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

Jason Abaluck and Abi Adams

*

May 4, 2017

Abstract

Consideration set models relax the assumption that consumers are aware of all available options when making their choice. Thus far, identification arguments in these models have relied either on auxiliary data on what options were considered or on instruments assumed to impact consideration probabilities or utility but not both. We prove that utility and consideration set probabilities can be separately identified without excluding variables from attention or utility even if auxiliary data is unavailable. This greatly expands the potential empirical applications of consideration set models. Our identification proof constructively recovers consideration probabilities from asymmetries in the responsiveness of choice probabilities to characteristics of rival goods. We show in a lab experiment that we can recover preferences and consideration probabilities using only data on which items were ultimately chosen. We show in hotel choice data from Expedia that the model determines that the randomly assigned ordering of hotels in search impacts attention and not utility and that bounds implied by the model predict the benefits of informative advertising. We replicate an earlier finding that health plan choices are more sensitive to characteristics of the plan chosen last year than rival goods and show that this implies that observed inertia is largely driven by inattention but there remain non-trivial adjustment costs.

*Thanks to Leila Bengali and Mauricio Caceres for excellent research assistance and to Dan Ackerberg, Joe Altonji, Dan Benjamin, Steve Berry, Judy Chevalier, Jeremy Fox, Jonathan Gruber, Phil Haile, Erzo Luttmer, Costas Meghir, Olivia Mitchell, Barry Nalebuff, Fiona Scott Morton, 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.

1 Introduction

Discrete choice models typically assume that consumers are aware of all available options. This prevents researchers from asking many questions of interest. Would goods with low sales face higher demand if more consumers noticed them? Will inertial consumers 'wake up' in response to a premium hike but remain unresponsive if rivals lower prices? 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 we can measure preferences conditional on consideration, we can assess the benefits of policies which 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. They allow for the possibility that consumers consider only a subset of possible options when making their choice. This 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, from search costs, or because consumers face (unobserved) constraints on what can be chosen (Beffy, Blundell, Bozio, and Laroque 2014; Gaynor, Propper, and Seiler 2016).[1]

Empirical models in this literature typically rely either on additional data or exclusion restrictions to separate the impact of observables on utility and consideration set probabilities. For example, Conlon and Mortimer (2013) assume that consideration sets are known in some periods, Honka (2014) and Honka, Hortaçsu, and Vitorino (2015) use auxiliary information on which options consumers are aware of,[2] and Goeree (2008) and Gaynor, Propper, and Seiler (2016) assume that some observables impact either attention or utility but not both. Crawford, Griffith, and Iaria (2016) show that identification is possible with panel data if one restricts how consideration sets change over time.

In this paper we prove that the restrictions on choice probabilities imposed by economic theory are sufficient to separately identify preferences and consideration probabilities. Our proof does not require panel data or auxiliary information on consideration sets, and 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 those that we are aware of in the applied literature and does not rely on assuming a particular functional form (e.g. Type 1 Extreme Value) for random utility errors.

Our results suggest that in many applications one could estimate consideration set models rather than the conventional discrete choice models that they nest. In cross-sectional data, one can use our

[1] In this paper, we use "attentive" as synonymous with "a good is in the consumer's consideration set" – this is a shorthand that subsumes the cases where consumers are inattentive, where they are unaware of a product, and where they are subject to unobserved constraints.

[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 – we discuss the relationship between our framework and their model in Section 3.

results to identify whether goods are demanded because they are high-utility or are more likely to be considered. In panel data, one can evaluate whether inertia reflects switching costs or inattention. 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.

The identification result builds on the insight that imperfect consideration results in asymmetric demand responses. Intuitively, imperfect attention breaks the symmetry between cross-price (or other characteristic) responses. For example, if consumers do not consider alternatives unless a default option becomes 'bad enough', choices will be more sensitive to the characteristics of the default option than to rival goods – you might switch if the price of the good you chose last year increases by \$100 but fail to do so if the price of everything else falls by \$100. The link between inattention and symmetry has been noted elsewhere (Chen, Levy, Ray, and Bergen 2008; Cabral and Fishman 2012; Gabaix 2011) but has not previously found its way into the applied literature. More generally, this paper relates to the literature that uses deviations from the neoclassical benchmark to disentangle preferences from other drivers of behaviour (Browning and Chiappori 1998; Blundell, Browning, and Crawford 2008; Abaluck and Gruber 2011; Hastings and Shapiro 2013; Adams, Blundell, Browning, and Crawford 2015).

In contrast to the applied literature to date, theoretical consideration set models have previously derived revealed preference conditions under which information on preferences and consideration can be separately retrieved from choice data (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014). However, under these approaches point identification of model primitives is either not possible or requires a degree of choice set variation that is not observed outside of experimental settings. For example, Manzini and Mariotti (2014) require that the researcher observes choice from every pairwise comparison of available options to pin down attention and the (deterministic) preference relation. Our contribution relative to this literature is to provide conditions for identification that can be applied in the field and which permit utility to be stochastic, as is standard in the discrete choice literature.

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 – our general identification proof subsumes hybrid models combining features of both alternatives. Our results show the possibility of non-parametric estimation, but in most cases, we still advocate estimating parametric generalizations of conventional models. Functional form assumptions in these model are serving the desired role of filling in gaps in the data and are not necessary for identification.

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 accurately recover the probabilities that each good was available as well as recovering the preferences we would estimate conditional on knowing which items were available. We show that our model implies different patterns of elasticities than flexible full-consideration models with a comparable number of parameters.

We also apply the ASC model to hotel choice data from Expedia in which hotel search order was randomized. We show that we can correctly recover that 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 the current demand is due to high utility (meaning that a hotel would be more popular if more people noticed it) or high attention (meaning that additional advertising is unlikely to be effective).

Our second special case is the "Default-Specific Consideration" (DSC) model which permits consideration probabilities to 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 so unsatisfactory that they must 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 through the 'Slutsky asymmetries' this generates; the degree of inattention we estimate is consistent with Heiss, McFadden, Winter, Wupperman, and Zhou (2016). Nonetheless, the remaining adjustment costs are sufficiently large that they offset the cost savings from assigning beneficiaries to the lowest cost plans. 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 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 the general argument and proof. Section 4 lays out the ASC model, validates it with our lab experiment, and applies it to Expedia data. Section 5 lays out 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 nearly all of the assumptions we make here are for expository purposes and will be relaxed in the general model presented in Section 3. For example, our general proof allows for $J > 2$ options and does not require a default good with observable characteristics. The model described here is a special case of the DSC model, which is itself a special case of our more general framework.

Consider a consumer i selecting from two possible products, $j = \{0, 1\}$, for example insurance plans. Each plan is defined as a bundle of characteristics, $x \in \mathbb{R}^K$: annual premiums, the size of the deductible, drug coverage, and so on. 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 good (or bad) the characteristics of the default good are. If the consumer does not pay attention, they pick the default. However, if the consumer also pays attention to the non-default good, then they pick the good that maximises their quasilinear utility function from the set of plans considered.[3]

Let $\phi_i(x_{i0})$ give the probability that consumer i pays attention to both products. The probability that consumer i picks plans j, s_{ij}, in this model can then be expressed as:

$$\begin{aligned} s_{i0}(x_{i0}, x_{i1}) &= (1 - \phi_i(x_{i0})) + \phi_i(x_{i0}) s^\star_{i0}(x_{i0}, x_{i1}) \\ s_{i1}(x_{i0}, x_{i1}) &= \phi_i(x_{i0}) s^\star_{i1}(x_{i0}, x_{i1}) \end{aligned} \tag{1}$$

where $s^\star_{ij}$ gives choice probabilities conditional on paying attention.

ϕ_i, $s^\star_{i0}$ and $s^\star_{i1}$ can all be separately identified in this model. The key to our identification argument is that maximising behaviour when considering all goods implies a symmetry in demand responses to the changes in the characteristics of rival goods, a result analogous to Slutsky symmetry in the continuous demand setting. In the two good case, this symmetry is very intuitive – with full attention and no income effects, consumers should only care about price *differences*. If we see that individuals readily switch when the price of the default changes but not when the price of the alternative changes in the opposite direction, this suggests inattention.

Symmetry of demand responses is violated if changes in product characteristics also impact consideration probabilities. Differentiating Equation 1 and using the fact that the market shares conditional on paying attention satisfy symmetry, we obtain:

$$\frac{\partial s_{i1}}{\partial x^k_{i0}} - \frac{\partial s_{i0}}{\partial x^k_{i1}} = \frac{\partial \phi_i}{\partial x^k_{i0}} s^\star_{i1} = \frac{\partial \log(\phi_i)}{\partial x^k_{i0}} s_{i1} \tag{2}$$

where the second equality follows from the fact that $s_{i1} = \phi_i(x_{i0}) s^*_{i1}$. Thus, changes in the probability of considering both goods are directly identified from data on choice probabilities:

$$\boxed{\frac{\partial \log(\phi_i)}{\partial x^k_{i0}} = \frac{1}{s_{i1}} \left[\frac{\partial s_{i1}}{\partial x^k_{i0}} - \frac{\partial s_{i0}}{\partial x_{i1k}} \right]} \tag{3}$$

To see the intuition for this result, first consider the probability of choosing a good conditional on considering both plans (the s^*_{ij}). A rise in the price of the default plan will increase the share choosing the non-default plan and a rise in the price of the non-default plan will increase the share choosing the default plan. Now consider the unconditional market shares. An increase in the price of the default plan will also increase the probability of paying attention to both goods if individuals only search when the default is particularly unattractive. Thus, the responsiveness of the non-default plan to a reduction in the price of the default plan will be greater than the responsiveness of the default plan to a reduction in the price of the non-default plan, a behavioural 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 $\phi_i = 1$, a

[3] This is a special case of the model in Ho, Hogan, and Scott Morton (2015) and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) which we consider more generally below.

point expanded upon in Section 3. Thus, integrating Equation 3 we obtain:

$$\boxed{\phi_i(x_{i0}) = \exp\left(-\int \frac{1}{s_{i1}}\left[\frac{\partial s_{i1}}{\partial x_{i0}^k} - \frac{\partial s_{i0}}{\partial x_{i1}^k}\right] dx_{i0}^k\right)} \tag{4}$$

Our broad approach, then, is to argue that economic theory imposes restrictions on choice behavior given the assumption that individuals consider every good. Observed deviations from the predictions of the full consideration model are informative about the structure of the underlying consideration probabilities. Of course, there are other reasons one might detect asymmetries in any given case such as misspecification of the parametric model or some other behavioral phenomenon. In the two-good case considered above, one can check the asymmetries for different characteristics are consistent with a common underlying attentive probability. More generally, the J-good model predicts a particular pattern of asymmetries across the different goods. In Sections 3 and 5, we provide tests to check whether the asymmetries observed in the data follow the distinctive pattern implied by inattention.

3 Model & Identification

In this section, we outline formally our analytic framework and nonparametric identification results, drawing the connections between our work and the prior literature. We begin by stating the assumptions that underpin our approach at its most general level before adapting the framework to apply to commonly estimated consideration set models in the applied literature.

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 characterised 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. 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:

$$\mathbb{P}(j) = \{c : c \in \mathcal{P}(\mathcal{J}) \quad \& \quad j \in c\} \tag{5}$$

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

$$s_{ij}(x_i) = \sum_{c \in \mathbb{P}(j)} \pi_{ic}(x_i) s^\star_{ij}(x_i|c) \tag{6}$$

where s_{ij} is the observed probability of 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:

$$\sum_{c \in \mathcal{P}(\{0,\ldots,J\})} \pi_{ic} = 1 \tag{7}$$

$$\sum_{j \in c} s^\star_{ij}(c) = 1 \tag{8}$$

The structural objects of interest are the consideration set probabilities, π_{ic}, and the unobserved latent choice probabilities, $s^\star_{ij}(x_i|c)$. In this paper we do not directly address the identification of preference parameters given knowledge of $s^\star_{ij}(x_i|c)$; the parameters of any utility model that is identified from choice behaviour with full consideration of all goods, can be identified given $s^\star_{ij}(x_i|c)$.

Modeling consideration sets as probabilistic is somewhat more common in the empirical than the theoretical literature on choice with inattention. For example, Eliaz and Spiegler (2011) model each good's membership in the consideration set as binary in their model of competitive marketing, while Masatlioglu, Nakajima, and Ozbay (2012) assume a deterministic mapping from the set of all feasible goods to the consideration set. However, our framework subsumes those developed by the empirical literature, in which consideration sets are more often modelled as stochastic (Goeree 2008; Honka 2014; Honka, Hortaçsu, and Vitorino 2015; Ho, Hogan, and Scott Morton 2015).

Parallels can also be drawn to the literature on mixture models, which are often used as a device to account for unobserved preference heterogeneity and measurement error. In this setting, rather than the true value of a variable or a consumer's 'type' being unobserved, it is the set of goods which individuals consider when making a choice decision that is unknown by the econometrician (see Compiani and Kitamura (2016) for a review of the recent mixtures literature). General results in this literature typically appeal to the availability of instruments or shape restrictions on the underlying structural functions for identification. Our approach differs in that we derive restrictions from consumer theory which are sufficient for point identification.

To make progress towards point identification of the structural functions of interest, we will have to place some restrictions on consideration set probabilities. If the π_{ic} are allowed to vary arbitrarily, then identification of the underlying structural functions is hopeless. This fact has been noted in a number of sources in the econometrics literature on mixtures and in theoretical work on consideration sets (Henry, Kitamura, and Salanié 2014). For example, Theorem 2 of Manzini and Mariotti (2014) proves that unrestricted dependence of consideration set probabilities on the set of goods considered yields a model with no observable restrictions. As a simple illustration, consider a case with two choice sets, $\{0,1\}$ and $\{0,1,2\}$ where:

$$s_{i1} = \pi_i(\{0,1\}) s^\star_{i1}(\{0,1\}) + \pi_i(\{0,1,2\}) s^\star_{i1}(\{0,1,2\}) \tag{9}$$

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. Thus, to uniquely identify consideration set probabilities, restrictions must be imposed on how these probabilities can vary with the underlying characteristics of available goods.

It should thus be unsurprising that restrictions on consideration set formation are already made in the applied literature and, in fact, arise naturally in many applications. For example, Goeree (2008) and Gaynor, Propper, and Seiler (2016) assume that each good has an independent consideration probability which depends only on the characteristics of that good, Ho, Hogan, and Scott Morton (2015) assumes that consideration probabilities depend only on characteristics of default goods, and Honka (2014) and Honka, Hortaçsu, and Vitorino (2015) assume consideration probabilities depend on a ranking of goods based on a index which depends only on characteristics of the good in question. What the above example rules out is a model where consideration probabilities for every good (and more generally, every set of goods) vary in a completely flexible with the characteristics of *all* alternative goods.

Below we outline a set of restrictions on π that encompass all the consideration set models we are aware of in the applied literature. Our key argument is not that one needs *no* restrictions to identify consideration set models but rather that current empirical papers rely on *additional* exclusion restrictions that are unnecessary for identification, such as excluding attributes such as advertising or price from entering either utility or consideration probabilities.

3.1 Key Assumptions

Individuals are assumed to make choices from any given consideration set to maximize their utility. We take a random utility approach, decomposing the overall utility of good j, u_{ij}, into a deterministic component that depends on the characteristics of good j and a random error term:

$$u_{ij} = v_{ij}(x_{ij}) + \epsilon_{ij} \tag{10}$$

We make the following assumptions:

Assumption 1. *Quasi-linearity* There exists a characteristic x^1_{ij} that enters the indirect utility function linearly and with homogenous coefficients across goods, although this coefficient can be heterogeneous across individuals.

$$\begin{aligned} u_{ij} &= v_{ij}(x_{ij}) + \epsilon_{ij} \\ &= \beta_i x^1_{ij} + w_{ij}(x^2_{ij}) + \epsilon_{ij} \end{aligned} \tag{11}$$

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 $j = 0, ..., J$.

While quasilinearity can be a restrictive assumption, it can be relaxed with additional functional form restrictions so long as the parameters of the indirect utility function stay constant across goods. Further, this assumption will be of little consequence if the budget share allocated to the goods in question is relatively small (Hausman and Newey 2016). Price is not required to be the quasilinear characteristic, and thus nonlinearities in price effects can still be incorporated. However, it is natural to choose price as the quasilinear characteristic given that maximisation of a quasilinear direct utility function implies that the marginal utility of income should be the same across goods.

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 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:

$$s^\star_{ij}(c) = Pr\left(v_{ij} + \epsilon_{ij} = \max_{j' \in c} v_{ij'} + \epsilon_{ij'}\right)$$
$$= \int\int\int_{-\infty}^{v_{ij}+e-v_{il}} ... \left[\int_{-\infty}^{v_{ij}+e-v_{ij}}\right] ... \int_{-\infty}^{v_{ij}+e-v_{il'}} f_c(z_l,...,e,...,z_{l'})\mathrm{d}z_{l'}...[\mathrm{d}z_j]...\mathrm{d}z_l\mathrm{d}e \; \mathrm{d}F(\beta_i) \tag{12}$$

Maximization given the structure placed on the utility function implies some specific restrictions on choice probabilities from a given consideration set, in addition to those imposed by probability theory.

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

$$\frac{\partial s^\star_{ij}}{\partial x^1_{ij'}} = \frac{\partial s^\star_{ij'}}{\partial x^1_{ij}} \tag{13}$$

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

$$s^\star_{ij}(x^1_i, x^2_i) = s^\star_{ij}(x^1_i + \delta, x^2_i) \tag{14}$$

Proof in Appendix A.

The identification results in this paper apply to models in which consideration set probabilities vary with product characteristics. We cannot say anything about identification of the underlying structural functions in situations where consideration set probabilities are independent of characteristics.

Assumption 5. π_{ic} is continuously differentiable with, for $\pi_{ic} < 1$:

$$\frac{\partial \pi_{ic}}{\partial x^1_{ij}} \neq 0 \tag{15}$$

This assumption is natural in most applied settings of interest and is only unlikely to hold when consideration is random. Further, what drives changes in attention is a question of interest in itself. The impact of advertising on consideration sets is a key question in the marketing literature (Shocker, Ben-Akiva, Boccara, and Nedungadi 1991; Hauser 2014) and the theoretical and empirical IO literature (Eliaz and Spiegler (2011) and Hastings, Hortaçsu, and Syverson (2013) give additional references). This metric is also relevant for welfare analysis. In the related literature on tax salience, Taubinsky and Rees-Jones (2016) show that new channels through which policies can influence demand behaviour and welfare arise when attention varies with economic incentives.

Slutsky Asymmetries & Nominal Illusion In this model, only one mechanism is available to generate asymmetries. Thus, with no additional restrictions on consideration probabilities, cross-derivative asymmetries and nominal illusion will always be evidence of imperfect consideration. More generally, other behavioral anomalies could result in asymmetric responses, and we describe and implement overidentification tests to check whether the pattern of asymmetries is consistent with our model of imperfect consideration.

Lemma 1. Asymmetries & Nominal Illusion Imply Imperfect Consideration. Given Assumptions 1-5, if

$$\frac{\partial s_{ij}}{\partial x^1_{ij'}} \neq \frac{\partial s_{ij'}}{\partial x^1_{ij}} \tag{16}$$

$$s_{ij}(x^1_i, x^2_i) \neq s_{ij}(x^1_i + \delta, x^2_i) \tag{17}$$

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

Our maintained assumptions allow us to make statements about the consistency of observed choice behaviour and a full consideration model. However, they are not sufficient for point identification of the structural functions of interest.

Most consideration set models in the applied literature take one of two forms. One strand of the literature assumes that each good has an independent probability of being considered which depends on characteristics of that good. This includes the models in Goeree (2008) and Gaynor, Propper, and Seiler (2016). Under this 'Alternative-Specific Consideration' (ASC) approach, consideration set probabilities take the form:

$$\pi_{ic} = \prod_{j \in c} \phi_{ij}(x_{ij}) \prod_{j' \notin c} \left(1 - \phi_{ij'}(x_{ij'})\right) \tag{18}$$

In this model, one needs to specify what good is chosen if consideration sets are empty. This can be an inside good or an outside good with utility normalised to 0 (what we refer to as an 'outside' default).

An alternative approach, the 'Default-Specific Consideration' (PSC) model, assumes the existence of a default good with observed characteristics and allows the probability of considering all alternative options to vary only as a function of the characteristics of that default.[4] Ho, Hogan, and Scott Morton (2015) and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) develop models in which 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. Under this approach, the market shares of the default (good 0) and non-default goods take the form:

$$s_{i0} = (1 - \phi_{i0}(x_{i0})) + \phi_{i0}(x_{i0}) s^{\star}_{i0}(\mathcal{J}) \tag{19}$$

$$s_{ij} = \phi_{i0}(x_{i0}) s^{\star}_{ij}(\mathcal{J}) \quad \text{for } j > 0 \tag{20}$$

where ϕ_{i0} gives the probability of considering all available products.

We adapt our general framework to encompass both the ASC and PSC models. Let the market share of the inside default, good-0, and non-default goods take the form:

$$s_{i0} = (1 - \phi_{i0}(x_{i0})) + \sum_{c \in \mathbb{P}(0)} \prod_{j \in c} \phi_{ij}(x_{ij}) \prod_{j' \notin c} \left(1 - \phi_{ij'}(x_{ij'})\right) s^{\star}_{i0}(c) \tag{21}$$

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{j \in c} \phi_{ij}(x_{ij}) \prod_{j' \notin c} \left(1 - \phi_{ij'}(x_{ij'})\right) s^{\star}_{ij}(c) \quad \text{for } j > 0 \tag{22}$$

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

Restricting $\phi_{ij} = 1$ for all $j > 0$ gives the PSC model. Restricting $\phi_{i0} = 1$ 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' is not available. For example, in the beer market, a local craft beer might not be available in all locations or, in 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 and not containing $j' > 0$ be given as:

$$\mathbb{P}(j/j') = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad j' \notin c \quad \& \quad 0 \in c\} \tag{23}$$

[4] We shall refer to a default option with observed characteristics as a "inside default". If the characteristics of the default are unobserved (e.g. the option not to buy anything), we refer to this as an "outside default".

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

$$s_{ij}(\mathcal{J}/j') = \sum_{c \in \mathbb{P}(j/j')} \prod_{l \in c} \phi_l \prod_{l' \notin \{c, j'\}} (1 - \phi_{l'})\, s^{\star}_{ij}(c) \tag{24}$$

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' plus the impact of j' on purchasing good j within each consideration set including j:

$$s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j') = \overbrace{\phi_{ij'}}^{\text{Probability consider } j'} \underbrace{\sum_{c \in \mathbb{P}(j/j')} \prod_{l \in c} \phi_l \prod_{l' \notin \{c, j'\}} (1 - \phi_{l'}) \left(s^{\star}_{ij}(c \cup j') - s^{\star}_{ij}(c)\right)}_{\text{Impact of adding } j' \text{ to all consideration sets including } j} \tag{25}$$

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, the size of the cross derivative differences depends on the degree to which consideration probabilities are altered, plus 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:

$$\frac{\partial s_{ij}}{\partial x^1_{ij'}} - \frac{\partial s_{ij'}}{\partial x^1_{ij}} = \frac{\partial \log(\phi_{ij'})}{\partial x^1_{ij'}} \left(s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')\right) - \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} \left(s_{ij'}(\mathcal{J}) - s_{ij'}(\mathcal{J}/j)\right) \tag{26}$$

Changes in the 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:

$$\frac{\partial s_{ij}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{ij}} = \frac{\partial \log(\phi_{i0})}{\partial x^1_{i0}} s_{ij}(\mathcal{J}) - \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} \left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)\right) \tag{27}$$

Equations 26 and 27 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:

$$y_i = D_i \beta_i \tag{28}$$

where y_i is the vector of cross derivative differences, β_i is the $J+1$-vector of log consideration probability derivatives, and D_i is the coefficient matrix of leave-one-out differences. To better understand the structure of these matrices, consider the just identified case with $J = 2$:

$$\begin{bmatrix} \frac{\partial s_{i1}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{i1}} \\ \frac{\partial s_{i2}}{\partial x^1_{i0}} - \frac{\partial s_{i0}}{\partial x^1_{i2}} \\ \frac{\partial s_{i1}}{\partial x^1_{i2}} - \frac{\partial s_{i2}}{\partial x^1_{i1}} \end{bmatrix} = \begin{bmatrix} s_{i1}(\mathcal{J}) & -(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1)) & 0 \\ s_{i2}(\mathcal{J}) & 0 & -(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2)) \\ 0 & -(s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)) & (s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)) \end{bmatrix} \begin{bmatrix} \frac{\partial \phi_{i0}}{\partial x_{i0}} \\ \frac{\partial \phi_{i1}}{\partial x_{i1}} \\ \frac{\partial \phi_{i2}}{\partial x_{i2}} \end{bmatrix} \tag{29}$$

This example is continued in Appendix A.

As there are typically more than $J+1$ cross-derivative differences, it is convenient to work with the system:[5]

$$D_i' y_i = D_i' D_i \beta_i \tag{30}$$

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

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

As discussed in Appendix A, the rank condition will be satisfied if $J \geq 2$ and[6]

$$s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j') \neq 0 \tag{31}$$

$$\frac{s_{il}(\mathcal{J}) - s_{il}(\mathcal{J}/j)}{s_{ij'}(\mathcal{J}) - s_{ij'}(\mathcal{J}/j)} \neq \frac{s_{il}(\mathcal{J}) - s_{il}(\mathcal{J}/j')}{s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')} \tag{32}$$

for all $j, j', l \in \mathcal{J}$ with $j, j' > 0$. Equation 31 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 32 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. If the rank condition holds, then the derivatives of log consideration probabilities are given as:

$$\beta_i = \left(D_i' D_i\right)^{-1} D_i' y_i \tag{33}$$

In other words, the derivatives of the consideration set probabilities can be recovered from a regression of cross-derivative asymmetries on an appropriately constructed matrix of changes in choice probabilities when goods are removed from the choice set.

Harnessing variation generated by removing goods from choice sets has a central role in revealed preference conditions for rationalizability by a consideration set model (Masatlioglu, Nakajima, and Ozbay (2012), Manzini and Mariotti (2014)). Our proof relies on much weaker assumptions than these papers about the amount of choice set variation required and (as we will show) still yields point identification of the underlying structural functions of interest. Leave-one-out choice set variation can be replaced by a large support assumption on x_{ij}^1 for nonparametric identification of changes in choice probabilities. In practice, parametric assumptions can also replace this type of variation in the data as we will discuss in Section 4.

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

Assumption 7A imposes that at poor 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 $x_{ij}^1 \to -\infty, s_{ij}^\star(c) \to 0$

[5] Any alternative weighting matrix can be used in place of D_i'.

[6] Note that in the two special cases of the model, the ASC and PSC frameworks, the order condition falls to $J \geq 1$, i.e. only one non-default good is required.

for all $c \in \mathbb{P}(j)$) even if it is available. Then,

$$s_{ij}(x^1_{ij'}, x^2_{ij'}, \{x^1_{il}, x^2_{il}\}_{l\neq j'}|\mathcal{J}) = s_{ij}(\mathcal{J}/j') + \phi_{ij'}(x^1_{ij'}, x^2_{ij'}) \sum_{c\in\mathbb{P}(j/j')} \prod_{l\in c}\phi_l \prod_{l'\notin\{c,j'\}} (1-\phi_{l'}) \left(s^\star_{ij}(c\cup j') - s^\star_{ij}(c)\right) \tag{34}$$

$$\to s_{ij}(\mathcal{J}/j') \quad \text{as} \quad x^1_{ij'} \to -\infty \tag{35}$$

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^1_{ij}:

$$\log(\phi_{ij}) = \int \beta_{il}\ \mathrm{d}x^1_{ij} + C \tag{36}$$

where $l = j+1$ for $j > 0$ as the $(j+1)^{th}$ element of β_i corresponds to $\partial \log(\phi_{ij})/\partial x^1_{ij}$.

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^1_{ij} 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), Magnac and Maurin (2007)), the identification of binary games and entry models (Tamer (2003), Fox, Hsu, and Yang (2012), Lewbel and Tang (2015)), 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.

Large support assumptions might be thought unattractive in practice because 'identification at infinity' can affect inference. However, this will not often be a problem in our intended applications. **?**) define 'thin set identification' as that which requires covariates to take on a range of values that have probability zero. When this results in slower than $\sqrt{N}$-rates of estimation, it is termed 'irregular identification'. So long as x^1_{ij} has a strictly positive probability of attaining the value at which attention is paid with probability one, this problem is overcome.

Assumption 7b. As $x^1_{ij} \to \infty$, $\phi_{ij} \to 1$.

Note that Assumption 7b is still much weaker than that used in empirical studies currently. A series of applied papers assume that there exist markets and/or subgroups of consumers in which *all* goods are perfectly considered in order to identify inattention in other markets. For example, new entrants to a market and 'active' switchers are assumed to pay attention to all available goods in Heiss, McFadden, Winter, Wupperman, and Zhou (2016). We only require that there exists one market in which good j is paid attention to for nonparametric identification; in the same market, other goods might not form part of the consumer's consideration set.

Identifying Full Consideration Market Shares There are 2^J independent $s^\star_{ij}$ to identify in our model. This identification problem — how do we identify $s^\star_{ij}$ given knowledge of ϕ_{ij} and s_{ij} — is analogous to the problem of identifying the 'long' regression in Cross and Manski (2002). While in similar settings the functions of interest are typically only partially identified without instruments (Henry, Kitamura, and Salanié 2014), we will again show that the restrictions implied by optimizing behavior again result in point identification of the objects of interest. We will treat ϕ_{ij} as known in this subsection given the argument above.

Nominal illusion facilitates the identification of the latent choice probabilities, $s^\star_{ij}$. Imagine that $N = 2^J$ level shifts in the quasilinear characteristic are observed. Let $k = 1, ..., C$ index the consideration sets of which j is a member. The probabilities of these consideration sets containing j will be given as $\pi_{j1}, ..., \pi_{jC}$. Then, for each good $j > 0$,[7] define the matrices:

$$\Pi_j(\delta) = \begin{bmatrix} \pi_{j1}(\delta_1) & \cdots & \pi_{jC}(\delta_1) \\ \vdots & \ddots & \vdots \\ \pi_{j1}(\delta_N) & \cdots & \pi_{jC}(\delta_N) \end{bmatrix} \tag{37}$$

$$s^\star_{ij} = \left[s^\star_{ij}(c_{j1}), ..., s^\star_{ij}(c_{jC})\right] \tag{38}$$

$$s_{ij}(\delta) = \left[s_{ij}(x^1_i + \delta_1, x^2_i), ..., s_{ij}(x^1_i + \delta_N, x^2_i)\right] \tag{39}$$

where

$$\pi_c(\delta) = \prod_{l \in c} \phi_{il}(x^1_{il} + \delta) \prod_{l' \notin c} \left(1 - \phi_{il'}(x^1_{il'} + \delta)\right) \tag{40}$$

with the dependence of ϕ_{ij} on x^2_{ij} suppressed for notational simplicity. Unobserved latent choice probabilities can then be identified as the unique solution to the following linear system:

$$\Pi(\delta) s^\star_i = s_i(\delta) \tag{41}$$

$$s^\star_i = \Pi^{-1}(\delta) s_i(\delta) \tag{42}$$

where

$$\Pi(\delta) = \begin{bmatrix} \Pi_1(\delta) & 0 & \cdots & 0 \\ 0 & \Pi_2(\delta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Pi_J(\delta) \end{bmatrix} \tag{43}$$

$$s^\star_i = [s^\star_{i1}, ..., s^\star_{iJ}] \tag{44}$$

$$s_i(\delta) = [s_{i1}(\delta), ..., s_{iJ}(\delta)] \tag{45}$$

There is a unique solution to this system, and thus all $s^\star_{ij}$ are identified, when $\Pi^{-1}(\delta)$ is full rank.

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

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

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

$$\frac{\phi_{ij}(x_{ij}^1+\delta_i)}{1-\phi_{ij}(x_{ij}^1+\delta_i)} \neq \frac{\phi_{ij}(x_{ij}^1+\delta_{i'})}{1-\phi_{ij}(x_{ij}^1+\delta_{i'})} \tag{46}$$

$$\phi_{ij}(x_{ij}^1+\delta_i) \neq \phi_{ij'}(x_{ij'}^1+\delta_i) \quad \text{at, at least one } i=1,...,N \tag{47}$$

for $i \neq i'$ and $j \neq j'$. This requires that consideration set probabilities vary with δ, that ϕ_{ij} is a nonlinear function of characteristics, and that ϕ_{ij} and $\phi_{ij'}$ are not perfectly exchangeable. These are natural conditions given that ϕ_{ij} is bounded between zero and one, the characteristics of goods vary (imperfect exchangeability), and in circumstances where consideration probabilities are less than one. Further, 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, $\phi_{ij}(x_{ij})$, and latent market shares conditional on consideration set c, $s_{ij}^{\star}(c)$ are identified for all $c \in \mathbb{P}(\mathcal{J})$ and $j \in \mathcal{J}$.

3.3 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 30, changes in consideration set probabilities, β_i, are defined by the linear system:

$$D_i\beta_i - y_i = 0 \tag{48}$$

where y_i is the vector of cross derivative differences, β_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 thus

$$\underbrace{\frac{1}{2}J(J+1)}_{\#\text{ Independent Cross Deriv. Diffs}} - \overbrace{(J+1)}^{\#\ \phi_{ij}\text{ Derivatives}} \tag{49}$$

overidentifying restrictions. Similar reasoning shows that there are $N - 2^J$ overidentifying restrictions for latent market shares. Let $\widetilde{D}_i$ give a subset of cross derivative difference conditions for which $\text{rank}(\widetilde{D}_i) = J+1$, with $[\widetilde{D}_i; \widehat{D}_i] = D_i$, $[\widetilde{y}_i; \widehat{y}_i] = y_i$ and

$$\widetilde{\beta}_i = \widetilde{D}_i^{-1}\widetilde{y}_i \tag{50}$$

In the absence of sampling variation, our framework would be rejected if:

$$\widehat{D}_i \widetilde{\beta}_i \neq \widehat{y}_i \tag{51}$$

In practice, cross derivative differences and market shares are estimated and subject to sampling variation. In this case, one could adapt Generalised Method of Moment tests for over-identifying restrictions. We describe and implement a practical overidentification test in Section 5 that checks explicitly whether the patterns of cross-derivatives we observe in the data scale in the way predicted by our consideration set model.

4 Alternative-Specific Consideration

Theorem 1 provides a powerful nonparametric identification result that can, in theory, be bought directly to data to recover consideration probabilities and latent market shares (and thus utility parameters). However, in most settings of applied interest there will not be sufficient data for nonparametric estimation of the structural functions of interest. We thus consider parametric versions of the two special cases which motivate our general model.

In this section, we develop a parametric version of the general ASC framework, in which consideration probabilities for each good depend on characteristics of that good. In the next section, we consider the Default-Specific Consideration model, in which there is a single consideration probability which depends on characteristics of the default good. The proof in Section 3 shows that one can, if desired, combine both models.[8]

The ASC model is appropriate in many settings where one observes cross-sectional data with no clear default good. In online applications, for example, the ranking of a product in search will depend on attributes of that product. In bricks and mortar retail, the shelf a product is on or the location in the store is likewise chosen based on observable attributes of that product. With a large number of products, the impact of the characteristics of any single rival product on consideration probabilities may be second order. This model has been applied in the literature (Goeree 2008; Gaynor, Propper, and Seiler 2016) but has thus far relied on additional exclusion restrictions for identification.

We test our ability to correctly recover preferences and consideration set probabilities in an experimental setting and then apply the model to Expedia hotel choice data. In the Expedia extract we use, the order of the hotels in search results was randomly assigned, and we use this variation to validate our model.

[8]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.

4.1 Nonparametric Identification

The ASC model defines market shares as:

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il}(x_{il}) \prod_{l' \notin c} (1 - \phi_{il'}(x_{il'}))\, s^{\star}_{ij}(c) \tag{52}$$

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 choice set c. This model is a special case of our general framework with the restriction $\phi_{i0} = 1$. To complete the model, one must specify what happens if consideration sets are empty – 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.

In scenarios where the default option is an inside good, a very simple closed form expression is available for changes in consideration probabilities.

$$\boxed{\frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} = \frac{\frac{\partial s_{i0}}{\partial x^1_{ij}} - \frac{\partial s_{ij}}{\partial x^1_{i0}}}{s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)}} \tag{53}$$

Intuitively, if the default is more responsive to changes in the characteristics of good j than vice-versa, this suggests that consideration set probabilities vary with good j's characteristics. A given asymmetry implies a larger change in attention for goods that generate little utility — this is revealed by the degree to which the demand for rival goods changes when good j is removed (the denominator of Equation 53).

Given Assumption 7A, a closed form expression for consideration probabilities is given as:

$$\boxed{\phi_{ij} = \exp\left(-\int \frac{\frac{\partial s_{i0}}{\partial x^1_{ij}} - \frac{\partial s_{ij}}{\partial x^1_{i0}}}{s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)}\, \mathrm{d}x^1_{ij}\right)} \tag{54}$$

4.2 Parametric Representation

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

$$u_{ij} = x_{ij}\beta_i + \epsilon_{ij} \tag{55}$$

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:

$$\phi_{ij} = Pr\left(x_{ij}\gamma - \eta_{ij} > 0\right) \tag{56}$$

$$= \frac{\exp(x_{ij}\gamma)}{1+\exp(x_{ij}\gamma)} \tag{57}$$

This model can be estimated by maximum likelihood, but becomes computationally intractable as J grows because the likelihood sums over the 2^J possible consideration sets. Goeree (2008) shows how this problem can be avoided using a simulated likelihood approach. We provide details in Appendix D. We refer to parametric specifications that assume type I extreme value errors for utility and a logit specification for consideration probabilities as "attentive logit" models.

These assumptions allow us to express this consideration set model as a full attention model where the utility of each good depends directly on the characteristics of rival goods:[9]

$$v_{ij} = x_{ij}\beta_i + \psi_{ij} + \epsilon_{ij} \tag{58}$$

where:

$$\begin{aligned}\psi_{ij} &= \ln\left(\frac{s_{ij}}{1-s_{ij}}\right) - \ln\left(\frac{s^{\star}_{ij}(j)}{1-s^{\star}_{ij}(j)}\right) \\ &= \ln\left(\frac{\phi_{ij}\sum_{k\neq j}\exp(x_{ik}\beta_i+\psi_{ik})}{(1-\phi_{ij})\exp(x_{ij}\beta_i)+\sum_{k\neq j}\exp(x_{ik}\beta_i+\psi_{ik})}\right)\end{aligned} \tag{59}$$

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. ψ_{ij} gives the difference between (a monotonic function of) the observed probability of choosing the each option given full and imperfect consideration. This representation builds 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.

Unlike a traditional logit model, the utility of good j in this representation depends directly on the characteristics of other goods via the ψ_{ij} term if the consumer has some probability of being inattentive to good j. This dependence effectively "undoes" the substitution that would occur in response to changes in the characteristics of rival goods in a full consideration model. 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.

[9] See Appendix B for a proof of this.

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 perhaps an unreasonable amount of data is needed before they can be estimated.

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 parameters fail to do so.

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 1. 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 this incentive scheme and its implication and they 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 E.1 reports results using only this subset of users who passed this test and shows that results are qualitatively unchanged.

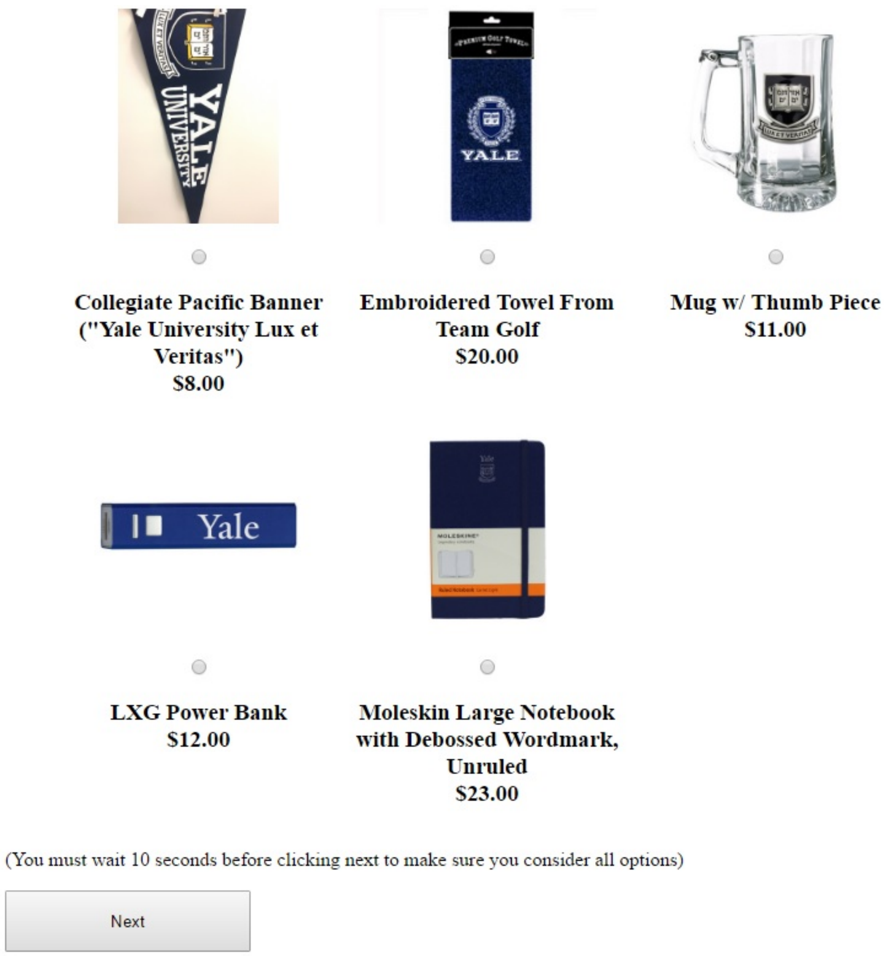
Table 1: Product Names and Prices

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

The probability that each good was in a given choice set was fixed in advance – this probability varied across goods and varied 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 (the logit coefficients that determine the consideration probabilities are given in Table 2). To increase the likelihood that participants considered all of these products, we required consumers to spend at least 10 seconds with each choice before finalizing their choices. This allows us to take choices given the actually available alternatives as representative of consumers' true preferences. A sample product selection screen is shown in Figure 1. 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.

Figure 1: Lab Experiment: Sample Product Selection Screen



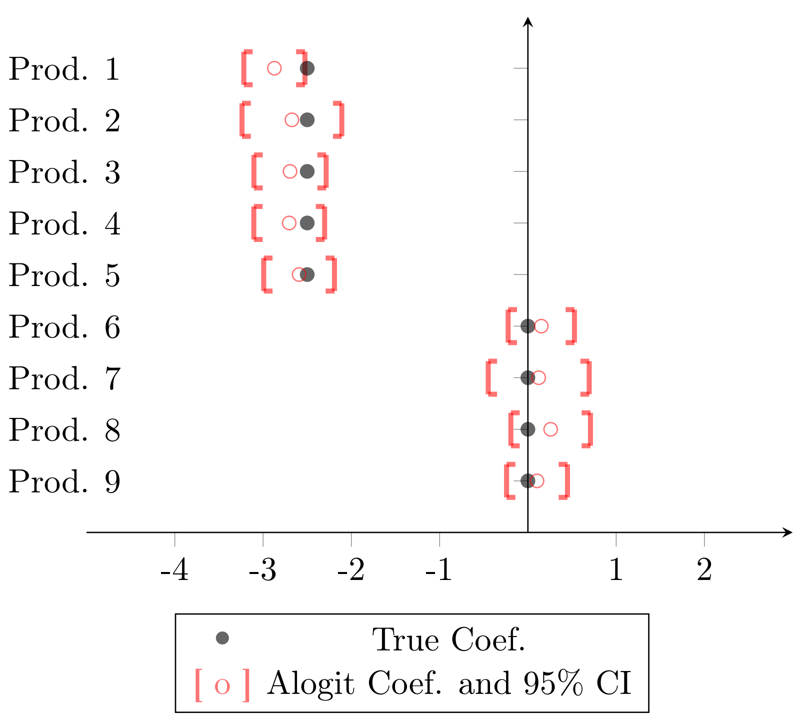
Estimation of the ASC model requires specifying a default good which is chosen if the consideration set is empty. We specify this as good 10. At the estimated parameter values, an empty consideration set has a 0.2% chance of occurring so the choice of default does not impact estimation. To recover the 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.[10]

[10]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

Results 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 consideration parameters are known in advance – the consideration probability for each good is a logit function of the indicated parameters. 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 first the consideration parameters (the second panel of Table 2). In the attention probability equation, 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). 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. Figure 2 shows this information graphically. Table 3 shows that the implied average consideration probabilities by good track the known consideration probabilities closely. 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.

Figure 2: Product Fixed Effects in Attention: Truth vs. ASC Model



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 effected 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 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.

Consider next the preference parameters shown in the top panel of Table 2 (we consider the implied elasticities below). The claim that the model recovers the true preference parameters requires assuming that the 'truth' in that case is a conditional logit model estimated using realised 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 which 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.

Table 3: Consideration Probabilities: Actual vs. Estimated

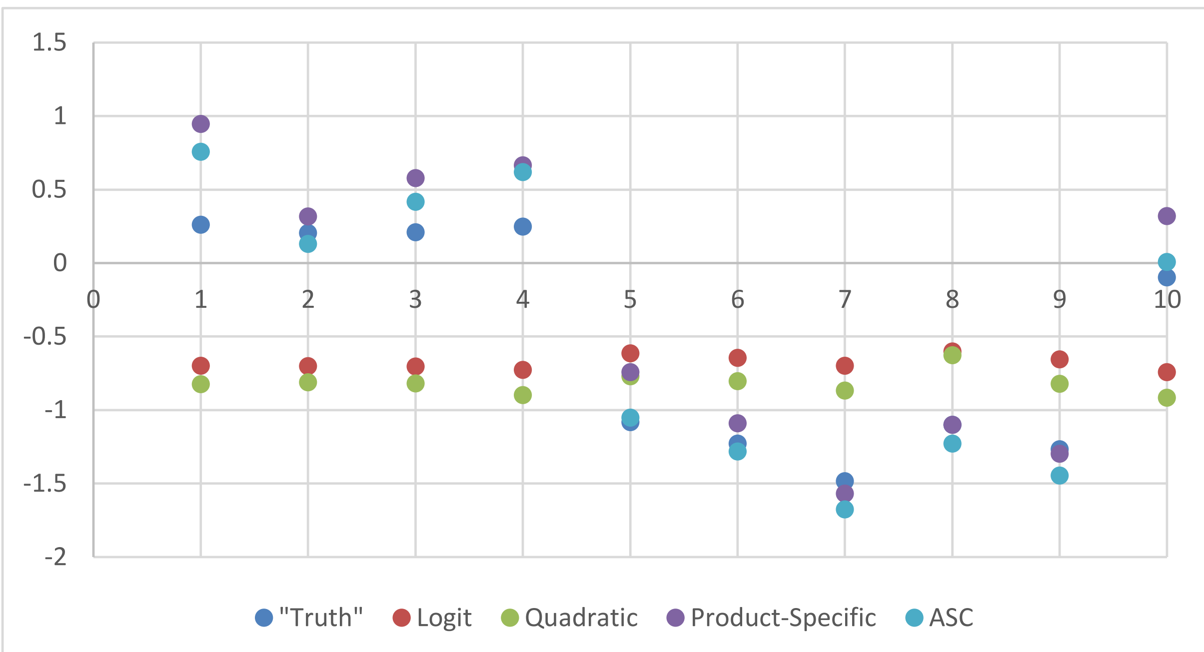
	Truth	Attentive Logit
Consideration Probability		
1	18.3%	10.7%
2	19.7%	13.5%
3	18.0%	12.3%
4	17.7%	12.2%
5	19.3%	14.4%
6	70.3%	68.1%
7	69.7%	67.4%
8	64.9%	65.2%
9	69.5%	67.1%
10	70.0%	64.9%

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

An additional question of interest is how the model performs in terms of accounting for price elasticities relative to 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 3 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.

Figure 3: Experimental Data: Own-Price Elasticities by Good



Of course, the game is rigged in that this setting is designed to perfectly match the underlying structural assumptions about consideration set probabilities. This exercise serves as a proof of principle that the attentive logit model can recover elasticities, preference parameters and consideration set probabilities given observed choices. But what about real-world settings where consideration probabilities arise organically as a function of the choice environment? 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 in these environments.

4.4 Informative Advertising and Hotel Choice

We will 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 our model identifies that search position impacts attention but not utility, as well as to validate the model as a tool for generating out of sample predictions of the efficacy of promotions that increases 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.[11]

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.

After restricting to the sample with a randomized hotel ordering in search results, we end up with 2,441 total queries for which we observe a final transaction which span 9,851 hotels. Summary statistics from our data after all sample selection restrictions are imposed are reported in Table 4. 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 they may believe that higher ranked hotels are better in some unobservable respect.

The estimation results for a conditional logit model and the attentive logit model, are shown in Table 5. First, note that in both models, all the coefficients have the expected sign – consumers dislike high prices and they like hotels with more stars, higher review scores, better locations and higher positions in 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.

[11] 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.

Table 4: Expedia Data: Summary Statistics

	A. All Hotels	*B. Chosen Hotels*
Price (dollars)	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 Hotels	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.

The attentive logit model shows that the impact of search position on choices comes entirely through the impact on attention rather than utility. The model also implies that consumers are much more likely to consider hotels which 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 6 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 seems to modestly attenuate 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).

Because 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 2 to 1 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 can 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

Table 5: 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)

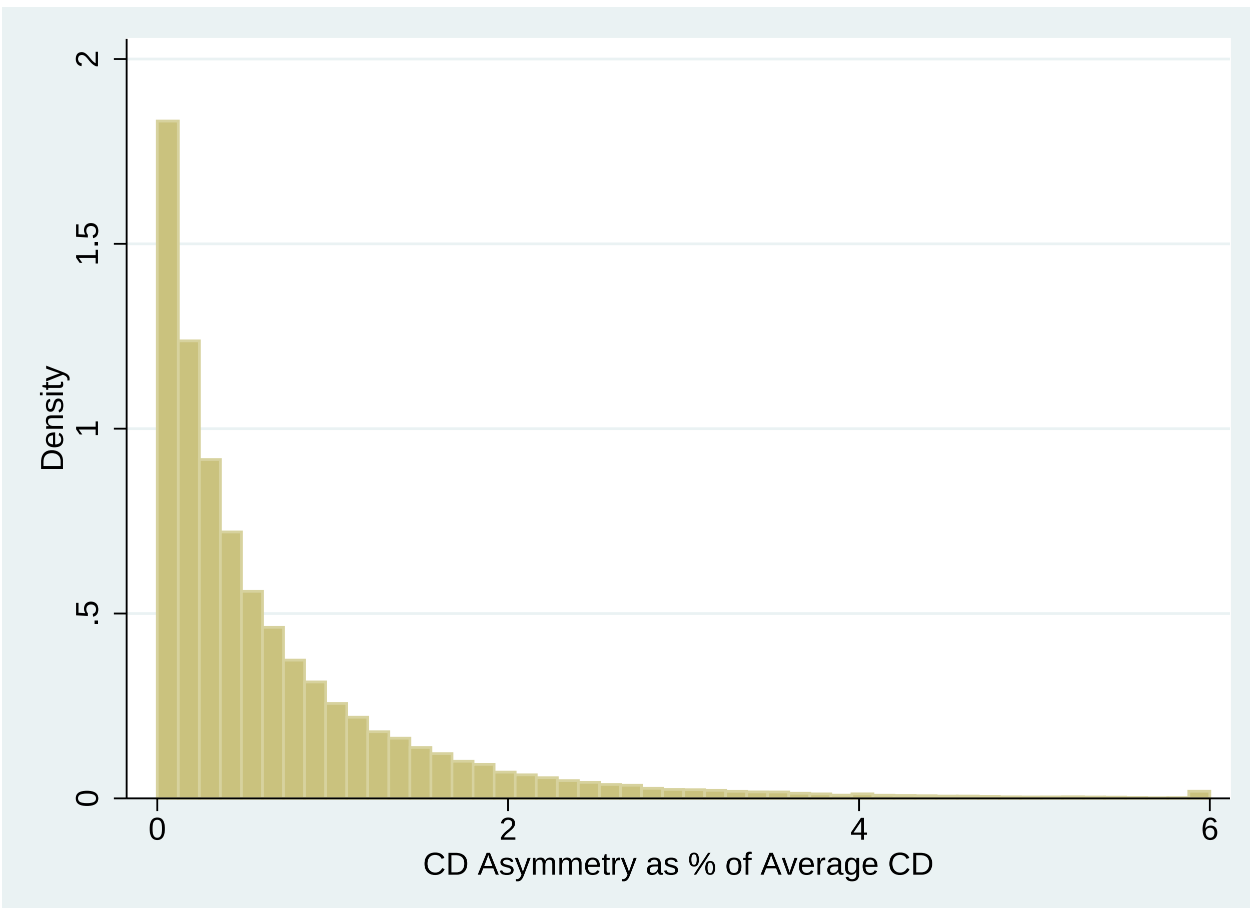
Notes: Table reports coefficient estimates from the Goeree (2008) model. Estimates are the coefficients in the utility and attention equations (not marginal effects). Standard errors are in parentheses. ***Denotes significance at the 1% level, **Denotes significance at the 5% level and *Denotes significance 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.

the asymmetry as a percentage of the average absolute cross-effect in the data. Figure 4 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.

Table 6: Expedia Data: Choice Probabilities and Elasticities

Search Position	Market Share	Attentive Probability	Conditional Logit Elasticity	Attentive Logit Elasticity
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

Figure 4: Estimated Asymmetries in Position Responses



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 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 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 5 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) and the "x" indicates the average demand observed in the data for hotels of each type. 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 7 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.

Figure 5: Expedia: Bound vs. Demand by Search Position

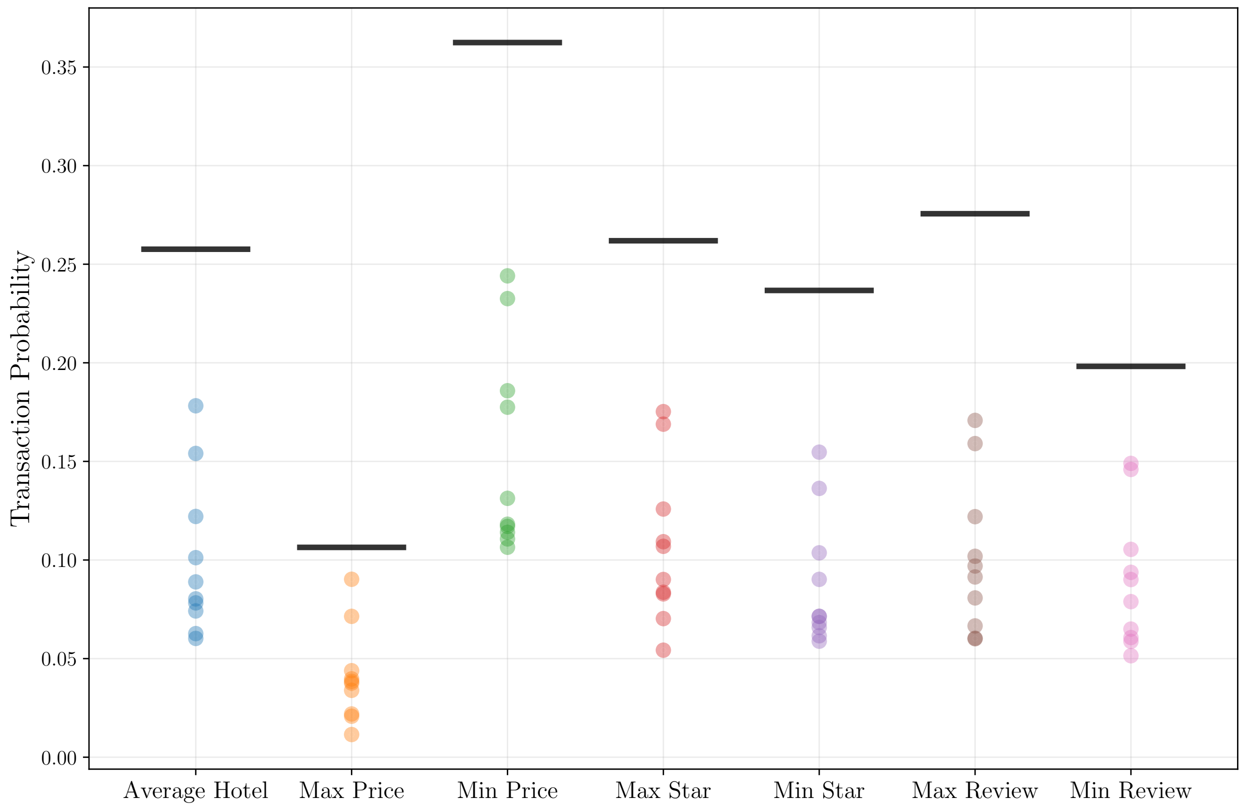


Table 7: Expedia Data: Regression of Choice Probability in Positions 1-2 on Hotel Characteristics

	(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 do 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.

One can derive the DSC model from a two-stage choice process. First, consumers decide whether to be attentive or not as a function of unobservables and the characteristics of the default good. Then, if they are attentive, they make an active choice among all goods.

5.1 Nonparametric Identification

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:

$$\pi_i\left(\mathcal{J}\right) = \phi_i(x_{id}) \tag{60}$$

$$\pi_i(d) = 1 - \phi_i(x_{id}) \tag{61}$$

$$\pi_i(c) = 0 \quad \text{for c} \notin \{\mathcal{J}, d\} \tag{62}$$

The probability of selecting option j is expressed as:

$$s_{ij} = (1 - \phi_i)\, 1\left(j = d\right) + \phi_i s^{\star}_{ij} \tag{63}$$

where $s^{\star}_{ij}$ denotes the probability of choosing j conditional on considering all available goods.

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

$$\boxed{\begin{aligned} \frac{\partial \log\left(\phi_i\right)}{\partial x^1_{id}} &= \frac{1}{s_{ij}}\left[\frac{\partial s_{ij}}{\partial x^1_{id}} - \frac{\partial s_{id}}{\partial x^1_{ij}}\right] && (64)\\ \phi_i &= \exp\left(-\int \frac{1}{s_{ij}}\left[\frac{\partial s_{ij}}{\partial x^1_{id}} - \frac{\partial s_{id}}{\partial x^1_{ij}}\right] dx^1_{id}\right) && (65)\end{aligned}}$$

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 the DSC model.

5.2 Parametric Representation

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

$$x_{id}\beta + \epsilon_{id} > f(z_i) + v_i \tag{66}$$

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. Then the probability of being inattentive is:

$$1 - \phi_i = \frac{\exp(x_{id}\beta)}{\exp(f(z_i)) + \exp(x_{id}\beta)} \tag{67}$$

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 this model by maximum likelihood.

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:

$$u_{ij} = x_{ij}\beta + \xi_{i,j=d} + \psi_{i,j=d} + \epsilon_{ij} \tag{68}$$

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:

$$\begin{aligned} \psi_{j=d} &= \ln\left(\frac{s_{id}}{1-s_{id}}\right) - \ln\left(\frac{s^\star_{id}}{1-s^\star_{id}}\right) \\ &= \ln\left(\frac{1 + (1-\pi_i)\sum_{k\neq d}\exp((x_{ik} - x_{id})\beta)}{\pi_i}\right) \end{aligned} \tag{69}$$

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.[12] 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 last term of equation 69 shows that $\psi_{i,j=d}$ can be written explicitly as a monotonic transformation of observed choice probabilities minus what choice probabilities would

[12] The inertia term ξ is not formally identical to a model where error terms are correlated over time. Abaluck and Gruber (2016) gives one example of a model which allows for both possibilities in the empirical setting of health plan choice we consider below. Nonetheless, in a model that does not allow for such a correlation, the ξ term may proxy for it.

be were consumers fully attentive.[13]

Suppose we observe that beneficiaries are inertial and want to know – is this because of adjustment costs or inattention? If it is because of adjustment costs ($\psi_i = 0$) then consumers will readily switch to alternative options if those options become more desirable. If it is because of inattention, consumers will be insensitive to characteristics of alternative plans. They may however still switch in response to changes in the characteristics of the default plan because these changes are allowed to impact the degree of inattention. Thus, in the health plan context, we may see the characteristic pattern documented in Ho, Hogan, and Scott Morton (2015) wherein consumers readily switch when the premiums of their prior year plan increase but do not switch when the premiums of alternative plans fall. Provided we observe enough determinants of attentive behavior (which again, need not include any characteristics beyond those which enter utility), we can separately identify ξ and ψ based on the asymmetry in how the share of rival plans responds to changes in the characteristics of the default good relative to how the share of the default good responds to changes in the characteristics of rival plans.

5.3 Overidentification Tests

This paper has shown that consideration probabilities are identified from cross-derivative asymmetries. However, this naturally raises the question – can such asymmetries arise for reasons other than a failure to consider all of possible choices? As discussed in Section 3, the consideration set models we consider are overidentified. This means that we can test whether the patterns observed in the cross-derivatives are consistent with the underlying model of consideration using only a subset of the observed cross-derivatives.

In this section, we provide a way of implementing these overidentification tests which allows one to visualize directly whether the model-predicted patterns of cross-derivative asymmetry hold in the data. In the general formulation of the DSC model, cross-derivative asymmetries are given by:

$$\frac{\partial s_{ij}}{\partial x_{id}^1} - \frac{\partial s_{id}}{\partial x_{ij}^1} = \frac{\partial \log(\pi_i)}{\partial x_{id}^1} s_{ij} \tag{70}$$

The model thus predicts that cross-derivative symmetries should scale across goods with observed market shares. Intuitively, asymmetries should be especially large for goods that have a high likelihood of being bought when they are considered – if an alternative health plan is especially attractive, while you might not notice if it cuts its price by \$100, you will have a high probability of switching to it if the plan you chose last year raises its price by \$100. Note that this prediction follows directly from the structure of the model rather than, for example, particular functional form assumptions on the distribution of error terms.

[13]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 $s_{ij}^{\star}$ 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 $s_{id}^{\star}$. 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.

In the formulation where ϕ_i is a logit function of the characteristics of the default good, we have:

$$\frac{\partial s_{ij}}{\partial x_{id}} - \frac{\partial s_{id}}{\partial x_{ij}} = \gamma(1-\phi_i)s_{ij} \tag{71}$$

We can test whether the cross-derivatives have this characteristic form. Since any specification test will reject with a sufficient quantity of data, a more relevant standard may be to assess graphically whether the cross-derivative patterns are well explained by equation 71 or whether we observe substantial deviations. To do so, we compare the predictions of the DSC model to an alternative discrete choice model that nests the DSC model and also includes flexible interaction terms between the characteristics of each good and the default good in order to flexibly model cross-elasticities.

More precisely, the test we propose is to first estimate the DSC model to recover $\hat{\beta}_i$, $\hat{\xi}_{i,j=d}$ and $\hat{\psi}_{ij}$, then estimate:

$$\begin{aligned} u_{id} &= A_{id} + \epsilon_{id} \\ u_{ij} &= A_{ij} + \sum_k \sum_{k'} x_{idk} x_{ijk'} \alpha_{k,k'} + \epsilon_{ij} \text{ for } j \neq d \end{aligned} \tag{72}$$

where $A_{ij} = x_{ij}\hat{\beta} + \hat{\xi}_{i,j=d} + \hat{\psi}_{i,j=d}$ is the predicted component of utility from the DSC model which 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 C, we derive an analytical expression for $\frac{\partial s_{ij}}{\partial x_{id}} - \frac{\partial s_{id}}{\partial x_{ij}}$ which we use to apply this overidentification test in Section 5.4.

5.4 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 plan (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 rival plans would save money. We estimate the DSC model described in Section 5.2 to separately identify an inertial term $\xi_{j=d}$ and an adjustment cost term. In addition, we perform the overidentification test described in Section 5.3 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 rely on assumptions that some variables impact attention and not utility. We instead rely 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.

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 8. We report the mean and standard deviation of a variety of characteristics for all plans and also for chosen plans.

Table 8: Part D Data: Summary Statistics

	A. All Plans	*B. Chosen Plans*
Annual Premium (dollars)	493	423
	(242)	(199)
Annual Out of Pocket Costs (dollars)	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).

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 E, 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 9 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 it in dollar terms).

The attentive logit model coefficients have (mostly) the same sign as the conditional logit coefficients 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.[14]

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%. Still, 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

[14] 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.

of $0 (the observed degree of inertia is almost fully explained by inattention),[15] 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.

Table 9: 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.

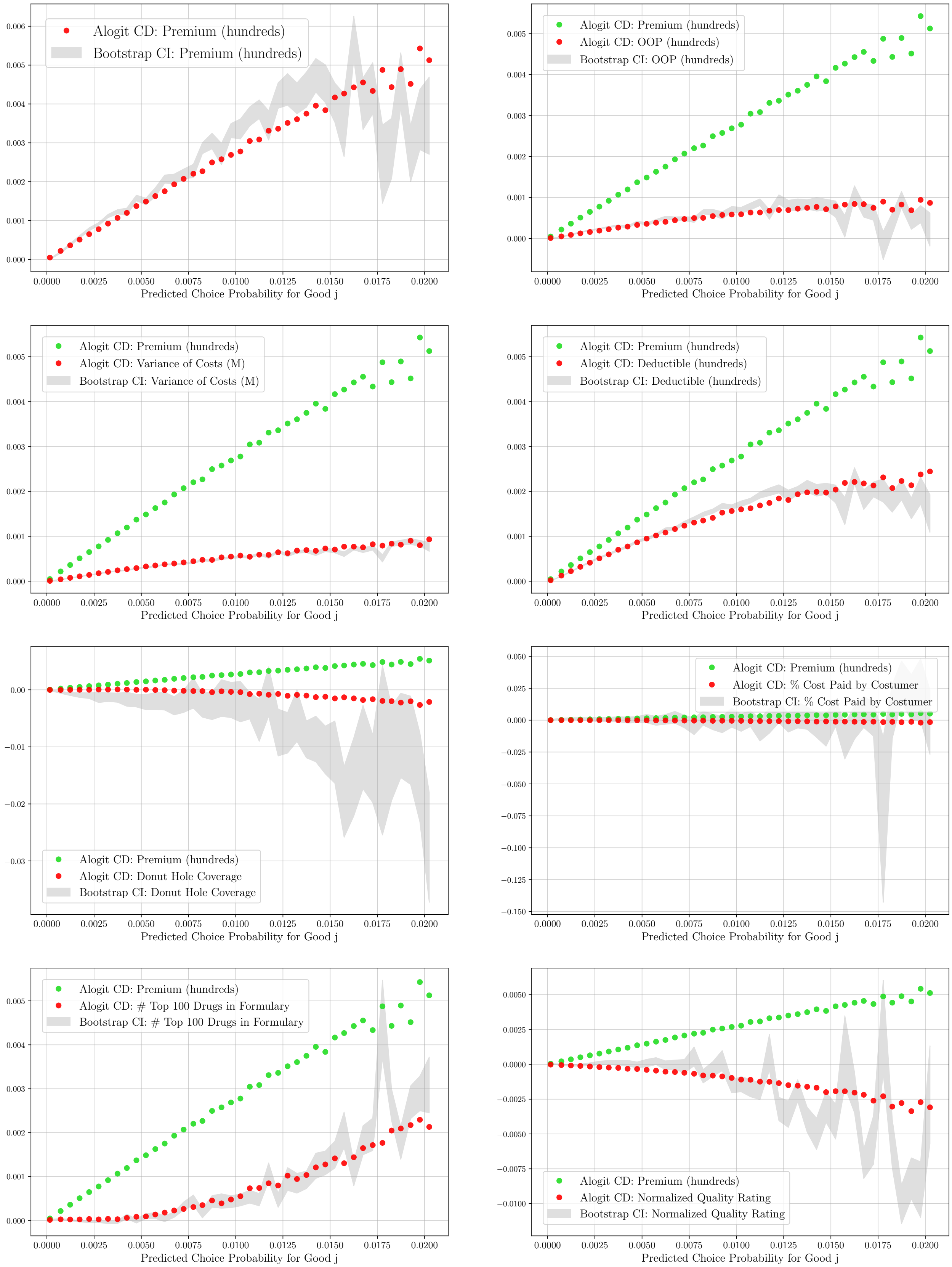
[15] In the attentive logit model, this shows up as a large negative and imprecisely estimated adjustment costs term. This is because, conditional on inattention fully explaining the observed degree of inertia, the model cannot distinguish between adjustment costs of zero and adjustment costs of negative 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.

The above results suggest that a model that permits asymmetries in demand responses fits the data better than one that does not. But do we see the specific pattern of asymmetries which is characteristic of inattention? To test this, we implement the overidentification test outlined in Section 5.3. We estimate a model which flexibly parametrizes the asymmetries in the data and compare the results to the more parsimonious attentive logit model.

Figure 6 gives the predicted cross derivative difference between default and non-default goods for included plan characteristics. We graph both the estimated cross derivatives from Equation 72 and the cross-derivatives implied by the DSC model ($\gamma_k(1-\phi_i)\hat{s}_{ij}$) 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 72 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 72. 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 6: 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 by either 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 can be constructively recovered 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 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 results impacts attention and not utility, and the model can 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 a model of inattention, and we show that this implies that, while most inertia is driven by inattention, there remain non-trivial adjustment costs.

Our identification result highlights one principle motivation for consideration sets – there are patterns of substitution which are not permitted by conventional models which are permitted in consideration set models. There may be large asymmetries in cross-derivatives with respect to some characteristics and these asymmetries may lead other parameters to be misspecified. Models estimated in practice – such as random coefficients logit models – also often impose strong assumptions on how own-price elasticities vary across goods. While these assumptions could in principle be relaxed by, for example, allowing for heteroscedasticity in the idiosyncratic error term, consideration set models relax these assumptions in a particularly parsimonious way while making systematic and testable restrictions on how asymmetries in the cross-derivatives vary across characteristics for different goods.

A second motivation for consideration set models is theoretical. We prefer random utility models to purely statistical models of choice because we believe the assumption of consumer optimization adds power and interpretability. But the assumption of perfect consideration is often suspect, and it may be more plausible to assume that consumers are optimizing given only a subset of the information available to the econometrician. Consideration set models are one way of formalizing this notion.

We hope that the results in this paper will make it possible to adapt consideration set models to a wider range of settings than they have previously been applied. These models enable us to consider counterfactuals and normative assumptions which are not possible in conventional models. We can ask, 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.

Adams, A., R. Blundell, M. Browning, and I. Crawford (2015). Prices versus preferences: taste change and revealed preference. Technical report, IFS Working Papers.

Beffy, M., R. Blundell, A. Bozio, and G. Laroque (2014). Labour supply and taxation with restricted choices. Technical report, IFS Working Papers.

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.

Blundell, R., M. Browning, and I. Crawford (2008). Best nonparametric bounds on demand responses. *Econometrica 76*(6), 1227–1262.

Browning, M. and P.-A. Chiappori (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, 1241–1278.

Cabral, L. and A. Fishman (2012). Business as usual: A consumer search theory of sticky prices and asymmetric price adjustment. *International Journal of Industrial Organization 30*(4), 371–376.

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.

Compiani, G. and Y. Kitamura (2016). Using mixtures in econometric models: a brief review and some new results. *The Econometrics Journal 19*(3).

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.

Cross, P. J. and C. F. Manski (2002). Regressions, short and long. *Econometrica 70*(1), 357–368.

Eliaz, K. and R. Spiegler (2011). Consideration sets and competitive marketing. *The Review of Economic Studies 78*(1), 235–262.

Fox, J. T., D. H. Hsu, and C. Yang (2012). Unobserved heterogeneity in matching games with an application to venture capital. Technical report, National Bureau of Economic Research.

Gabaix, X. (2011). A sparsity-based model of bounded rationality. Technical report, National Bureau of Economic Research.

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.

Hastings, J. S., A. Hortaçsu, and C. Syverson (2013). Sales force and competition in financial product markets: The case of mexicos social security privatization. Technical report, National Bureau of Economic Research.

Hastings, J. S. and J. M. Shapiro (2013). Fungibility and consumer choice: Evidence from commodity price shocks. *The quarterly journal of economics*, qjt018.

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

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

Hausman, J. A. and W. K. Newey (2016). Individual heterogeneity and average welfare. *Econometrica 84*(3), 1225–1248.

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.

Lewbel, A. and X. Tang (2015). Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics 189*(1), 229–244.

Magnac, T. and E. Maurin (2007). Identification and information in monotone binary models. *Journal of Econometrics 139*(1), 76–104.

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.

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.

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

Taubinsky, D. and A. Rees-Jones (2016). Attention variation and welfare: Theory and evidence from a tax salience experiment. Technical report, National Bureau of Economic Research.

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

A Model & Identification Proof

A.1 Results to Complement Section 3

Corollary 1. *Symmetry of Cross Derivatives*:

$$\frac{\partial s^{\star}_{ij}}{\partial x^1_{ij'}} = \frac{\partial s^{\star}_{ij'}}{\partial x^1_{ij}} \tag{73}$$

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

$$\frac{\partial s^{\star}_{ij}}{\partial x^1_{ij'}} = \int \beta_i \int ... \left[\int_{-\infty}^{v_{ij}+e-v_{ij}}\right] ... \left[\int_{-\infty}^{v_{ij}+e-v_{ij'}}\right] ... \int_{-\infty}^{v_{ij}+e-v_{iJ}} f(z_0, ..., e, ..., v_{ij}+e-v_{ij'}, ..., z_J)\mathrm{d}z_J...[\mathrm{d}z'_j]...[\mathrm{d}z_j]...\mathrm{d}z_0\mathrm{d}e\ \mathrm{d}F(\beta_i) \tag{74}$$

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

$$\begin{aligned}\frac{\partial s^{\star}_{ij}}{\partial x^1_{ij'}} &= \int \beta_i \int \int_{-\infty}^{v_{ij'}+t-v_{i0}} ... \left[\int_{-\infty}^{v_{ij'}+t-v_{ij'}}\right] ... \left[\int_{-\infty}^{v_{ij'}+t-v_{ij}}\right] ... \int_{-\infty}^{v_{ij'}+t-v_{iJ}} \\ & \quad f(z_0, ..., v_{ij'}+t-v_{ij}, ..., t, ..., z_J)\mathrm{d}z_J...[\mathrm{d}z'_j]...[\mathrm{d}z_j]...\mathrm{d}z_0\mathrm{d}t\ \mathrm{d}F(\beta_i) \\ &= \frac{\partial s^{\star}_{ij'}}{\partial x^1_{ij}}\end{aligned} \tag{75}$$

Corollary 2. *Absence of Nominal Illusion*:

$$\begin{aligned}s^{\star}_{ij}(x^1_{ij}, x^2_{ij}) &= Pr\left(v_{ij} + \epsilon_{ij} = \max_{j'\in\{0,...,J\}} v_{ij'} + \epsilon_{ij'}\right) &(76)\\ &= Pr\left(\beta_i x^1_{ij} + w_{ij}(x^2_{ij}) + \epsilon_{ij} = \max_{j'\in\{0,...,J\}} \beta_i x^1_{ij'} + w_{ij}(x^2_{ij'}) + \epsilon_{ij'}\right) &(77)\\ &= Pr\left(\beta_i x^1_{ij} + \beta_i\delta + w_{ij}(x^2_{ij}) + \epsilon_{ij} = \max_{j'\in\{0,...,J\}} \beta_i x^1_{ij'} + \beta_i\delta + w_{ij}(x^2_{ij'}) + \epsilon_{ij'}\right) &(78)\\ &= Pr\left(\beta_i\left(x^1_{ij}+\delta\right) + w_{ij}(x^2_{ij}) + \epsilon_{ij} = \max_{j'\in\{0,...,J\}} \beta_i\left(x^1_{ij'}+\delta\right) + w_{ij}(x^2_{ij'}) + \epsilon_{ij'}\right) &(79)\\ &= s^{\star}_{ij}(x^1_{ij}+\delta, x^2_{ij}) &(80)\end{aligned}$$

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

$$\mathbb{P}(j,j') = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad j' \in c\}, \tag{81}$$

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

$$\frac{\partial s_{ij}}{\partial x^1_{ij'}} - \frac{\partial s_{ij'}}{\partial x^1_{ij}} = \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}}{\partial x^1_{ij'}} s^\star_{ij}(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}}{\partial x^1_{ij}} s^\star_{ij'}(c') + \sum_{c'' \in \mathbb{P}(j,j')} \pi_{ic''} \left(\frac{\partial s^\star_{ij}(c'')}{\partial x^1_{ij'}} - \frac{\partial s^\star_{ij'}(c'')}{\partial x^1_{ij}} \right) \quad (82)$$

$$= \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_{ic}}{\partial x^1_{ij'}} s^\star_{ij}(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{ic'}}{\partial x^1_{ij}} s^\star_{ij'}(c') \quad (83)$$

$$\neq 0 \quad \text{for } \pi_i(\mathcal{J}) < 1 \quad (84)$$

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 $\delta \neq 0$,

$$s_{ij}(x^1_i, x^2_i) = \sum_{c \in \mathbb{P}(j)} \pi_{ic}(x^1_i, x^2_i) Pr\left(v_{ij} + \epsilon_{ij} = \max_{j' \in c} v_{ij'} + \epsilon_{ij'} \right) \quad (85)$$

$$\neq \sum_{c \in \mathbb{P}(j)} \pi_{ic}(x^1_i + \delta, x^2_i) Pr\left(v_{ij} + \epsilon_{ij} = \max_{j' \in c} v_{ij'} + \epsilon_{ij'} \right) \quad (86)$$

$$= s_{ij}(x^1_i + \delta, x^2_i) \quad \text{for } \pi_i(\mathcal{J}) < 1 \quad (87)$$

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:

$$\frac{1}{2} J(J+1) \geq J + 1 \quad (88)$$

$$J \geq 2 \quad (89)$$

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:

$$s_{ij}(\mathcal{J}) \neq s_{ij'}(\mathcal{J}) \quad (90)$$

$$\frac{s_{il}(\mathcal{J}) - s_{il}(\mathcal{J}/j)}{s_{ij'}(\mathcal{J}) - s_{ij'}(\mathcal{J}/j)} \neq \frac{s_{il}(\mathcal{J}) - s_{il}(\mathcal{J}/j')}{s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')} \quad (91)$$

$$s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j') \neq 0 \quad (92)$$

for all $j, j', l \in \mathcal{J}$ with $j, j' > 0$. Equation 92 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 91 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:

$$\begin{bmatrix} -(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1)) & 0 & s_{i1}(\mathcal{J}) \\ 0 & -(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2)) & s_{i2}(\mathcal{J}) \\ -(s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)) & (s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \phi_{i1}}{\partial x_{i1}} \\ \frac{\partial \phi_{i2}}{\partial x_{i2}} \\ \frac{\partial \phi_{i0}}{\partial x_{i0}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_{i1}}{\partial x_{i0}^1} - \frac{\partial s_{i0}}{\partial x_{i1}^1} \\ \frac{\partial s_{i2}}{\partial x_{i0}^1} - \frac{\partial s_{i0}}{\partial x_{i2}^1} \\ \frac{\partial s_{i1}}{\partial x_{i2}^1} - \frac{\partial s_{i2}}{\partial x_{i1}^1} \end{bmatrix} \tag{93}$$

The determinant of D_i is:

$$\det(D_i) = s_{i2}(\mathcal{J})\,(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1))\,(s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)) - s_{i1}(\mathcal{J})\,(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2))\,(s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)) \tag{94}$$

When D_i is singular:

$$\frac{1}{s_{i1}(\mathcal{J})}\left(\frac{s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/1)}{s_{i2}(\mathcal{J}) - s_{i2}(\mathcal{J}/1)}\right) = \frac{1}{s_{i2}(\mathcal{J})}\left(\frac{s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/2)}{s_{i1}(\mathcal{J}) - s_{i1}(\mathcal{J}/2)}\right) \tag{95}$$

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

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

$$\frac{\phi_{ij}(x_{ij}^1 + \delta_i)}{1 - \phi_{ij}(x_{ij}^1 + \delta_i)} \neq \frac{\phi_{ij}(x_{ij}^1 + \delta_{i'})}{1 - \phi_{ij}(x_{ij}^1 + \delta_{i'})} \tag{96}$$

$$\phi_{ij}(x_{ij}^1 + \delta_i) \neq \phi_{ij'}(x_{ij'}^1 + \delta_i) \quad \text{at, at least one } i = 1, ..., N \tag{97}$$

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:

$$\Pi(\delta) = \begin{bmatrix} \phi_{i0}(\delta_1)\phi_{i1}(\delta_1)(1 - \phi_{i2}(\delta_1)) & \phi_{i0}(\delta_1)\phi_{i1}(\delta_1)\phi_{i2}(\delta_1) & 0 & 0 \\ \phi_{i0}(\delta_2)\phi_{i1}(\delta_2)(1 - \phi_{i2}(\delta_2)) & \phi_{i0}(\delta_2)\phi_{i1}(\delta_2)\phi_{i2}(\delta_2) & 0 & 0 \\ 0 & 0 & \phi_{i0}(\delta_1)\phi_{i2}(\delta_1)(1 - \phi_{i1}(\delta_1)) & \phi_{i0}(\delta_1)\phi_{i1}(\delta_1)\phi_{i2}(\delta_1) \\ 0 & 0 & \phi_{i0}(\delta_2)\phi_{i2}(\delta_2)(1 - \phi_{i1}(\delta_2)) & \phi_{i0}(\delta_2)\phi_{i1}(\delta_2)\phi_{i2}(\delta_2) \end{bmatrix} \tag{98}$$

$$= \begin{bmatrix} \Pi_1(\delta) & 0 \\ 0 & \Pi_2(\delta) \end{bmatrix} \tag{99}$$

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

$$\det(\Pi(\delta)) = \det(\Pi_1(\delta))\det(\Pi_2(\delta)) \tag{100}$$

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

$$\frac{1-\phi_{i2}(\delta_1)}{\phi_{i2}(\delta_1)} = \frac{1-\phi_{i2}(\delta_2)}{\phi_{i2}(\delta_2)} \tag{101}$$

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

$$\frac{1-\phi_{i1}(\delta_1)}{\phi_{i1}(\delta_1)} = \frac{1-\phi_{i1}(\delta_2)}{\phi_{i1}(\delta_2)} \tag{102}$$

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:

$$s_{ij} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_{il}(x_{i0}, x_{il}) \prod_{l' \notin c} (1 - \phi_{il'}(x_{i0}, x_{il})) \tag{103}$$

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:

$$\frac{\partial s_{ij}}{\partial x_{i0}^1} - \frac{\partial s_{i0}}{\partial x_{ij}^1} = \frac{\partial \log(\phi_{ij})}{\partial x_{i0}^1} s_{ij}(\mathcal{J}) + \sum_{j' \neq \{j,0\}} \frac{\partial \log(\phi_{ij'})}{\partial x_{i0}^1} \left(s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/j')\right) - \frac{\partial \log(\phi_{ij})}{\partial x_{ij}^1} \left(s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)\right) \tag{104}$$

Thus there are now $2J$ derivatives of log consideration probabilities to identify: $\partial \log(\phi_{ij})/\partial x_{ij}^1$ and $\partial \log(\phi_{ij})/\partial x_{i0}^1$ 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):

$$\frac{1}{2} J(J+1) \geq 2J \tag{105}$$

$$J \geq 3 \tag{106}$$

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:

$$\frac{1}{2}J(J-1) \geq J \tag{107}$$

$$J \geq 3 \tag{108}$$

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:

$$s_{i0} = \prod_{j\in\mathcal{J}} (1-\phi_{ij}(x_{ij})) + \sum_{c\in\mathbb{P}(0)} \prod_{l\in c} \phi_{il}(x_{il}) \prod_{l'\notin c} (1-\phi_{il'}(x_{il'}))\, s^{\star}_{i0}(c) \tag{109}$$

$$s_{ij} = \sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_{il}(x_{il}) \prod_{l'\notin c} (1-\phi_{il'}(x_{il'}))\, s^{\star}_{ij}(c) \tag{110}$$

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:

$$s_{i0} = \prod_{j\in\mathcal{J}} (1-\phi_{ij}) + \sum_{c\in\mathbb{P}(0)} \prod_{l\in c} \phi_{il} \prod_{l\notin c} (1-\phi_{il'}) s^{\star}_{i0}(c) \tag{111}$$

$$s_{ij} = \sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_{il} \prod_{l\notin c} (1-\phi_{il'}) s^{\star}_{ij}(c) \tag{112}$$

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

$$\frac{\partial s_{i0}}{\partial x^1_{ij}} - \frac{\partial s_{ij}}{\partial x^1_{i0}} = \frac{\partial \log(\phi_{ij})}{\partial x^1_{ij}} (s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)) - \frac{\partial \log(\phi_{i0})}{\partial x^1_{i0}} (s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/0)) \tag{113}$$

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.

B Proof of Utility Representations for Consideration Set Models

Consider first the ASC model. Let's start by assuming there is a single plan to which you might be inattentive, plan 1 and an alternative, plan 0 to which you are attentive. Suppose that the probability of choosing good 1 is given by:

$$P(Y_{i1} = 1) \quad = \quad P(A_{i1}|x_{i1})P(Y^*_{i1} = 1) \tag{114}$$

where $P(Y^*_{i1} = 1)$ is the probability of choosing good 1 conditional on paying attention. In a logit setting, this is equivalent to a model where:

$$u_{i1} = x_{ij}\beta + \psi_{j=1} + \epsilon_{ij} \tag{115}$$

where $\psi_{j=1}$ takes the value ψ_1 for plan 1 and is 0 otherwise and ψ_1 is given by:

$$\psi_1 = \ln\left(\frac{P(A_1)\exp(x_{i0}\beta)}{(1 - P(A_1))\exp(x_{i1}\beta) + \exp(x_{i0}\beta)}\right) \tag{116}$$

We prove that an analogous result holds in a J good model. Specifically, suppose there are $J-1$ goods to which you might be inattentive and a default good 0 to which you are attentive with certainty (this is a normalization). For each of the $J-1$ goods to which you might be inattentive:

$$P(Y_{ij} = 1) \quad = \quad P(A_{ij}|x_{ij})P(Y^a_{ij} = 1) \tag{117}$$

where Y^a_{ij} is the probability of choosing good j conditional on paying attention to that good. In a logit setting, this can be represented by writing:

$$u_{ij} = x_{ij}\beta + \psi_j + \epsilon_{ij} \tag{118}$$

where ψ_j is 0 for the default plan and is given by:

$$\psi_j = \ln\left(\frac{P(A_j)\sum_{k\neq j}\exp(x_{ik}\beta + \psi_k)}{(1 - P(A_j))\exp(x_{ij}\beta) + \sum_{k\neq j}\exp(x_{ik}\beta + \psi_k)}\right) \tag{119}$$

And if you are fully attentive to good J, then the same representation holds with $\psi_J = 0$.

Let's prove this by induction. We showed above that this representation holds for a 2 plan choice set. Next, suppose it holds for a $J-1$ plan choice set. If we add a Jth plan to which you might be inattentive:

$$P(Y_{iJ} = 1) \quad = \quad P(A_{iJ}|x_{iJ})P(Y^a_{iJ} = 1) \tag{120}$$

By the inductive hypothesis, we have:

$$P(Y^a_{iJ} = 1) = \frac{\exp(x_{iJ}\beta)}{\exp(x_{iJ}\beta) + \sum_{k\neq J}\exp(x_{ik}\beta + \psi_k)} \tag{121}$$

And therefore:

$$P(Y_{iJ}=1) \quad = \quad \frac{P(A_{iJ}|x_{iJ})\exp(x_{iJ}\beta)}{\exp(x_{iJ}\beta)+\sum_{k\neq J}\exp(x_{ik}\beta+\psi_k)} \tag{122}$$

And it is straightforward to confirm that we obtain this representation if we set:

$$\psi_J = \ln\left(\frac{P(A_J)\sum_{k\neq J}\exp(x_{ik}\beta+\psi_k)}{(1-P(A_J))\exp(x_{iJ}\beta)+\sum_{k\neq J}\exp(x_{ik}\beta+\psi_k)}\right) \tag{123}$$

Next, consider the PDM.

$$\begin{aligned} P(Y_{id}=1) &= P(I_i|x_{id}) + (1-P(I_i|x_{id}))P(Y^*_{id}=1) \\ P(Y_{ij}=1) &= (1-P(I_i|x_{id}))P(Y^*_{ij}=1) \text{ for } j\neq d \end{aligned} \tag{124}$$

where $Y^*_{ij}=1$ are the choices given by maximizing:

$$u^*_{ij} = x_{ij}\beta_i + \xi_{i,j=d} + \epsilon_{ij} \tag{125}$$

We want to show that this is equivalent to a model given by:

$$u_{ij} = x_{ij}\beta_i + \xi_{i,j=d} + \psi_{i,j=d} + \epsilon_{ij} \tag{126}$$

We will derive an expression for $\psi_{i,j=d}$ in the case where the ϵ_{ij} are i.i.d. type I extreme value. Let ψ_i and ξ_i denote the values of $\psi_{i,j=d}$ and $\xi_{i,j=d}$ when $j=d$. In this case, the probability of choosing the default plan is given by:

$$\begin{aligned} P(Y^*_{id}=1) &= \int \frac{\exp(x_{id}\beta_i+\xi_i)}{\sum_k \exp(u_{ik})} f(\beta,\xi)d\beta d\xi \\ &= \int \frac{1}{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)} f(\beta,\xi)d\beta d\xi \end{aligned} \tag{127}$$

We want to solve for ψ_i satisfying:

$$\begin{aligned} \int \frac{1}{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i-\psi_i)} f(\beta,\xi)d\beta d\xi = \\ P(I_i|x_{id}) + (1-P(I_i|x_{id}))\int \frac{1}{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)} f(\beta,\xi)d\beta d\xi \end{aligned} \tag{128}$$

We can rewrite the 2nd term as:

$$\int \frac{P(I_i|x_{id})(1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)+(1-P(I_i|x_{id}))}{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)} f(\beta,\xi)d\beta d\xi \tag{129}$$

Thus, our problem reduces to finding ψ_i which satisfies:

$$\frac{1}{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i-\psi_i)} = \frac{P(I_i|x_{id})(1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)+(1-P(I_i|x_{id}))}{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)} \tag{130}$$

Taking reciprocals of both sides yields:

$$1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i-\psi_i) = \frac{1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)}{P(I_i|x_{id})(1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)+(1-P(I_i|x_{id}))} \tag{131}$$

Subtracting 1 gives:

$$\exp(-\psi_i)\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i) = \frac{(1-P(I_i|x_{id}))\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)}{P(I_i|x_{id})(1+\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)+(1-P(I_i|x_{id}))} \tag{132}$$

And so:

$$\psi_i = \ln\left(\frac{1+P(I_i|x_{id})\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta_i-\xi_i)}{(1-P(I_i|x_{id}))}\right) \tag{133}$$

C Overidentification Test in the HHS Model

In the HHS model:

$$\begin{aligned} s_{id} &= (1-\pi_i) + \pi_i s^*_{id} \\ s_{ij} &= \pi_i s^*_{ij} \end{aligned} \tag{134}$$

As noted in Section **??**, the HHS model with linear utility and logit errors can be written as a random utility model where:

$$u_{ij} = x_{ij}\beta + \psi_{i,j=d} + \epsilon_{ij} \tag{135}$$

where $\psi_{i,j=d}$ is the consideration term which may vary as a function of own and rival characteristics. and is given by: where:

$$\begin{aligned} \psi_{j=d} &= \ln\left(\frac{1+(1-\pi_i)\sum_{k\neq d}\exp((x_{ik}-x_{id})\beta)}{\pi_i}\right) \\ &= \ln\left(\frac{(1-\pi_i)+\pi_i s^*_{id}}{\pi_i s^*_{id}}\right) \end{aligned} \tag{136}$$

The test we propose is to first estimate the HHS model to recover $\hat{\beta}_i$ and $\hat{\psi}_{ij}$, then estimate:

$$\begin{aligned} u_{id} &= A_{id} + \epsilon_{id} \\ u_{ij} &= A_{ij} + \sum_k \sum_{k'} x_{idk} x_{ijk'} \alpha_{k,k'} + \epsilon_{ij} \text{ for } j \neq d \end{aligned} \tag{137}$$

where $A_{ij} = x_{ij}\hat{\beta} + \hat{\psi}_{i,j=d}$ is the predicted component of utility from the HHS model which is now regarded as a known constant. Let $v_{id} = A_{id}$ and $v_{ij} = A_{ij} + \sum_k \sum_{k'} x_{idk} x_{ijk'} \alpha_{k,k'}$. Let $q_{ij} = \sum_k \sum_{k'} x_{idk} x_{ijk'} \alpha_{k,k'}$. Assume logit errors. Then:

$$\begin{aligned} \frac{\partial s_{id}}{\partial x_{ijk}} &= \frac{\partial s_{id}}{\partial v_{id}}\frac{\partial A_{id}}{\partial x_{ijk}} + \frac{\partial s_{id}}{\partial v_{ij}}\beta_k + \sum_{l\neq d}\frac{\partial s_{id}}{\partial v_{il}}\frac{\partial q_{il}}{\partial x_{ijk}} \\ \frac{\partial s_{ij}}{\partial x_{idk}} &= \sum_{l\neq d}\frac{\partial s_{ij}}{\partial v_{il}}\frac{\partial q_{il}}{\partial x_{idk}} + \frac{\partial s_{ij}}{\partial v_{id}}\frac{\partial A_{id}}{\partial x_{idk}} \end{aligned} \tag{138}$$

With $\alpha_{k,k'} = 0$, these equations hold for the estimated alogit parameters. Let $\hat{s}^*_{ij}$ denote the estimated choice probabilities conditional on paying attention. From the choice probability definitions, we have: $\frac{\partial s_{id}}{\partial x_{ijk}} = \pi_i \frac{\partial \hat{s}^*_{id}}{\partial x_{ijk}} = -\beta_k \pi_i \hat{s}^*_{id}\hat{s}^*_{ij}$ and $\frac{\partial s_{ij}}{\partial x_{idk}} = \pi_i \frac{\partial \hat{s}^*_{ij}}{\partial x_{idk}} + s^*_{ij}\frac{\partial \pi_i}{\partial x_{idk}} = -\beta_k \pi_i \hat{s}^*_{id}\hat{s}^*_{ij} + \gamma_k \hat{s}_{ij}(1-\pi_i)$. Combining these with the above equations gives:

$$\begin{aligned} -\beta_k \pi_i \hat{s}^*_{id}\hat{s}^*_{ij} &= \hat{s}_{id}(1-\hat{s}_{id})\frac{\partial A_{id}}{\partial x_{ijk}} - \hat{s}_{id}\hat{s}_{ij}\beta_k \\ -\beta_k \pi_i \hat{s}^*_{id}\hat{s}^*_{ij} + \gamma_k \hat{s}_{ij}(1-\pi_i) &= -\hat{s}_{id}\hat{s}_{ij}\frac{\partial A_{id}}{\partial x_{idk}} \end{aligned} \tag{139}$$

Rearranging gives:

$$\begin{aligned}\frac{\partial A_{id}}{\partial x_{ijk}} &= \frac{\beta_k \hat{s}_{ij}(\hat{s}_{id}-\hat{s}^*_{id})}{\hat{s}_{id}(1-\hat{s}_{id})}\\ \frac{\partial A_{id}}{\partial x_{ijk}} &= \frac{\beta_k \hat{s}^*_{id}-\gamma_k(1-\pi_i)}{\hat{s}_{id}}\end{aligned} \tag{140}$$

Moving on to the "α" terms, we have:

$$\begin{aligned}\sum_{l\neq d}\frac{\partial s_{id}}{\partial v_{il}}\frac{\partial q_{il}}{\partial x_{ijk}} &= \frac{\partial s_{id}}{\partial v_{ij}}\frac{\partial q_{ij}}{\partial x_{ijk}} + \sum_{l\neq j,d}\frac{\partial s_{id}}{\partial v_{ij}}\frac{\partial q_{il}}{\partial x_{ijk}}\\ &= -\hat{\hat{s}}_{id}\hat{\hat{s}}_{ij}\sum_{k'} x_{idk'}\alpha_{k',k}\\ \sum_{l\neq d}\frac{\partial s_{ij}}{\partial v_{il}}\frac{\partial q_{il}}{\partial x_{idk}} &= \frac{\partial s_{ij}}{\partial v_{ij}}\frac{\partial q_{ij}}{\partial x_{idk}} + \sum_{l\neq j,d}\frac{\partial s_{ij}}{\partial v_{il}}\frac{\partial q_{il}}{\partial x_{idk}}\\ &= \hat{\hat{s}}_{ij}(1-\hat{\hat{s}}_{ij})\sum_{k'} x_{ijk'}\alpha_{k,k'} - \sum_{l\neq d,j}\hat{\hat{s}}_{ij}\hat{\hat{s}}_{il}\sum_{k'} x_{ilk'}\alpha_{k,k'}\\ &= \hat{\hat{s}}_{ij}\sum_{k'}(x_{ijk'}-\tilde{x}_{ik'})\alpha_{k,k'}\end{aligned} \tag{141}$$

where $\tilde{x}_{ik'} = \sum_{l\neq d}\hat{\hat{s}}_{il}x_{ilk}$.

Thus,

$$\begin{aligned}\frac{\partial s_{ij}}{\partial x_{idk}} - \frac{\partial s_{id}}{\partial x_{ijk}} &= \frac{\partial A_{id}}{\partial x_{idk}}\frac{\partial s_{ij}}{\partial v_{id}} - \frac{\partial A_{id}}{\partial x_{ijk}}\frac{\partial s_{id}}{\partial v_{id}} - \frac{\partial s_{id}}{\partial v_{ij}}\beta_k\\ &\quad + \frac{\partial s_{ij}}{\partial v_{ij}}\left(\sum_{k'} x_{ijk'}\alpha_{k,k'}\right) - \frac{\partial s_{id}}{\partial v_{ij}}\left(\sum_{k'} x_{idk'}\alpha_{k',k}\right)\\ &= \frac{\beta_k\hat{s}^*_{id}-\gamma_k(1-\pi_i)}{\hat{s}_{id}}\cdot -\hat{\hat{s}}_{id}\hat{\hat{s}}_{ij} - \frac{\beta_k\hat{s}_{ij}(\hat{s}_{id}-\hat{s}^*_{id})}{\hat{s}_{id}(1-\hat{s}_{id})}\cdot\hat{\hat{s}}_{id}(1-\hat{\hat{s}}_{id}) + \hat{\hat{s}}_{id}\hat{\hat{s}}_{ij}\beta_k\\ &\quad + \hat{\hat{s}}_{ij}\left[\sum_{k'}(x_{ijk'}-\tilde{x}_{ik'})\alpha_{k,k'} + \hat{\hat{s}}_{id}\sum_{k'} x_{idk'}\alpha_{k'k}\right]\\ &= \frac{\beta_k(\hat{s}_{id}-\hat{s}^*_{id})\hat{\hat{s}}_{id}\Big((1-\hat{s}_{id})\hat{\hat{s}}_{ij} - (1-\hat{\hat{s}}_{id})\hat{s}_{ij}\Big) + \gamma_k(1-\pi_i)(1-\hat{s}_{id})\hat{\hat{s}}_{id}\hat{\hat{s}}_{ij}}{\hat{s}_{id}(1-\hat{s}_{id})}\\ &\quad + \hat{\hat{s}}_{ij}\left[\sum_{k'}(x_{ijk'}-\tilde{x}_{ik'})\alpha_{k,k'} + \hat{\hat{s}}_{id}\sum_{k'} x_{idk'}\alpha_{k'k}\right]\end{aligned} \tag{142}$$

where $\hat{\hat{s}}_{ij}$ are the predicted market shares from the 2nd stage estimation of equation 137. Note that when all $\alpha_{k,k'} = 0$, we have: $\hat{\hat{s}}_{ij} = \hat{s}_{ij}$ and this expression simplifies to: $\gamma_k(1-\pi_i)\hat{s}_{ij}$. To perform the overidentification test, we compute the expression on the RHS of equation 142 by quantile of $\hat{\hat{s}}_{ij}$.

D Estimation

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

$$P(Y_{ij} = 1|\theta) = P(c = \emptyset|\theta) \cdot \mathbb{1}_{j=d} + \sum_{c \in C} \prod_{l \in c} \phi_{il}(\theta) \prod_{k \notin c} (1 - \phi_{ik}(\theta)) P(Y_{ij} = 1|c, \theta) \tag{143}$$

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 $\sum_{c \in C} \prod_{l \in c} \phi_{il}(\theta) \prod_{k \notin c} (1 - \phi_{ik}(\theta)) P(Y_{ij} = 1|c, \theta)$ by simulating R consideration sets per individual where, for each 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:

$$\hat{P}_{ij} = \frac{1}{R} \sum_r P(Y_{ij} = 1|c_r, \theta) \tag{144}$$

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

$$\begin{aligned} \hat{P}_{ij} \quad &\to_p \quad \frac{1}{R} \sum_r \sum_{c \in C} \prod_{l \in c} \phi_{il}(\theta) \prod_{k \notin c} (1 - \phi_{ik}(\theta)) P(Y_{ij} = 1|c, \theta) \\ &= \quad \sum_{c \in C} \prod_{l \in c} \phi_{il}(\theta) \prod_{k \notin c} (1 - \phi_{ik}(\theta)) P(Y_{ij} = 1|c, \theta) \end{aligned} \tag{145}$$

This procedure would still be computationally burdensome because it would require computing $P(Y_{ij} = 1|c_r)$ 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 144 using:

$$\hat{P}_{ij} = \frac{1}{R} \sum_r \prod_{l \in c} \phi_{il} \prod_{k \notin c} (1 - \phi_{ik}) \frac{P(Y_{ij} = 1|c_0, \theta)}{\phi^0_{ir}(\theta_0)} \tag{146}$$

where $\phi^0_{ir}(\theta_0) = \prod_{l \in S_0} \phi_{il} \prod_{k \notin S_0} (1 - \phi_{ik})$ and each consideration set is sampled with probability $\phi^0_{ir}(\theta_0)$.[16]

[16] Recall that 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

$f(x)$.

E Robustness

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

ASC Robustness Table E.1 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 5 in the text.

DSC Robustness Table 10 reports coefficient estimates from the DSC model with brand fixed effects added.

Figure E.1: 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***	2.5
		(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 10: 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.