap estimation of Equation 5.12 and graph the resulting confidence interval.

In all graphs, the green dots indicate the empirical cross-derivatives with respect to premiums – this is exactly the same data in all graphs, and is included for scale (the green dots are absent in the premium graph itself since they would overlap perfectly with the red dots). For each variable, the red dots indicate the predicted cross-derivative difference from the DSC model and the grey confidence region indicates the "empirical" cross-derivative difference from the more flexible specification in Equation 5.12. We can see that in nearly all cases, the DSC model cross-derivatives match up well with empirical cross-derivatives. There are a few exceptions – for example, there are some nonlinearities in the cross-derivatives with respect to the quality rating which are not well-accounted for by the underlying model of inattention. But overall, the patterns in the cross-derivatives are extremely well-explained by the relatively parsimonious model of inattention.

Figure D.4: Empirical vs. Model Predicted Cross-derivatives



E Practical Overidentification Test in the DSC Model

The DSC model models choice probabilities as:

$$\begin{aligned} s_{id} &= (1-\mu_i) + \mu_i s^*_{id} \\ s_{ij} &= \mu_i s^*_{ij} \end{aligned} \tag{E.1}$$

In our empirical applications, we focus on linear logit specifications. Thus,

$$\begin{aligned} \mu_i &= \frac{\exp(x_{id}\gamma)}{1+\exp(x_{id}\gamma)} \\ s^\star_{ij} &= \frac{\exp(x_{ij}\beta)}{\sum_{j'=0}^{J}\exp(x_{ij'}\beta)} \end{aligned} \tag{E.2}$$

As noted in Section 5, the DSC model with linear utility and logit errors can be written as a random utility model where utility depends on the characteristics of rival goods:

$$u_{ij} = x_{ij}\beta + \psi_{i,j=d} + \epsilon_{ij} \tag{E.3}$$

$$= v_{ij} + \epsilon_{ij} \tag{E.4}$$

where $\psi_{i,j=d}$ is the a term that reflects the impact of imperfect consideration that varies as a function of own and rival characteristics: where:

$$\begin{aligned} \psi_{j=d} &= \ln\left(\frac{1+(1-\mu_i)\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta)}{\mu_i}\right) \\ &= \ln\left(\frac{(1-\mu_i)+\mu_i s^*_{id}}{\mu_i s^*_{id}}\right) \end{aligned} \tag{E.5}$$

The test we propose is to first estimate the DSC model to recover $\hat{\beta}$ and $\hat{\gamma}$, and thus $\hat{\psi}_{ij}$, to form:

$$\hat{v}_{ij} = x_{ij}\hat{\beta} + \hat{\psi}_{i,j=d} \tag{E.6}$$

This 'first stage' yields predicted market shares $\hat{s}_{ij}$ and predicted latent shares, $\hat{s}^\star_{ij}$ (i.e. full consideration predictions). Predicted cross-derivative differences then take the form;

$$\widehat{\frac{\partial s_{ij}}{\partial x_{idk}} - \frac{\partial s_{id}}{\partial x_{ijk}}} = \hat{\gamma}_k(1-\hat{\mu}_i)\hat{s}_{ij} \tag{E.7}$$

To determine whether the DSC model is sufficiently flexible to be able to capture the patterns

in empirical cross-derivatives, we next estimate the following model with a rich set of interaction terms:

$$\widetilde{u}_{ij} \quad = \quad x_{ij}\hat{\beta} + \sum_{k}\sum_{k'} x_{idk}x_{ijk'}\alpha_{k,k'} + \epsilon_{ij} \text{ for } j \neq d \tag{E.8}$$

with predicted market shares $\widetilde{s}_{ij}$ and predicted latent shares, $\widetilde{s}^{\star}_{ij}$.

Cross derivative differences with our more flexible specification take the form:

$$\begin{aligned}\widetilde{\frac{\partial s_{ij}}{\partial x_{idk}}} - \frac{\partial s_{id}}{\partial x_{ijk}} \quad = \quad & \frac{\hat{\beta}_k(\hat{s}_{id} - \hat{s}^{*}_{id})\widetilde{s}_{id}\left((1-\hat{s}_{id})\widetilde{s}_{ij} - (1-\widetilde{s}_{id})\hat{s}_{ij}\right) + \hat{\gamma}_k(1-\hat{\mu}_i)(1-\hat{s}_{id})\widetilde{s}_{id}\widetilde{s}_{ij}}{\hat{s}_{id}(1-\hat{s}_{id})} \\ & + \widetilde{s}_{ij}\left[\sum_{k'}(x_{ijk'} - \tilde{x}_{ik'})\alpha_{k,k'} + \widetilde{s}_{id}\sum_{k'} x_{idk'}\alpha_{k'k}\right]\end{aligned} \tag{E.9}$$

where $\tilde{x}_{ik} = \sum_{j\neq d}\widetilde{s}_{ij}x_{ijk}$. Note that when all $\alpha_{k,k'} = 0$, we have: $\widetilde{s}_{ij} = \hat{s}_{ij}$ and there is no difference in estimated cross-derivative differences at the first and second stages. Thus, if there are no significant differences between these cross-derivative difference estimates, we conclude that the DSC model fits the data well. In practise, we estimate the difference in Equations E.7 and E.9 by quantile of $\widetilde{s}_{ij}$.

F Robustness

In this section, we report several robustness checks for the empirical specifications in Sections 4 and 5.

ASC Robustness Table 10 reports estimates of the ASC model for the subset of experimental participants who correctly answered the question testing their understanding of the instructions. The results are very comparable to Table 3 in the text.

Table 11 reports estimates of the ASC model on the Expedia data with 20 goods. The conditional logit model estimated in the main text implies a willingness to pay of \$7 per night for each one position increase in search position on the site. The attentive logit model instead implies a willingness to pay of \$1.77 per night with 20 goods and \$0.08 with 10 goods. In both cases, the predominant impact of search position is on attention.

Table 12 reports estimates of the ASC model on the Expedia data with an outside option included. In this specification, the conditional logit model implies a willingness to pay of \$17.5 per search position, while the attentive logit model implies a willingness to pay of \$3.88 per night. Thus, in all cases, we find that the attentive logit model reduces the apparent willingness to pay for search

Table 10: Experimental Data Estimation Results

	Conditional Logit	Attentive Logit	Truth
Utility:			
Price (dollars)	-0.052***	-0.16***	-0.17***
	(0.004)	(0.033)	(0.005)
Product 1	-1.129***	1.561**	0.751***
	(0.084)	(0.769)	(0.109)
Product 2	-1.577***	0.143	-0.026
	(0.101)	(0.661)	(0.119)
Product 3	-1.331***	0.287	0.329***
	(0.091)	(0.582)	(0.111)
Product 4	-1.544***	0.393	0.234*
	(0.099)	(0.701)	(0.12)
Product 5	-1.162***	1.429*	0.664***
	(0.086)	(0.832)	(0.108)
Product 6	0.26***	0.487***	0.327***
	(0.056)	(0.136)	(0.066)
Product 7	-0.675***	-0.996***	-0.898***
	(0.073)	(0.181)	(0.081)
Product 8	-0.615***	-1.067***	-0.875***
	(0.07)	(0.2)	(0.079)
Product 9	-0.215***	-0.168	-0.311***
	(0.063)	(0.157)	(0.072)
Attention:			
Price (dollars)		0.158***	1.5
		(0.029)	
Product 1		-3.302***	-2.5
		(0.399)	
Product 2		-2.855***	-2.5
		(0.484)	
Product 3		-2.629***	-2.5
		(0.392)	
Product 4		-2.97***	-2.5
		(0.439)	
Product 5		-3.344***	-2.5
		(0.395)	
Product 6		-0.326	0
		(0.317)	
Product 7		0.638	0
		(0.795)	
Product 8		0.725	0
		(0.578)	
Product 9		-0.244	0
		(0.325)	

Notes: Table reports coefficient estimates from conditional logit and attentive logit models. Estimates are the coefficients in the utility and attention equations (not marginal effects). The conditional logit coefficients are recovered from estimating a model assuming all 10 possible goods are considered. The "true" utility parameters are estimated using a conditional logit model given the actual choice set consumers faced. The true attention parameters are known in advance. The attentive model also includes a constant. *** Denotes significance at the 1% level, ** significance at the 5% level and * significance at the 10% level.

Table 11: Expedia Estimation Results: 20 Goods

	Conditional Logit	Attentive Logit
Utility:		
Price (dollars)	-0.015 ***	-0.019 ***
	(0.001)	(0.001)
Hotel Stars (1-5)	0.655 ***	0.688 ***
	(0.033)	(0.078)
Hotel Review Score (1-5)	0.406 ***	0.460 ***
	(0.036)	(0.084)
Popular Brand Indicator	0.146 ***	0.333 ***
	(0.045)	(0.098)
Location Score (normalized)	0.812 ***	0.149 **
	(0.035)	(0.067)
Ongoing Promotion Indicator	0.324 ***	0.005
	(0.042)	(0.094)
Position in Search	-0.083 ***	-0.034 ***
	(0.003)	(0.008)
New Default		0.284
		(0.207)
Attention:		
Price (dollars)		0.000
		(0.001)
Hotel Stars (1-5)		0.162 *
		(0.097)
Hotel Review Score (1-5)		0.053
		(0.095)
Popular Brand Indicator		-0.210
		(0.144)
Location Score (normalized)		1.200 ***
		(0.092)
Ongoing Promotion Indicator		0.573 ***
		(0.138)
Position in Search		-0.086 ***
		(0.009)
New Default		-0.543 **
		(0.242)
Constant		-0.354
		(0.464)

Notes: Table reports coefficient estimates from the Goeree (2008) model. Estimates are the coefficients in the utility and attention equations (not marginal effects). Standard errors are in parentheses. ***Denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. The model is also includes a default which is a randomly chosen alternative for each consumer. Given the estimated attention probabilities, this default is chosen less than 1% of the time.

position by at least 75%.

DSC Robustness Table 13 reports coefficient estimates from the DSC model with brand fixed effects added. Table 14 reports coefficients on the changes on Part D characteristics over time from the specification that allows these changes to impact both utility and attention.

Table 12: Expedia Estimation Results: Outside Option

	Conditional Logit	Attentive Logit
Utility:		
Price (dollars)	-0.008 ***	-0.018 ***
	(0.000)	(0.001)
Hotel Stars (1-5)	0.142 ***	0.271 ***
	(0.024)	(0.083)
Hotel Review Score (1-5)	0.359 ***	0.817 ***
	(0.030)	(0.107)
Popular Brand Indicator	0.241 ***	0.433 ***
	(0.037)	(0.096)
Location Score (normalized)	0.271 ***	-0.007
	(0.018)	(0.065)
Ongoing Promotion Indicator	0.045	-0.094
	(0.037)	(0.099)
Position in Search	-0.140 ***	-0.072 ***
	(0.006)	(0.015)
Outside	5.215 ***	7.291 ***
	(0.108)	(0.474)
Attention:		
Price (dollars)		-0.003 ***
		(0.001)
Hotel Stars (1-5)		0.370 ***
		(0.087)
Hotel Review Score (1-5)		-0.122
		(0.087)
Popular Brand Indicator		-0.151
		(0.124)
Location Score (normalized)		1.093 ***
		(0.088)
Ongoing Promotion Indicator		0.454 ***
		(0.135)
Position in Search		-0.122 ***
		(0.016)
Outside		3.046 ***
		(0.404)
Constant		-0.145
		(0.396)

Notes: Table reports coefficient estimates from the Goeree (2008) model. Estimates are the coefficients in the utility and attention equations (not marginal effects). Standard errors are in parentheses. ***Denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. The model is also includes a default which is a randomly chosen alternative for each consumer. Given the estimated attention probabilities, this default is chosen less than 1% of the time.

Table 13: Part D Results w/ Brand Fixed Effects

	2007		2008		2009	
	Clogit	Alogit	Clogit	Alogit	Clogit	Alogit
Utility:						
Annual Premium (hundreds)	-0.415***	-0.909***	-0.596***	-1.074***	-0.599***	-1.245***
	(0.012)	(0.029)	(0.013)	(0.026)	(0.015)	(0.027)
Annual Out of Pocket Costs (hundreds)	-0.418***	-0.661***	-0.691***	-0.923***	-0.433***	-0.484***
	(0.020)	(0.028)	(0.029)	(0.047)	(0.034)	(0.054)
Variance of Costs (millions)	-2.131***	-3.359***	-1.809***	-2.351***	-2.056***	-0.702
	(0.178)	(0.248)	(0.299)	(0.448)	(0.326)	(0.526)
Deductible (hundreds)	-0.208***	-0.355***	-0.737***	-0.792***	-0.231***	-0.590***
	(0.024)	(0.032)	(0.027)	(0.037)	(0.030)	(0.043)
Donut Hole Coverage	-0.178***	0.505***	-0.263***	-0.798***	1.335***	1.917***
	(0.055)	(0.074)	(0.065)	(0.120)	(0.083)	(0.142)
Average Consumer Cost Sharing %	0.704**	-0.071	-2.002***	-4.274***	0.798**	-1.898***
	(0.280)	(0.376)	(0.333)	(0.450)	(0.358)	(0.541)
# of Top 100 Drugs in Formulary	0.641***	1.078***	0.749***	0.826***	-0.060***	0.022*
	(0.040)	(0.071)	(0.046)	(0.057)	(0.008)	(0.013)
Normalized Quality Rating	0.087***	0.319***	0.299***	0.688***	0.564***	0.659***
	(0.017)	(0.025)	(0.018)	(0.028)	(0.017)	(0.026)
Prior Year Plan	5.930***	-15.619	6.380***	3.370***	6.525***	2.410***
	(0.025)	(846.880)	(0.034)	(0.122)	(0.038)	(0.208)
Attention:						
Annual Premium (dollars)		0.240***		0.364***		0.068**
		(0.016)		(0.023)		(0.027)
Annual Out of Pocket Costs (dollars)		0.141***		0.186***		-0.029
		(0.038)		(0.051)		(0.064)
Variance of Costs (millions)		2.037***		-0.113		1.777***
		(0.315)		(0.455)		(0.589)
Deductible (hundreds)		0.373***		0.182***		0.075
		(0.046)		(0.053)		(0.065)
Donut Hole Coverage		0.829***		-1.364***		-0.268*
		(0.082)		(0.128)		(0.142)
Average Consumer Cost Sharing %		1.321**		-5.493***		0.060
		(0.538)		(0.693)		(0.733)
# of Top 100 Drugs in Formulary		-0.211***		0.429***		0.099***
		(0.065)		(0.102)		(0.021)
Normalized Quality Rating		0.002		0.034		-0.600***
		(0.024)		(0.036)		(0.032)

Notes: "Clogit" refers to the conditional logit model; "alogit" refers to the attentive logit model. The table reports coefficient estimates from the DSC model. Estimates are the coefficients in the utility and attention equations (not marginal effects). The coeffic1ents in the attention equation are the coefficients on the listed characteristics of the default good (demeaned). Standard errors are in parentheses. The attentive model also includes a constant. *** denotes significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level. Standard errors in parentheses.

Table 14: Health Plan Data: Validity of Exclusion Restriction on Changes

Coefficient on Changes:	Utility	Attention
Annual Premium (dollars)	-0.106***	0.055***
	(0.040)	(0.012)
Annual Out of Pocket Costs (dollars)	0.025	-0.003
	(0.018)	(0.006)
Variance of Costs (millions)	-0.240	0.012
	(0.241)	(0.008)
Deductible (hundreds)	-0.112	-0.011
	(0.105)	(0.026)
Donut Hole Coverage	0.176	-0.290***
	(0.278)	(0.080)
Average Consumer Cost Sharing %	1.596	0.589**
	(1.578)	(0.301)
# of Top 100 Drugs in Formulary	-0.089	0.037***
	(0.080)	(0.007)
Normalized Quality Rating	-0.090	0.038
	(0.102)	(0.025)

Notes: Table reports coefficient estimates from the DSC model. Estimates are the coefficients on the changes in the reported variables in the utility and attention equations (not marginal effects). The coefficients in the attention equation are the coefficients on the listed characteristics of the default good (demeaned). The levels are also included in both equations and their coefficients are not shown (for reference, they are quite similar to the specification which omits differences). Standard errors are in parentheses. The attentive model also includes a constant. *** Denotes significance at the 1% level, ** significance at the 5% level and * significance at the 10% level.