

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

Jason Abaluck and Abi Adams

*

May 5, 2018

Abstract

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

1 Introduction

Discrete choice models generally assume that consumers consider all available options when making their choices. This prevents researchers from asking many questions of interest. What factors lead consumers to become aware of more options? Will inertial consumers ‘wake up’ in response to a price increase but remain unresponsive if rivals lower prices? Which products will respond well to advertising because they have high market shares conditional on being noticed? Normatively, whether people eat the same foods and go to the same stores year after year because they like those

*Thanks to Leila Bengali and Mauricio Caceres for excellent research assistance and to Dan Akerberg, Joe Altonji, Dan Benjamin, Doug Bernheim, Steve Berry, Judy Chevalier, Aureo de Paula, Jonathan Feinstein, Jeremy Fox, Xavier Gabaix, Jonathan Gruber, Phil Haile, Erzo Luttmer, Paola Manzini, Marco Mariotti, Olivia Mitchell, Francesca Molinari, Fiona Scott Morton, Barry Nalebuff, Simon Quinn, Alan Sorensen, Joe Shapiro, K. Sudhir, Chris Taber and participants in the Heterogeneity in Supply and Demand Conference, the Roybal Annual Meeting, and seminar participants at Berkeley, Harvard, Oxford, Stanford, the University of Chicago, Wharton, and Yale. We acknowledge financial support from NIA grant number R01 AG031270 and the Economic and Social Research Council, Grant ES/N017099/1.

options or because they do not know what else exists has first-order consequences for welfare. If one can measure preferences conditional on consideration, we can assess the benefits of policies that help consumers make more considered choices.

Consideration set models are a generalization of discrete choice models that relax the assumption that individuals consider all goods. These models instead specify a probability that each subset of options is considered (Manski 1977). The approach has long been applied in the marketing literature (Hauser and Wernerfelt 1990; Shocker, Ben-Akiva, Boccara, and Nedungadi 1991) and has become increasingly popular in both theoretical and applied literatures in economics. Consideration sets might arise due to inattention or bounded rationality (Treisman and Gelade 1980), from search costs (Caplin, Dean, and Leahy 2016), or because consumers face (unobserved) constraints on what options can be chosen (Gaynor, Propper, and Seiler 2016).¹

Identification is an immediate concern in consideration set models – if changes in prices or other characteristics perturb demand, can we tell whether this impact comes via consideration or utility? The results in this paper highlight a new source of identifying variation in two widely used classes of consideration set models that have been the focus of much applied and theoretical work. In the first class of model, which we call the “Default-Specific Consideration” (DSC) model, consumers are either “asleep” and choose a default option or they “wake up” and make an active choice from all products (Ho, Hogan, and Scott Morton 2015; Hortaçsu, Madanizadeh, and Puller 2015; Heiss, McFadden, Winter, Wupperman, and Zhou 2016). In the second class of model, which we call the “Alternative-Specific Consideration” (ASC) model, each good has an independent consideration probability that depends on characteristics of the good in question (Swait and Ben-Akiva 1987; Ben-Akiva and Boccara 1995; Goeree 2008; Van Nierop, Bronnenberg, Paap, Wedel, and Franses 2010; Manzini and Mariotti 2014; Kawaguchi, Uetake, and Watanabe 2016).

Empirical models of both types have previously relied either on auxiliary data or on additional exclusion restrictions for point identification of the structural functions of interest. For example, Conlon and Mortimer (2013) assume that consideration sets are known in some periods, while Draganska and Klapper (2011) and Honka and Chintagunta (2016) use survey data on what products are and are not considered when choosing. Goeree (2008), Hortaçsu, Madanizadeh, and Puller (2015), Gaynor, Propper, and Seiler (2016), and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) assume that there exists a set of variables that impact consideration but not utility and vice versa, while Dardanoni, Manzini, Mariotti, and Tyson (2017) rule out unobserved preference

¹Given the two-stage framework in this paper, “attention”, “awareness”, and “consideration” all synonymously mean “a good is in the choice set from which consumers then maximize utility”. We assume that conditional on considering a good, one observes all of its relevant attributes. For theoretical frameworks that relax this assumption see, for example, Kőszegi and Szeidl (2012), Bordalo, Gennaioli, and Shleifer (2013), and Gabaix (2014) among others.

heterogeneity.² These exclusion restrictions are often questionable: does advertising only impact choices via informing consumers about which goods exist? (Goeree 2008; Van Nierop, Bronnenberg, Paap, Wedel, and Franses 2010). Agnosticism over what variables impact utility and which impact attention is usually associated with only partial identification of the objects of interest (Lu 2016; Barseghyan, Maura, Molinari, and Teitelbaum 2018).

We prove that the restrictions on choice probabilities imposed by economic theory are sufficient for point identification of preferences and consideration probabilities in the DSC and ASC frameworks as well as hybrid models combining features of both alternatives. Our method does not require auxiliary information on consideration sets and it allows all observables to impact both consideration and utility. We provide simple closed form expressions for consideration set probabilities in terms of differences in cross-derivatives (the discrete choice analogue of ‘Slutsky asymmetries’). Our framework subsumes many of the consideration set models in the applied literature and does not rely on assuming a particular functional form for random utility errors. We also show that ad hoc attempts to model consideration sets such as fixed effects in utility for products on different shelves or interactions between prices and such fixed effects still typically still lead to misspecified models.

Our identification result builds on the insight that imperfect consideration breaks the symmetry between cross-price responses (or more generally, cross-characteristic responses). For example, in a model with a default, symmetry would ordinarily require that switching decisions be equally responsive to an increase in the price of the default good by \$100 or a decrease in the price of all rival goods by \$100. Suppose instead that consumers will be inattentive and choose the default option unless the default good becomes sufficiently unsuitable. Now, switching decisions will be unresponsive to changes in the price of rival goods but more responsive to changes in the price of the default to the degree that these changes perturb attention. While the link between imperfect attention and Slutsky asymmetry has been discussed in the theoretical literature, notably in Gabaix (2014), this link has not previously been utilized in applied work.³

Our approach builds on a recent literature in behavioral decision theory that highlights the

² Crawford, Griffith, and Iaria (2016) show that identification of preference parameters is possible without excluding any variables from utility or consideration with panel data and logit preferences, but only if one assumes consideration sets do not change over time. Separate identification of preferences and consideration more generally requires both excluding variables from consideration that impact utility *and* excluding variables from utility that impact consideration (Compiani and Kitamura 2016). Barseghyan, Molinari, and Thirkettle (2018) show that one only requires variables that influence utility to be excluded from consideration in combination with an ‘identification at infinity’ type argument for identification of a special case of our framework.

³Chen, Levy, Ray, and Bergen (2008) also note a connection between inattention and asymmetries in theoretical models. Davis and Schiraldi (2014) provide generalizations of multinomial logit models that permit asymmetries, but they explicitly note that these models cannot be rationalized by an underlying random utility interpretation and do not attempt to use these asymmetries to identify inattention. Aguiar and Serrano (2017) use deviations from Slutsky symmetry to quantify violations of rationality but do not use these for constructive identification of behavioral phenomena.

identifying power of theoretical restrictions in the context of consideration set models (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014; Cattaneo, Ma, Masatlioglu, and Suleymanov 2018). However, identification results in this literature have thus far relied on substantial variation in product availability. For example, in a special case of our framework, Manzini and Mariotti (2014) prove that consideration probabilities and a consumer’s preference relation can be uniquely identified from individual choice data if one observes choices from every possible non-degenerate subset of feasible alternatives. These results have thus had limited impact on applied work to date.

Our results imply that, in many applications, one could estimate consideration set models rather than the conventional discrete choice models that they nest.⁴ In cross-sectional data, our results can be used to identify whether goods are demanded because they are high-utility or because they are more likely to be considered. In panel data, one can evaluate whether inertia reflects utility-relevant factors 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. We also show that the consideration set framework considered in this paper is generally over-identified, and if instruments are available, one can use our results to test the validity of additional exclusion restrictions.

Our identification proof is constructive and so, in theory, consistent nonparametric estimators could be based upon it. However, in most applications of interest, we advocate estimating parametric generalizations of conventional models. We develop an estimation strategy grounded in the Slutsky asymmetries at the heart of our paper to demonstrate the identifying power of this variation in practice. We estimate the structural consideration and preference parameters by indirect inference. That is, we specify a flexible auxiliary model that permits a very general pattern of asymmetries, and then estimate the parameters of our consideration set model to fit the asymmetries we see in the data. We also show how to use such flexible auxiliary models to test whether parameters estimated by maximum likelihood explain well the asymmetries observed in reduced form data.

We validate our approach in a lab experiment in which participants made a series of choices from (known) proper subsets of 10 possible goods. Using only data on choices and ignoring information on what items were available, matching Slutsky asymmetries enables us accurately to recover the probabilities that each good was available as well as recovering the preference parameters that we would estimate conditional on knowing which items were available. Conventional models with a comparable number of parameters misspecify own- and cross-price elasticities relative to the elasticities computed using data on which items were actually available. The average absolute error

⁴A Stata command which implements several special cases of our model is available for download as “alogit”; a User’s Guide as well as sample datasets can be downloaded at <https://sites.google.com/view/alogit/home>.

in cross-elasticities in conventional models is 2.5 times larger than in our consideration set model or 45.5 percentage points larger as a fraction of the average absolute cross-elasticity.

We apply the Default Specific Consideration model to health plan choice data from Medicare Part D. We replicate in nationally representative data the finding in Ho, Hogan, and Scott Morton (2015) that switching decisions are far more sensitive to characteristics of the default plan than characteristics of rival plans, and we show that this implies that the observed degree of inertia is largely due to inattention given the cross-derivative asymmetries this generates. While we find that most inertia is due to inattention, adjustment costs are sufficiently large that they offset the cost savings from assigning beneficiaries to the lowest cost plans. In conventional models, we estimate switching costs of \$1,000-\$1,400; after accounting for inattention, the DSC model implies utility-relevant switching costs of \$0-\$300. We also demonstrate that the specific patterns of asymmetries we observe in the data are consistent with our underlying model of inattention. The degree of inattention we estimate is consistent with that in Heiss, McFadden, Winter, Wupperman, and Zhou (2016). Nonetheless, when we directly test their identifying assumptions, we reject several: we find that changes in characteristics over time impact utility even conditional on the level of characteristics, and that preferences vary with demographics such as age and gender.

The rest of this paper proceeds as follows. In Section 2, we work through a simple example to illustrate the core insight that drives our identification result. Section 3 lays out our general model and identification result. Section 4 discusses estimation of our consideration set framework, developing an indirect inference estimator in which structural parameters are chosen to match cross-derivative asymmetries in the data. We then take two special cases of our framework to data. Section 5 validates the practical relevance of our result using data from a lab experiment to estimate the Alternative-Specific Consideration model. Section 6 applies the Default-Specific model to data from Medicare Part D to separately identify switching costs from consumer inattention. Section 7 concludes.

2 Motivating Example

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

3, where we consider a more general consideration set model.⁵

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

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

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

where s_j^* gives choice probabilities conditional on paying attention.

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

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

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

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

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

⁵In the existing literature, the model outlined here resembles those in Hortaçsu, Madanizadeh, and Puller (2015), Ho, Hogan, and Scott Morton (2015), Heiss, McFadden, Winter, Wupperman, and Zhou (2016) and which we consider more generally below.

Recovering the derivative of the attention probability identifies the level of attention up to a constant. This constant is determined by the fact that cross-derivatives are symmetric at $\mu(x_0) = 1$, a point expanded upon in Section 3. Integrating Equation 2.3 over the support of x_0 we obtain:

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

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

3 Model & Identification

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

We consider an individual i in a market m who makes a discrete choice among $J + 1$ products, $\mathcal{J} = \{0, 1, \dots, J\}$, with $J \geq 1$. Following the literature on nonparametric identification of discrete choice models (Berry and Haile 2009; Berry and Haile 2014), we will use the term “market” as synonymous with the feasible (although not necessarily considered) choice set. Thus a market might be defined according to individual, geographic or temporal variation. Each product j is characterized as a bundle of $K \geq 1$ characteristics, x_{jm} , with support $\chi \subseteq \mathbb{R}^K$. Our framework naturally incorporates additional consumer microdata, z_i , and interactions between consumer and product characteristics. However, variation in demographic characteristics is not required for our identification result and thus we will suppress the dependence of choice on z_i in what follows.

We allow for individuals to consider an (unobserved) subset of available goods when making their choice. The set of goods that a consumer considers is called the consideration set. At this point, we place no restrictions on consideration set formation except that there exists a default option, good-0, that is always considered. The default may be the ‘outside’ good (without observed characteristics) or an ‘inside’ good (with observed characteristics). Let $\mathcal{P}(\mathcal{J})$ represent the power set of goods, with any given element of $\mathcal{P}(\mathcal{J})$ indexed by C . The set of consideration sets containing good j is then

given as:

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

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

$$s_{jm} = s_j(x_m) = \sum_{C \in \mathbb{P}(j)} \pi_C(x_m) s_j^*(x_m|C) \quad (3.2)$$

where s_{jm} is the observed probability of good j being bought given observables $x_m = [x_{0m}, \dots, x_{Jm}]$, $\pi_C(x_m)$ gives the probability that the set of goods C is considered given observable characteristics, and $s_j^*(x_m|C)$ gives the probability good j is chosen from the consideration set C . As $\pi_C(x_m)$ and $s_j^*(x_m|C)$ represent proper probabilities, we have:

$$\sum_{C \in \mathcal{P}(\mathcal{J})} \pi_C(x_m) = 1 \quad , \quad \sum_{j \in C} s_j^*(x_m|C) = 1 \quad (3.3)$$

Throughout this section we treat the probability of selecting good j conditional on observables x_m , s_{jm} , as known. Loosely, we are considering a scenario in which we have enough markets and individuals that choice probabilities conditional on observables can be nonparametrically estimated.⁶ However, we do not know the extent to which observed choice probabilities reflect consideration versus preferences. The structural objects of interest are thus the consideration set probabilities, $\pi_C(x_m)$, and the unobserved latent choice probabilities, $s_j^*(x_m|C)$.⁷ We do not directly address the identification of preference parameters given knowledge of $s_j^*(x_m|C)$ nor the identification of, for example, search costs given consideration probabilities. The parameters of any utility model that are identified from choice behavior with full consideration, and the parameters of models that provide microfoundations for consideration sets given consideration probabilities will follow from our identification results. Our aim is to provide general identification results that can be tailored by applied researchers to special cases of the framework considered here.

3.1 Key Assumptions

In this subsection, we state assumptions that suffice for cross-derivative symmetry with full consideration and, thus, assumptions under which asymmetries imply imperfect consideration. Individuals

⁶With microdata in which attributes vary across individuals, saying we observe many individuals in each “market” means that we observe many individuals with each possible choice set (where choice sets are characterized by a set of alternatives and attributes for those alternatives).

⁷In practice, one rarely has enough data to nonparametrically estimate s_{jm} , but the purpose of our identification proof is to show that practically necessary functional form restrictions are not required for identification (following Berry and Haile (2014)).

make choices from any given consideration set to maximize their utility. We take a standard random utility approach, decomposing individual i 's utility from good j in market m , u_{ijm} , into a deterministic component that depends on the good's characteristics and a random error term:

$$u_{ijm} = v_{ij}(x_{jm}) + \epsilon_{ijm} \quad (3.4)$$

ASSUMPTION 1. *Additive Separability* There exists a characteristic x_{jm}^1 that is additively separable in the indirect utility function:

$$u_{ijm} = v_{ij}(x_{jm}) + \epsilon_{ijm} \quad (3.5)$$

$$= g_i(x_{jm}^1) + w_{ij}(x_{jm}^2) + \epsilon_{ijm} \quad (3.6)$$

$$= \beta_i x_{jm}^1 + w_{ij}(x_{jm}^2) + \epsilon_{ijm} \quad (3.7)$$

where $x_{jm}^2 \in \mathbb{R}^{K-1}$, $\beta_i \sim F(\beta_i)$ where β_i is independent of x_{jm} for all $j = 0, \dots, J$ and all m .

While point identification of consideration probabilities requires only additive separability (Equation 3.6), in the proof in the main text we will make the stronger assumption of quasilinearity (Equation 3.7) for expositional simplicity. The restriction that consumers value the separable characteristic equally across choices can be substantive, although it is often theoretically well-motivated as we discuss in Section 3.6.

ASSUMPTION 2. *Exogenous Characteristics*: $\epsilon_{ijm} \perp x_{j'm'}$ for $\forall i, j' \neq j$ and $m' \neq m$. 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_{jm}^1 is continuously distributed and the distribution of $x_{jm}^1 | x_{jm}^2$ has a positive density everywhere on χ for all $j \in \mathcal{J}$.

ASSUMPTION 4. $F(\epsilon_{i0m}, \dots, \epsilon_{iJm})$ is absolutely continuous with respect to the Lebesgue measure and gives rise to a density function that is everywhere positive on \mathbb{R} with $\lim_{\epsilon_{ijm} \rightarrow -\infty} F(\cdot) = 0$, i.e. there are no mass points at negative infinity.

With \square denoting exclusion, the probability that option j is chosen given consideration of the set of options C , with $j \in C$, is given by:

$$s_j^*(x_m|C) = Pr \left(v_{ij}(x_{jm}) + \epsilon_{ijm} = \max_{j' \in c} v_{ij'}(x_{j'm}) + \epsilon_{ij'm} \right) \quad (3.8)$$

Conditional on a consideration set, choice probabilities that result from maximizing a utility function satisfying our assumptions will satisfy cross-derivative symmetry and an absence of nominal illusion.

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

$$\frac{\partial s_j^*(x_m|C)}{\partial x_{j'm}^1} = \frac{\partial s_{j'}^*(x_m|C)}{\partial x_{jm}^1} \quad (3.9)$$

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

$$s_j^*(x_m^1, x_m^2|C) = s_j^*(x_m^1 + \delta, x_m^2|C) \quad (3.10)$$

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

Slutsky Asymmetries & Nominal Illusion In our baseline model, only one mechanism is available to generate cross-derivative asymmetries: imperfect consideration. Lemma 1 makes this point formally. The link between Slutsky asymmetries, nominal illusion and imperfect attention has previously been made in the theoretical literature (Gabaix 2014), although the identifying power of these violations has not yet been harnessed for empirical work. Later in this section, we describe overidentification tests to determine whether the pattern of asymmetries is consistent with our model of imperfect consideration as opposed to resulting from violations of our underlying assumptions or other behavioral anomalies.

LEMMA 1. ASYMMETRIES & NOMINAL ILLUSION IMPLY IMPERFECT CONSIDERATION.

Given Assumptions 1-4, if

$$\frac{\partial s_j(x_m)}{\partial x_{j'm}^1} \neq \frac{\partial s_{j'}(x_m)}{\partial x_{jm}^1} \quad (3.11)$$

$$s_j(x_m^1, x_m^2) \neq s_j(x_m^1 + \delta, x_m^2) \quad (3.12)$$

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

Lemma 1 leaves open the question of whether cross-derivative asymmetries and nominal illusion are also necessary conditions for imperfect consideration. Assumption 5 will imply that imperfect consideration typically yields asymmetries.⁸

ASSUMPTION 5. $\pi_C(x_m)$ is continuously differentiable for all $C \in \mathcal{P}(\mathcal{J})$ with, for $\pi_C(x_m) < 1$ and $j = 1, \dots, J$:

$$\frac{\partial \pi_C(x_m)}{\partial x_{jm}^1} \neq 0 \quad (3.13)$$

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

3.2 Consideration Set Framework

To make progress towards point identification of the structural functions of interest, we must place some additional restrictions on consideration set probabilities. If $\pi_C(x_m)$ are allowed to vary arbitrarily, then point identification of the underlying structural functions is hopeless without additional information on what consumers considered (Manzini and Mariotti 2014).⁹

The majority of consideration set models found in the applied literature to date have taken one of two forms. The ‘Default-Specific Consideration’ (DSC) model assumes the existence of an inside default good and allows the probability of considering all alternative options to vary only as a

⁸The qualifier “typically” here is formally captured in our proof by rank conditions stated in Assumption 6 which rule out knife-edge cases where the degree to which attributes perturb consideration exactly offsets in the observed choice probabilities. For example, Matejka and McKay (2014) show that when actions are homogeneous a priori and exchangeable in the decision maker’s prior, and the information strategy is time invariant, a rational inattention model provides a foundation for the multinomial logit (which yields symmetric cross-derivatives).

⁹Several papers in the literature produce partial identification results in more general cases, such as Masatlioglu, Nakajima, and Ozbay (2012), Cattaneo, Ma, Masatlioglu, and Suleymanov (2018), and Barseghyan, Maura, Molinari, and Teitelbaum (2018).

function of the characteristics of that default (Ho, Hogan, and Scott Morton 2015; Heiss, McFadden, Winter, Wupperman, and Zhou 2016; Hortaçsu, Madanizadeh, and Puller 2015). This model can be straightforwardly microfounded in a rational inattention framework: only if the characteristics of the default get sufficiently bad do consumers pay a cost to search among all available products. For simplicity, we assume in the text a homogenous default to avoid introducing an i subscript. In Appendix A, we show that our results extend without complication to the case with heterogeneous defaults across consumers. Under this approach, the market shares of the default (good 0) and non-default goods take the form:

$$\begin{aligned} s_{0m} &= (1 - \mu_m) + \mu_m s_{0m}^*(\mathcal{J}) \\ s_{jm} &= \mu_m s_{jm}^*(\mathcal{J}) \quad \text{for } j > 0 \end{aligned} \tag{3.14}$$

where $\mu_m \equiv \mu(x_{0m})$ and gives the probability of considering all available products, while $s_{0m}^*(\mathcal{J}) \equiv s_0^*(x_m|\mathcal{J})$ and gives choice probabilities conditional on considering all products.

A second strand of the literature assumes that each good has an independent probability of being considered that depends on the characteristics of that good. This includes the models in Goeree (2008), Manzini and Mariotti (2014), Gaynor, Propper, and Seiler (2016), and Kawaguchi, Uetake, and Watanabe (2016) and has been a popular model in marketing for many years (Ben-Akiva and Boccara 1995; Swait and Ben-Akiva 1987; Van Nierop, Bronnenberg, Paap, Wedel, and Franses 2010). This model can be microfounded by assuming consumers conduct a “non-rivalrous” search. For example, when searching online for PCs, they might consider all of the products that meet a certain set of specifications as in Goeree (2008). Under this ‘Alternative-Specific Consideration’ (ASC) approach, consideration set probabilities take the form:

$$\pi_C(x_m) = \prod_{j \in C} \phi_j(x_{jm}) \prod_{j' \notin C} (1 - \phi_{j'}(x_{j'm})) \tag{3.15}$$

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

The ASC and DSC models impose structure on the data that we will exploit in the constructive identification proof to come. However, applied work to date has (until now) had to rely on a further set of assumptions on what variables are excluded from utility and from consideration for identification of these models. In this paper we show that these *additional* exclusion restrictions are unnecessary. In the remainder of this section we show that consideration probabilities are identified from cross-derivative asymmetries in a hybrid model subsuming the ASC and DSC models. How-

ever, if interest lies in scenarios that cannot be nested within this hybrid framework, in Appendix A we show in a more general environment that features of consideration probabilities are identified using the same methods.¹⁰

THE HYBRID CONSIDERATION SET MODEL. Let the market share of the inside default, good-0, and non-default goods take the form:

$$\begin{aligned} s_{0m} &= (1 - \mu_m) + \mu_m \sum_{C \in \mathbb{P}(0)} \prod_{l \in C} \phi_{lm} \prod_{l' \notin C} (1 - \phi_{l'm}) s_{0m}^*(C) \\ s_{jm} &= \mu_m \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_{lm} \prod_{l' \notin C} (1 - \phi_{l'm}) s_{jm}^*(C) \quad \text{for } j > 0 \end{aligned} \quad (3.16)$$

where $\phi_0(x_{0m}) = 1$ for all $x_{0m} \in \chi$.

Restricting $\phi_{jm} = 1$ for all $j > 0$ gives the DSC model. Restricting $\mu_m = 1$ 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.

3.3 Identifying 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. In the hybrid model that we consider here, the probability of considering a particular set of goods is a function of the characteristics of all products in the market, albeit in the manner constrained by our theoretical framework. In this richer model, cross derivative differences involving the default take the following form:¹¹

$$\frac{\partial s_{jm}}{\partial x_{0m}^1} - \frac{\partial s_{0m}}{\partial x_{jm}^1} = \frac{\partial \mu_m}{\partial x_{0m}^1} s_{jm} - \frac{\partial \phi_{jm}}{\partial x_{jm}^1} \mu_m \sum_{\substack{C : C \in \mathbb{P}(0) \\ \&j \notin C}} \prod_{l \in C} \phi_{lm} \prod_{l' \notin \{C, j\}} (1 - \phi_{l'm}) (s_{0m}^*(C \cup j) - s_{0m}^*(C)) \quad (3.17)$$

¹⁰More precisely, when consideration probabilities can be written as a function of good-specific indices, so $\pi_C = \pi(v_{1m}, \dots, v_{Jm})$ with $\frac{\partial v}{\partial x_{jm}^1}$ constant across goods, we can recover the v_{jm} up to a monotonic transformation. This identifies the relative impact of different characteristics on good-specific indices.

¹¹We focus on the default here only to simplify the exposition. See Equation 3.21 for the general case.

There are now two forces influencing the size of cross derivative differences: first, how much consideration probabilities are affected by changes in characteristics; second, the latent utilities of goods that affect the magnitude of $s_{jm}^*(C \cup j) - s_{jm}^*(C)$.

To make progress, first imagine that consumer choice is observed in a market where it is known that good $j' > 0$ is not available. For example, in the beer market, a local craft beer might not be available in all locations or, for health insurance, a plan introduced at time t is not available at $t - 1$. With a slight abuse of notation, let the set of consideration sets containing good $j > 0$ and not containing $j' > 0$ be given as:

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

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

$$s_{jm}(\mathcal{J}/j') = \mu_m \sum_{C \in \mathbb{P}(j/j')} \prod_{l \in C} \phi_{lm} \prod_{l' \notin \{C, j'\}} (1 - \phi_{l'm}) s_{jm}^*(C) \quad (3.19)$$

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

$$s_{jm}(\mathcal{J}) - s_{jm}(\mathcal{J}/j') = \overbrace{\mu_m \phi_{j'm}}^{\text{Probability consider } j'} \underbrace{\sum_{C \in \mathbb{P}(j/j')} \prod_{l \in C} \phi_{lm} \prod_{l' \notin \{C, j'\}} (1 - \phi_{l'm}) (s_{jm}^*(C \cup j') - s_{jm}^*(C))}_{\text{Impact of adding } j' \text{ to all consideration sets including } j} \quad (3.20)$$

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

$$\boxed{\frac{\partial s_{jm}}{\partial x_{j'm}^1} - \frac{\partial s_{j'm}}{\partial x_{jm}^1} = \frac{\partial \log(\phi_{j'm})}{\partial x_{j'm}^1} (s_{jm}(\mathcal{J}) - s_{jm}(\mathcal{J}/j')) - \frac{\partial \log(\phi_{jm})}{\partial x_{jm}^1} (s_{j'm}(\mathcal{J}) - s_{j'm}(\mathcal{J}/j))} \quad (3.21)$$

Note the power of this expression: it relates unobservable changes in consideration probabilities to observed cross derivative differences and market shares.

Cross derivatives with respect to characteristics of the default good take a slightly different form

as the default is present in all choice sets. Cross derivative differences with $j' = 0$ are given by the linear system:

$$\frac{\partial s_{jm}}{\partial x_{0m}^1} - \frac{\partial s_{0m}}{\partial x_{jm}^1} = \frac{\partial \log(\mu_m)}{\partial x_{0m}^1} s_{jm}(\mathcal{J}) - \frac{\partial \log(\phi_{jm})}{\partial x_{jm}^1} (s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/j)) \quad (3.22)$$

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

$$c_m = D_m \theta_m \quad (3.23)$$

where c_m is the vector of cross derivative differences, θ_m is the $J + 1$ -vector of log consideration probability derivatives, and D_m is the coefficient matrix of leave-one-out differences.¹² As there are typically more than $J + 1$ cross-derivative differences, it is convenient to work with the system:¹³

$$D'_m c_m = D'_m D_m \theta_m \quad (3.24)$$

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

ASSUMPTION 6. (RANK CONDITION) The matrix $D'_m D_m$ is full rank.

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

$$\theta_m = (D'_m D_m)^{-1} D'_m c_m \quad (3.25)$$

What if You Do Not Observe Leave-One-Out Variation? While the above proof relies on leave-one-out variation, the amount of choice set variation required is substantially less than earlier work (Manzini and Mariotti 2014; Masatlioglu, Nakajima, and Ozbay 2012): one only needs to observe $J + 1$ choice sets to identify all $J + 1$ derivatives of interest.¹⁴ However, this variation can

¹² See Appendix A for illustrations of the structure of these matrices.

¹³ Alternative weighting matrices, W_m , can be used: $D'_m W_m D_m$.

¹⁴ Kawaguchi, Uetake, and Watanabe (2016) show that less choice set variation is needed for identification of the Manzini and Mariotti (2014) model if a variable that influences utility is left excluded from consideration. Their condition relates the percentage change in product demand when a single product is unavailable to consideration probabilities using an identification at infinity argument. In contrast, we rely on minimal choice set variation (of the sort assumed in Kawaguchi, Uetake, and Watanabe (2016)) and no additional exclusion restrictions.

be replaced by a large support assumption on x_{jt}^1 for nonparametric identification of changes in consideration probabilities.

ASSUMPTION 7A. As $x_{jm}^1 \rightarrow -\infty$, $s_{jm} \rightarrow 0$.

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

In some scenarios one might observe only a subset of leave-one-out choice sets, for example, a scenario in which only a single product gets added or dropped in a market over the time period in which data is available. In this scenario, can still identify changes in consideration probabilities without making use of an identification at negative infinity argument if one assumes that the consideration function is invariant across products :

$$\phi_j(x_{jm}) = \phi(x_{jm}) \tag{3.26}$$

Note that this allows consideration probabilities to vary across products as a function of their characteristics. In this case, one only needs to observe a market with a single non-default good missing to identify changes in μ_m and ϕ_{jm} .

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

$$\log(\phi_{jm}) = \int \frac{\partial \log(\phi_{jm})}{\partial x_{jm}^1} dx_{jm}^1 + \alpha \tag{3.27}$$

Identifying the level of consideration requires an additional assumption to pin down the constant of integration, α . Assuming that consumers are prompted to consider good j when x_{jm}^1 reaches an extreme value enables the level of attention to be identified. However, note that importantly we do not require consumers to be paying attention with probability one to *all* goods in the market at any point. This is in contrast to approaches that assume there exist some markets in which consumers consider all goods to facilitate identification of preferences (Conlon and Mortimer 2013). We require only the much weaker assumption that there is *at least one good* which, for some values of observables, is considered.

This assumption is analogous to those made in the literature on nonparametric identification of

multinomial discrete choice models Berry and Haile (2009, Lewbel (2000), treatment effects (Heckman and Vytlacil 2005; Lewbel 2007), the identification of binary games and entry models (Tamer 2003), and the use of special regressors more generally. What distinguishes our approach from other identification results on consideration sets that use identification at infinity type arguments is that we do not require additional exclusion restrictions on top of this assumption. Further, the assumption that an option is considered once x_{jm}^1 reaches a particular level is testable in our setting by checking that cross derivative differences are symmetric at that value of the covariate.¹⁵

ASSUMPTION 7B. As $x_{jm}^1 \rightarrow \infty$, $\phi_{jm} \rightarrow 1$.

THEOREM 1. (IDENTIFICATION OF CONSIDERATION PROBABILITIES) Given Assumptions 1-7, consideration probabilities in the Hybrid Consideration Set Model, $\mu(x_{0m})$ and $\phi_j(x_{jm})$ are identified for all $j \in \mathcal{J}$.

3.4 Identifying Full Consideration Market Shares

Nominal illusion facilitates the identification of the 2^J independent latent choice probabilities, $s_{jm}^*(C)$.¹⁶ We will treat ϕ_{jm} as known in this subsection given the argument above. Imagine that $N = 2^J$ level shifts in the separable characteristic are observed. These shifts alter consideration probabilities but do not alter latent choice probabilities conditional on consideration. Let $k = 1, \dots, \kappa$ index the consideration sets of which j is a member. The probabilities of these consideration sets containing j are given as $\pi_{j1}, \dots, \pi_{j\kappa}$. For each good $j > 0$,¹⁷ define the matrices:

$$\Pi_j = \begin{bmatrix} \pi_{j1}(\delta_1) & \cdots & \pi_{j\kappa}(\delta_1) \\ \vdots & \ddots & \vdots \\ \pi_{j1}(\delta_N) & \cdots & \pi_{j\kappa}(\delta_N) \end{bmatrix} \quad (3.28)$$

$$s_{jm}^* = [s_{jm}^*(C_{j1}), \dots, s_{jm}^*(C_{j\kappa})] \quad (3.29)$$

$$s_{jm}^\delta = [s_{jm}(x_m^1 + \delta_1, x_m^2), \dots, s_{jm}(x_m^1 + \delta_N, x_m^2)] \quad (3.30)$$

¹⁵ ‘Thin set identification’ will not be a problem in our intended applications so long as x_{jm}^1 has a strictly positive probability of attaining the value at which attention is paid with probability one.

¹⁶This identification problem is analogous to the problem of identifying the ‘long’ regression. While the functions of interest are typically only partially identified without instruments (Henry, Kitamura, and Salanié 2014), we show that optimizing behavior here results in point identification of the objects of interest.

¹⁷The latent market shares of the default good are given by adding up within each consideration set.

where

$$\pi_{jC}(\delta) = \mu(x_{0m}^1 + \delta) \prod_{l \in C} \phi_l(x_{lm}^1 + \delta) \prod_{l' \notin C} (1 - \phi_{l'}(x_{l'm}^1 + \delta)) \quad (3.31)$$

with the dependence of ϕ_{jm} on x_{jm}^2 suppressed for notational simplicity and $\phi_{0m} = 1$ for all x_{0m} . Unobserved latent choice probabilities are defined as the solution to the following linear system:

$$\Pi_j s_{jm}^* = s_{jm}(\delta) \quad (3.32)$$

$$s_{jm}^* = \Pi_j^{-1} s_{jm}^\delta \quad (3.33)$$

There is a unique solution to this system, and thus all s_{jm}^* are identified, when all Π_j are full rank.

ASSUMPTION 8. (RANK CONDITION) Π_j is full rank for $j = 1, \dots, J$.

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

THEOREM 2. (IDENTIFICATION OF HYBRID CONSIDERATION SET MODEL) Given Assumptions 1-8, consideration probabilities, $\mu(x_{0m})$ and $\phi_j(x_{jm})$, and latent market shares conditional on consideration set C , $s_j^*(x_m|C)$ are identified for all $C \in \cup_{j=0}^J \mathbb{P}(j)$.

3.5 Overidentification

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

$$D_m \theta_m - c_m = 0 \quad (3.34)$$

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

$$\underbrace{\frac{1}{2}J(J+1)}_{\# \text{ Independent Cross Deriv. Diffs}} - \underbrace{(J+1)}_{\# \phi_{jm} \text{ Derivatives}} \quad (3.35)$$

overidentifying restrictions. Similar reasoning shows that there are $N - 2^J$ overidentifying restrictions for latent market shares.

3.6 Other Sources of Asymmetry

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

Income Effects and Nonlinear Price Responses The proof above assumed linearity of utility in the separable characteristic. This was required for Slutsky symmetry to hold within a consideration set at all points in the support of characteristics. If one instead allows for utility to take the form:

$$u_{ijm} = g(x_{jm}^1, z_i) + w_{ij}(x_{jm}^2, z_i) + \epsilon_{ijm} \quad (3.36)$$

where $g(\cdot)$ is a smooth function, then our proof of the identification of consideration probabilities proceeds with minimal changes. We can recover the derivative of consideration probabilities at point where we observe different goods with the same value of the separable characteristic as at these points Slutsky symmetry will hold within consideration sets:

$$\left. \frac{\partial s_{jm}^*}{\partial x_{j'm}^1} \right|_{z_i=z, x_{jm}^1=x_{j'm}^1=x^1} = \left. \frac{\partial s_{j'm}^*}{\partial x_{jm}^1} \right|_{z_i=z, x_{jm}^1=x_{j'm}^1=x^1} \quad (3.37)$$

Identification of consideration probabilities then follows analogously to Section 3.2. However, to recover latent choice probabilities using our nominal illusion argument, one must assume that $g(\cdot)$

is a known function if utility takes the form in Equation 3.36.

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

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

One robustly documented pattern is that consumers respond more to larger proportional changes in prices. This is sometimes referred to as the “Weber-Fechner law of psychophysics” (Thaler 1980). This effect could be captured by allowing indirect utility to be a nonlinear function of price and can be analyzed in an analogous manner to our discussion of income effects; utility will be symmetric conditional on two goods having the same price level. Additionally, this model predicts a different pattern of asymmetries than our consideration set model. In our consideration set model, asymmetries scale with latent utilities. In the Weber-Fechner model, cross-price asymmetries scale with the difference in prices between two goods. In Section 6, we show in one application that cross-derivatives scale in the manner predicted by our consideration set model.

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

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

$$u_{ijm} = \beta(p_d + \theta_j^p(p_{jm} - p_d)) + w_j(x_{dm} + \theta_j^x(x_{jm} - x_d)) + \epsilon_{ijm} \quad (3.38)$$

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

Discussing consumer choice, Gabaix (2014) treats m_j^p as structural parameters that are fixed independently of the realized characteristics of each good. In this model, the ratio of cross-derivatives is constant but not generally equal to 1 (implying asymmetric cross-derivatives):¹⁸

$$\frac{\frac{\partial s_{jm}^x}{\partial p_{j'm}}}{\frac{\partial s_{j'm}^x}{\partial p_{jm}}} = \frac{\theta_{j'}^p}{\theta_j^p} \neq 1 \quad \text{when } \theta_{j'} \neq \theta_j \quad (3.39)$$

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

Thus, while alternative behavioral stories can generate asymmetries, they are typically distinguishable from the models of good-specific consideration we consider here.

3.7 Extensions

Our framework implicitly assumes independence of the unobservables driving utility and attention. Our insights can be combined with nonparametric identification results for mixture models to make progress when this assumption is violated, but allowing utility and attention to have correlated unobservables requires additional exclusion restrictions. Specifically, let there be a finite set of types

¹⁸We here focus on cross-derivative ratios rather than differences as they take a particularly simple form in the Gabaix (2014) model.

$n = 1, \dots, N$ such that:

$$s_{jm} = \sum_{n=1}^N \omega^n(w) s_j^n(x_m) \quad (3.40)$$

where $\omega^n(w)$ gives the probability of an individual being of type n given covariates w and $s_j^n(x_m)$ gives the probability that a consumer of type n buys good j given characteristics x_m :

$$s_{jm}^n = \mu_m^n \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l^n \prod_{l' \notin C} (1 - \phi_{l'}^n) s_{jm}^{n*}(C) \quad (3.41)$$

For example, types might be indexed for their latent utility of a particular good such that they are: a) more likely to consider that good; b) more likely to buy the good conditional on consideration. Results on the identification of mixtures can be applied in this case. Given the exclusion restrictions embedded within Equation 3.40 (namely that contemporaneous values of good characteristics do not affect the distribution of types and that there exist variables w that influence the distribution of types but do not directly affect preferences or consideration conditional on consumer type), Compiani and Kitamura (2016) shows that the distribution of types and choice probabilities conditional on types are identified from demand data. One can then apply our results to choice probabilities conditional on a given type.

4 Estimation

In principle, nonparametric estimation of consideration probabilities and the consideration set-dependent choice probabilities is possible given market share data and application of the analogy principle. The constructive nature of our identification approach suggests nonparametric estimation could proceed by substituting nonparametric estimators of choice probabilities and cross derivatives into the population formulas.

However, in practice, the curse of dimensionality renders this approach infeasible in most applied settings of interest. Furthermore, consideration probabilities and s_{ij}^* are complicated nonlinear functions of observables, so measurement error will make estimation challenging even for low-dimensional problems. Thus, as with most practically-sized discrete choice models, parametric assumptions will be necessary for estimation. Throughout this section and our empirical applications, we will therefore assume that our consideration set model is characterized by the finite-dimensional parameter

vector $\beta \in \mathbb{R}^k$ as follows:

$$u_{ijm} = v(x_{jm}; \beta) + \epsilon_{ijm}(\beta) \quad (4.1)$$

$$\mu_m = Pr(h(x_{0m}; \beta) + \eta_{i0m}(\beta) > 0) \quad (4.2)$$

$$\phi_{jm} = Pr(g_j(x_{jm}; \beta) + \eta_{ijm}(\beta) > 0) \quad \text{for } j > 0 \quad (4.3)$$

and $\phi_{0m} = 1$.

Existing applications of consideration set models are typically estimated by maximum likelihood (Goeree 2008).¹⁹ In our setting, the principle downside of this approach is the lack of transparency regarding what variation is driving our ultimate results. Are the estimated consideration probabilities driven by the asymmetries in the choice probabilities or by parametric assumptions made in specifying the model?

A natural way to deal with this problem is to flexibly estimate the cross-derivative differences in the data and check whether they match those implied by the underlying consideration set model. In Section 6, we pursue this approach in the context of maximum likelihood estimation of the DSC model. An alternative approach, which we use in Section 5, is to estimate the structural parameters of interest directly from a flexible model of cross-derivatives via indirect inference (Smith 1993; Gourieroux, Monfort, and Renault 1993). Specifically, indirect inference involves specifying a flexible auxiliary model, estimating that model on the observational data, and then choosing structural parameters so that simulated data from the underlying structural model leads to the same auxiliary model estimates. This procedure makes the link between an auxiliary model of cross-derivatives and the structural parameters of interest direct.

Following Keane and Smith (2003), we define a flexible auxiliary model characterized by a set of a parameters θ with $a > k$. The auxiliary model can be estimated using the observed data to obtain parameter estimates $\hat{\theta}$:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(y; x, \theta) \quad (4.4)$$

where $y \equiv \{y_{ij}\}_{i=1, \dots, N}^{j=1, \dots, J}$ and $x \equiv \{x_{ij}\}_{i=1, \dots, N}^{j=1, \dots, J}$, with y_{ij} being an indicator variable for whether individual i bought option j .

Given exogenous variables x and structural parameters β , we use our consideration set model to simulate M statistically independent simulated data sets, $\{\tilde{y}_{ij}^m(\beta)\}$, $m = 1, \dots, M$, by redrawing the structural error terms ϵ_{ij} and η_{ij} from their parametric distributions. The auxiliary model is then estimated on each of the M simulated data sets to obtain a set of estimated parameter vectors

¹⁹Please see our stata command `alogit` for estimation by maximum likelihood.

$\tilde{\theta}^m(\beta)$. Formally, $\tilde{\theta}^m(\beta)$ solves:

$$\tilde{\theta}^m(\beta) \in \arg \max_{\theta} \mathcal{L}(\tilde{y}^m; x, \theta) \quad (4.5)$$

Indirect inference generates an estimate $\hat{\beta}$ of the structural parameters that minimizes the distance between the parameters of the auxiliary model estimated on the observed and simulated data. Loosely speaking, the approach harnesses the insight that if one has the right data generating process, operations performed on observed and simulated data should give the same answer. Let $\tilde{\theta}(\beta) = M^{-1} \sum \tilde{\theta}^m(\beta)$. Formally, $\hat{\beta}$ solves:

$$\hat{\beta} = \arg \min_{\beta} \left(\hat{\theta} - \tilde{\theta}(\beta) \right)' W \left(\hat{\theta} - \tilde{\theta}(\beta) \right) \quad (4.6)$$

where W is a positive definite weighting matrix. Note that the set of structural errors used to generate the simulated data sets are held fixed for different values of β . As the sample size grows large, $\hat{\theta}$ and $\tilde{\theta}(\beta)$ both converge to the same “pseudo true” value, θ_0 , underlying the consistency of the approach (Gourieroux, Monfort, and Renault 1993).²⁰

4.1 Auxiliary Model Specification

The choice of auxiliary model is, unsurprisingly, an important one for the performance of the estimator (Gourieroux, Monfort, and Renault 1993). While consistent estimation does not require the auxiliary model to provide a correct statistical description of the observed data, if it does, then indirect inference has the same asymptotic efficiency as maximum likelihood. As discussed in Bruins, Duffy, Keane, and Smith Jr (2015), the specification of the auxiliary model should balance statistical and computational efficiency; one should choose an auxiliary model that is flexible enough to give a good description of the data, whilst also being relatively quick to estimate.

Our identification proof points to the importance of specifying an auxiliary model that permits asymmetric cross elasticities. A functional form we find works well in practice (in simulations and our application below) is to specify a flexible logit model, where we begin with a conventional logit model (which imposes symmetry) and then add additional interaction terms between the characteristics of

²⁰Some computational difficulties arise from the fact that our choice variable is discrete. Therefore, as it stands our objective function is not a smooth function of our structural parameters as small changes in β result in discrete changes in our simulated data and thus auxiliary parameters. This renders standard gradient-based optimization methods unsuitable and thus we estimate the model using the Nelder-Mead simplex method (Nelder and Mead 1965).

alternative goods to allow for asymmetries. That is:

$$\tilde{u}_{ijm} = x_{ijm}\theta_j^0 + \sum_{j'} \sum_{k'} \sum_k \theta_{j,j'}^{k,k'} x_{ijm}^k x_{ij'm}^{k'} + e_{ijm} \quad (4.7)$$

where k denotes the attribute of a product.²¹ We then specify $\mathcal{L}(y; x; \theta)$ as the likelihood function for the conditional logit with the utility function \tilde{u}_{ijm} .

This specification has several desirable properties. We show in Appendix B that the consideration set model with logit utility can be rewritten as a full-consideration model where the utility of each alternative j depends directly on the attributes of rival goods.²² In the ASC case, we can derive our auxiliary model as a 2nd-order Taylor-expansion with respect to attributes of rival goods around the point where this dependence is 0 (which yields the logit choice probabilities).²³ Additionally, this specification nests the conventional logit model, which yields a symmetric substitution matrix, as a special case. Note that δ_j varies by alternative, so this specification is not as restrictive as models which assume $\delta_j = \delta$ thereby imposing that own-price elasticities must be closely tied to market shares. One could of course specify an even more flexible model (for example, by permitting all of the δ and α to be random coefficients). One reason to prefer the specification above is its computational convenience.

The representation result in Appendix B that motivates this auxiliary specification has a few other notable implications. The fact that consideration set models can be rewritten as random utility models where attributes of rival goods directly enter utility suggests a shortcoming of so-called “BLP instruments” where the exclusion of rival characteristics from the utility of each good is relied on for identification (Berry, Levinsohn, and Pakes 1995). Additionally, this representation shows that including fixed effects in a conventional model is not sufficient for consistent estimation given consideration sets. In the related literature on choice-based sampling, the econometrician sees only a subset of goods from which consumers choose. In these models, one can sometimes consistently estimate preferences by controlling for alternative-specific constant (Manski and Lerman 1977; Bierlaire, Bolduc, and McFadden 2008). In our framework, this approach does not work since such constants cannot capture the direct dependence of utility on attributes of rival goods.

²¹When estimating a DSC model, we find that it suffices to use:

$$\tilde{u}_{ijm} = x_{ijm}\theta_j^0 + \sum_k \sum_{k'} \theta_j^{k,k'} x_{ijm}^k x_{i0m}^{k'} + e_{ijm}$$

²²This is a similar insight to that pursued in Crawford, Griffith, and Iaria (2016).

²³Note that this provides one reason to prefer this auxiliary model to a flexible linear model. The flexible linear model is a Taylor expansion around a constant, whereas this model is a Taylor-expansion around the logit choice probabilities.

5 Application 1 : Experimental Validation

In this section, we investigate whether we can recover consideration probabilities from real choices with varying prices. While our proof shows that consideration probabilities can in principle be identified from observed choices, one might worry that our theoretical framework places ‘too much structure’ on the observed data. Perhaps our results are highly sensitive to a small amount of misspecification or require an unreasonable amount of data to recover accurately the structural functions of interest. Thus, rather than run a numerical simulation exercise, we evaluate the practical use of our identification result using a lab experiment in which consumers make choices from known subsets of a superset of 10 goods. We ask whether we can recover the known consideration probabilities as well as preferences conditional on consideration using information only on observed choices.

We show that we can accurately recover consideration probabilities from cross-price asymmetries. Furthermore, we recover statistically indistinguishable full-consideration own-price and cross-price elasticities from those that we estimate using all information on what options consumers considered. Some goods appear low utility in conventional models because they are rarely considered – our empirical strategy correctly recovers that they have high utility conditional on being considered.

Set-up We selected 10 goods sold at the Yale Bookstore with list prices ranging from \$19.98-\$24.98. These goods and their list prices are shown in Table 5 in Appendix C. 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. In total, we ran the experiment with 149 participants, resulting in 7,450 choices.²⁴ Prior to the experiment, participants were given several examples to illustrate the incentive scheme and were quizzed on their understanding. 70% correctly answered our test of understanding (and all participants were told why their answer was correct or incorrect). Appendix Table 7 reports results using only this subset of users who passed this test and shows that results are qualitatively unchanged.

The probability that each good appeared in a given choice set was fixed by us in advance – this

²⁴There were 150 participants in total, but one participant’s data was not recorded properly because they refreshed the browser several times during the experiment – this participant is dropped from the final analysis. When a participant refreshed the browser, the choice recorded in our data was whatever choice they made from the previous choice set. In 12 of the 7,450 remaining choices, we observe the recorded choice was not available in the choice set likely because of refreshing. We would not be able to observe cases where the browser was refreshed and last period’s choice was still available this period, but since that occurs about half the time, the total number of affected choices was likely around 25, or less than 0.35% of all choices. Dropping the cases we can identify has no impact on the results.

probability varied across goods and with prices such that goods were more likely to be considered if they had a higher price (perhaps mimicing the behavior of a retailer who places their highest margin products where they are most likely to be noticed). The probability that good j was in a participant i 's round r consideration set was specified as:²⁵

$$\phi_{ijr} = Pr(\delta_j + p_{ijr}\gamma - \eta_{ijr} > 0) \tag{5.1}$$

$$= \frac{\exp(\delta_j + p_{ijr}\gamma)}{1 + \exp(\delta_j + p_{ijr}\gamma)} \tag{5.2}$$

where η_{ijr} is distributed logistic, p_{ijr} give's the product's price, and δ_j is a product-specific fixed effect. The coefficients were chosen so that most choice sets would include between 2 and 7 products. See Table 1 for the precise coefficients. Note that in the experiment we do not specify an inside default good as in the DSC model and thus $\mu(p_{i0r}) = 1$.²⁶

To increase the likelihood that participants considered all of the products that they were presented with, we required them to spend at least 10 seconds looking at the screen before finalizing their choices. This allows us to take choices from the generated choice set as representative of consumers' true preferences. A sample product selection screen is shown in Figure 1. Consumers were shown images of all the products in their consideration 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 Here

Finally, we specify the utility that consumer i derives from product j in round r as:

$$u_{ijr} = x_{ijr}\beta + \xi_j + \epsilon_{ijr} \tag{5.3}$$

where ξ_j is a product fixed effect and ϵ_{ijr} is distributed iid Type 1 Extreme Value. Note that this is intentionally a very restrictive model of preferences which rules out heterogeneity and much else. We thus ask if we are able to recover the process generating consideration sets even when preferences are modeled in this simplistic way.

We report results from both empirical strategies presented in Section 4: parameters estimated by matching the 119 coefficients of a flexible auxiliary model that allows for asymmetries in cross-price

²⁵This is a similar empirical specification to that applied in the ASC literature to date (Goeree 2008)

²⁶However, closing the model requires us to specify a good that 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.

elasticities, and parameters estimated by maximum likelihood. The indirect inference estimation approach highlights the role of asymmetries in identifying the estimated consideration probabilities; we compare these estimates with the more conventional maximum likelihood estimates (which are implicitly identified by the same asymmetries). We also compare our results to a variety of flexible full-consideration specifications with a similar or larger number of parameters such as models with an alternative-specific price coefficient or random coefficients.

Results Table 1 compares the estimated parameters from a conditional logit model (estimated by maximum likelihood as if all 10 goods are considered) and from our consideration set model estimated by a) maximum likelihood and b) indirect inference. We also report preference parameters estimated by maximum likelihood using information on the actual choice sets that consumers faced. In contrast, our consideration set model 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. Like in any real-world setting, the underlying utility model could in reality be a nested logit, a random coefficients model, a multinomial probit or something more exotic – but our experiment shows that we can nonetheless recover the underlying attention probabilities.

Table 1 here

First, consider the maximum likelihood preference parameters shown in the top panel of Table 1 (we consider the implied elasticities below). The conditional logit model estimated based on a considered choice set of 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 consideration set model gives a value of -0.20 (0.03) – the confidence interval includes the ‘true’ value of -0.17. The conditional logit fixed effects are systematically biased because they conflate attention and utility. Products that are rarely in the choice set are assumed to be low utility. In contrast, the consideration set model recovers fixed effects consistent with those estimated using all information on what products were considered. The intervals are relatively wide, but that is a feature, not a bug relative to the conditional logit model: the consideration set model correctly recognizes that rare products are rare and that only limited information is available about how much consumers value them. The consideration set model confidence intervals on the less rare products (products 6-9 in Table 1) are reasonably precise.

The second panel of Table 1 reports the coefficients which give rise to the consideration probabilities and Figure 2 shows this information graphically. Across products, the confidence intervals on the fixed effects in the consideration equation include the true values with the exception of product 1, which lies close to the boundary of the confidence interval. We also correctly recover the impact of price on consideration. 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 consideration set model, we estimate 0.137 (.017).

Figure 2 here

Indirect Inference Table 1 also reports structural parameters estimated by indirect inference using a flexible auxiliary model that permits a broad class of asymmetric demand responses.²⁷ These estimates are slightly less precise than the maximum likelihood estimates, but otherwise share the qualitative features noted above. This approach allows us to test directly whether the ASC model can account for the patterns of asymmetries we see in the data.

Figure 3 shows the cumulative distribution of the p-values of a test of equality of the coefficients of the auxiliary model when estimated on the real data and data simulated from the consideration set model at the indirect inference parameters. A 45 degree line would indicate that any difference in these parameters is due to noise. 5.04% these tests are rejected at the 5% significance level revealing a good fit of the model to the data.

Figure 3 here

Figure 4 shows directly that different values of consideration probabilities yield very different patterns of asymmetries from what we see in the data. We first use the auxiliary model to compute cross-price elasticity asymmetries between goods 1 and 5 as the price of good 5 varies from its lowest to its highest price. This “real” cross-price elasticity (estimated from the fully flexible auxiliary model) is on the x-axis. The black diamonds plot the ASC model predicted cross-price asymmetries against these “real” elasticities. We can see that, as desired, these lie on a 45 degree line. The figure also shows that if we change the constant in the consideration probability of the ASC model, we get very different predictions for the relevant cross-derivatives that are far from the correct values. The red dots perturb the constant so that consideration probabilities are higher than we see in the data

²⁷See Section 4 and Appendix 4 for details.

(88% on average). In this case, cross-derivative asymmetries tend to be smaller in magnitude than what we see in the data. The blue dots show the model predictions when consideration is lower (26% on average). Now we see that cross-derivative asymmetries predicted by the ASC model are more pronounced than those estimated in the flexible auxiliary model.

Figure 4 here

Elasticities An additional question of interest is whether the implied price elasticities of consideration set models differ from full consideration approaches. We can compute the price elasticities given preferences estimated assuming consideration sets are known and compare them to the price elasticities implied by a variety of models. We consider a few alternatives: our consideration set model (the ASC model), a random coefficients model which allows each individual to have a separate price coefficient,²⁸ and standard conditional logit models with quadratic and alternative-specific price parameters. We compare these to the “Full Information” elasticities, where preferences are estimated using a logit model with known consideration sets and elasticities are computed given the known function relating consideration to prices (the elasticity reported is still the reduced form elasticity – how demand changes as prices change, combining the impact of prices on consideration and the impact of prices on preferences).

The ASC model has 2 price parameters (β and γ) and 20 fixed effects (one for each good in consideration and utility), the random coefficients model has 149 price parameters (one for each individual) and 10 fixed effects, the quadratic model has 2 price parameters and 10 fixed effects, and the product-specific model has 10 price parameters (one for each good) and 10 fixed effects.

Figure 5 shows the average 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 1 shows that price responses are negative as expected; the reduced form elasticities in Figure 5 are positive in some cases if the increase in consideration probability at higher prices outweighs the decrease in utility.

The logit, random coefficients and quadratic models all badly fail to characterize how elasticities vary across goods. With a separate price coefficient for each good, the product-specific model is able

²⁸ “Random coefficients” is something of a misnomer here, since the panel nature of our data allows us to estimate a separate price coefficient for each individual. This flexibility permits the substitution patterns normally allowed for in a random coefficients model.

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 full information cross-elasticities in the data is 0.090. The logit model has an average absolute deviation of 0.083, the random coefficients logit model has an average absolute deviation of 0.168, the quadratic model has an average deviation of 0.068, the product-specific model has an average deviation of 0.080, and the ASC model has an average deviation of 0.027, less than half of any of the alternative models. As a function of the full-information elasticities, the bias is on average 45.5 percentage points smaller in the ASC model than in any other model.

Figure 5 here

6 Application 2: Inertia and Inattention in Medicare Part D

In this section we present a field application of our framework to evaluate whether the observed inertia in Medicare Part D plans is due to inattention, adjustment costs or both. We also conduct reduced form analyses and overidentification tests to illustrate that the asymmetries predicted by our theoretical framework are consistent with what we see in the data.

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 that 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 stand-alone prescription drug insurance plans (PDP plans). 90% of beneficiaries of these plans choose to remain enrolled in the same plan as last year (Abaluck and Gruber 2016). An important question is whether this is because those beneficiaries would be worse off if they switched plans (because they like the plan they chose or have high adjustment costs) or because they are not paying attention and would switch if they understood that they could save money with rival plans.

In line with the existing literature on health plan choice, we assume that choice is characterized by the Default Specific Model (DSC) in which consumers are either inattentive and choose the default good or, if the default good becomes sufficiently unsuitable, they “wake up” and make an active choice from the full feasible choice set. The identification challenge is separately to identify the probability that consumers are not actively searching from (utility-driven) switching costs. Ho, Hogan, and Scott Morton (2015) assume that all inertia reflects inattention. Heiss, McFadden, Winter, Wupperman, and Zhou (2016) attempt separately to identify inattention and utility-driven switching costs using additional exclusion restrictions for identification – they impose a priori that

some variables impact attention but not utility and vice versa. We instead rely only on the asymmetry between how the market share of the default plan responds to prices of alternative plans relative to how the market shares of alternative plans respond to prices of the default plans. We show that the DSC model accounts well for the asymmetries that we observe in the data, and we use this variation to test several of the exclusion restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016).

6.1 Parametric Assumptions

In the Medicare Part D data, we observe variation across individuals and over time in defaults and product characteristics. Thus, attributes will also be indexed by i (rather than market m) and we will denote the default good with the index d . To nest the DSC model within our general framework, let $\phi_{ijt} = 1$ for all $j > 0$. The probability of considering all goods, $\mu_{it} \equiv \mu_t(x_{id})$ takes the following form:

$$\mu_t(x_{idt}) = \Pr(\delta_t + x_{idt}\gamma_t \geq \eta_{it}) \tag{6.1}$$

$$= \frac{\exp(\delta_t + x_{idt}\gamma_t)}{1 + \exp(\delta_t + x_{idt}\gamma_t)} \tag{6.2}$$

The probability of selecting option j is expressed as:

$$s_{ijt} = (1 - \mu_{it}) 1(j = d) + \mu_{it}s_t^*(x_{ijt}) \tag{6.3}$$

where s^* denotes the probability of choosing j conditional on considering all available goods.

We further assume that the utility of individual i from choosing plan j is given by:

$$u_{ijt} = x_{ijt}\beta + \xi_{j=d,t} + \epsilon_{ijt} \tag{6.4}$$

where ϵ_{ijt} are i.i.d. Type 1 Extreme Value and $\xi_{j=d,t}$ is a dummy for whether the plan is the default. The term $\xi_{j=d,t}$ represents all of the reasons why an attentive consumer might nonetheless prefer to choose the same plan – for example, because there are switching costs or persistent unobserved heterogeneity.²⁹

²⁹The distinction between switching costs and persistent unobserved heterogeneity is relevant for some counterfactuals but not others. Separately identifying these factors would be important to predicting how many consumers would switch back to their original plan were they defaulted into an alternative plan, but these need not be separated if we only want to predict whether consumers would switch were they fully informed about possible alternatives.

6.2 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 that 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 reported in Table 6 in Appendix C. We report the mean and standard deviation of a variety of characteristics for all plans and also for chosen plans.

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

6.3 Results

Reduced Form Results We first provide reduced form evidence of asymmetries in demand responsiveness to changes in the characteristics of default and non-default goods. Specifically, we

run a panel regression of an indicator for whether i switched plans in year t on attributes of the default plan and average attributes of alternative plans (with year and beneficiary fixed effects).

$$y_{it} = x_{idt}\alpha_d + \bar{x}_{ijt}\alpha_x + \delta_i + \delta_t + e_{it} \tag{6.5}$$

where y_{it} is a binary indicator for whether an individual switched from the default at t and \bar{x}_{ijt} is the average of non-default plans attributes at t .

Table 2 reports estimation results. For all the variables which have significantly different effects between the default and rival plans, we see that switching decisions are more responsive to attributes of the default, as predicted by the DSC model. An increase in premiums of \$100 for the default plan increases switching probabilities by 8.5 percentage points, while a decrease in premiums of \$100 for rival plans increases switching probabilities by 2.9 percentage points. An increase in the deductible by \$100 for the default plan likewise increases switching probabilities by 8.5 percentage points, while a decrease in the deductible of rival plans has no significant effect on switching probabilities (and we can rule out more than a 1 percentage point change). These patterns of variation are consistent with a model where many consumers do not actively search each period but can be induced to make an active choice if the default plan becomes bad enough – in this case, we will see greater responsiveness to attributes of the default plan which impact choices both via utility and via prompting consumers actively to consider other available plans.

Table 2 here.

Structural Results Table 3 shows the results of estimating a conditional logit model and the DSC model in the Part D data. The conditional logit results resemble those in Abaluck and Gruber (2016). Consumers dislike premiums and out of pocket costs, and even conditional on the out of pocket cost consequences they dislike deductibles. A few coefficients have unexpected signs relative to prior work – for example, in 2007 and 2009 consumers appear to favor plans with less favorable average cost-sharing features. Most notably for our purposes, they are willing to pay between \$1,000 and \$1,400 depending on the year to choose the same plan they chose in the previous year (obtained by dividing the coefficient on the prior year plan dummy by the coefficient on premiums to express the effect in dollar terms).

The DSC 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.³⁰

The DSC 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 DSC model implies that most of the observed degree of inertia is due to inattention, but utility-relevant factors still play a non-negligible role. The average attentive probability in the data is 11.4%, which would imply an inertia rate of 88.6% from inattention alone. The actual inertial rate is 90.74%. This implies that of the 11.4% of consumers making an active choice, almost 24% chose the default plan. Thus, the model continues to imply non-trivial adjustment costs, at least in some years. In 2007, we estimate adjustment costs of \$0 (the observed degree of inertia is almost fully explained by inattention),³¹ 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 3 here.

Asymmetries in Cross-Elasticities We now consider whether the asymmetries that we recover in the context of the DSC model are consistent with the pattern of asymmetric demand responses

³⁰ 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 DSC specification deals with this in a more principled way through an explicit model of inattention. For this reason, only the DSC 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 DSC model.

³¹In the DSC 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.

that we observe in the data. To do so we apply a strategy inspired by the indirect inference estimation approach that was outlined in Section 4. We estimate a model where the utility of good j is allowed to depend directly on attributes of the default good in a more flexible manner than that permitted by the DSC model, while nesting this framework (see Appendix D for the formal details). We then test whether the cross derivative differences implied by the more flexible model are statistically different from those estimated conditional on assuming the Default Specific Model.

Figure 6 gives the predicted cross derivative difference between default and non-default goods for included plan characteristics for four variables; the charts for all variables are in Appendix C. We graph the estimated cross derivative difference yielded by the flexible discrete choice model and by the DSC model against the predicted market share of plan j , \hat{s}_j . To capture the uncertainty in the estimated cross-derivatives, we bootstrap estimation of the flexible discrete choice model and graph the resulting confidence interval.

In all graphs, the green dots indicate the empirical cross-derivatives with respect to premiums estimated conditional on the DSC framework – 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 region indicates the 95% confidence interval on the cross-derivative difference from the more flexible specification. 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 here

Testing the Validity of Additional Exclusion Restrictions We can also rely upon our identification result to test the additional exclusion restrictions imposed in Heiss, McFadden, Winter, Wupperman, and Zhou (2016). While cross-derivative asymmetries also provide identifying power for Heiss, McFadden, Winter, Wupperman, and Zhou (2016), they impose several additional exclusion restrictions. Among others, they assume that: 1) changes in premiums, out of pocket costs and deductibles impact attention but do not impact utility conditional on paying attention, and that age, ethnicity and experience impact attention (via acuity) but not preferences directly.

We test these assumptions by estimating the DSC model and allowing each of the attributes listed

above to potentially impact both attention and utility. The coefficients relevant for these tests are reported in Table 4. Several of the exclusion restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016) are rejected. Conditional on levels, the changes in premiums and deductibles appear to impact utility directly. This might occur in a world where consumers have an extra distaste for plans that increase their premiums or deductibles – in other words, even if two plans have the same premiums this year, I am less likely to choose one which has just increased its premiums. Perhaps consumers infer from the fact that premiums used to be lower that something unobservable about the plan is worse. In any case, the fact that changes matter for utility conditional on levels violates the identifying assumption in Heiss, McFadden, Winter, Wupperman, and Zhou (2016) also made in Hortaçsu, Madanizadeh, and Puller (2015) in a different context.³² The remainder of Table 4 reports interactions between each of the preference parameters and age dummies, non-white dummies and experience dummies (experience is defined as the number of years since 2006 for which you enrolled in Part D). Heiss, McFadden, Winter, Wupperman, and Zhou (2016) assume that all of these terms are 0 – in other words, preferences are invariant to these attributes. We find several cases where this assumption is rejected – more experienced beneficiaries are more sensitive to premiums, and older and non-white beneficiaries are less sensitive to high deductibles. In Appendix Table 9, we show that, as hypothesized in Heiss, McFadden, Winter, Wupperman, and Zhou (2016), many of these variables have the predicted impact on attention. More experienced beneficiaries are more likely to be attentive in their choice – but they have systematically different preferences as well, violating the exclusion restriction.

Despite this, we find that imposing the above exclusion restrictions in the model has only a small effect on the estimated attention probability, which increases from 13.1% in the model with these additional terms allowed to impact attention but not utility to 15.4% when they are excluded from utility and thus contribute to identification. Recall that while Heiss, McFadden, Winter, Wupperman, and Zhou (2016) impose additional exclusion restrictions, they are also implicitly getting identification from the asymmetries implicit in the DSC model. Thus, while we do see violations of the additional exclusion restrictions imposed in Heiss, McFadden, Winter, Wupperman, and Zhou (2016), we find that these are not large enough to qualitatively change their results.

Table 4 here.

³²An earlier draft of this paper found smaller violations of the assumption that changes do not impact utility conditional on levels. The principle difference is that our estimates here follow Heiss, McFadden, Winter, Wupperman, and Zhou (2016) in only including in the model changes in a subset of variables – in this specification which more closely matches the original paper, we find large exclusion restriction violations.

7 Conclusion

Discrete choice models with consideration sets relax the strong assumption that beneficiaries consider all of the options available to them before making a choice. In the literature to date, such models have been identified either by bringing in auxiliary information on what options consumers consider or assuming that some characteristics impact attention or utility but not both. This paper shows that these assumptions are unnecessary. We show that a broad class of such models are identified from variation already available in the data. Consideration set probabilities are constructively identified by deviations from Slutsky symmetry, i.e. asymmetries in the matrix of cross-derivatives of choice probabilities with respect to characteristics of rival goods.

We illustrate a number of practical applications of the model. In a lab experiment, using only data on observed choices, we recover consideration probabilities and obtain accurate estimates of the elasticities that we would estimate if we observed consideration sets. In data from Medicare Part D, we validate the model further by demonstrating that the cross-derivative asymmetries follow the specific pattern predicted by our model of inattention. Our model implies that, while most inertia is driven by inattention in this market, there remain non-trivial adjustment costs. We also show that the model can be used to test identifying restrictions used in other work, many of which are rejected.

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

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

function. While it is sometimes argued that relaxing full rationality leads to a lack of discipline, we show that the consideration set models we develop are over-identified and their validity can thus be empirically determined.

While we show that deviations from Slutsky symmetry are indicative of imperfect attention in a large class of models, our constructive identification results use the additional structure imposed by the widely applied Default-Specific Consideration and Alternative-Specific Consideration frameworks. One direction for future work is to characterize more generally when consideration probabilities can be recovered from choice data. One important case are the K-rank models considered in Honka (2014) and Honka, Hortaçsu, and Vitorino (2015) in which consumers consider the K-goods which are highest ranked according to some index, thus violating the independence assumption of the ASC model. We hope that the sufficient conditions given here will make it possible to adapt consideration set models to a wider range of settings than they have previously been applied. A second extension, which we study in work in progress, is to assess whether one can identify which attributes of goods consumers attend to using only data on their choices in order to provide a general framework to identify salience effects (Kőszegi and Szeidl 2012; Bordalo, Gennaioli, and Shleifer 2013).

With additional structure, consideration set models can be used to identify parameters of interest such as search costs, and they enable us to construct counterfactuals and explore normative questions that would not be possible in conventional models. We can ask, for example, how might beneficiaries choose if they considered all available options? What is the potential value of information? When choices correlate with cognitive ability, is this because cognitive ability impacts preferences or because it impacts consumers' ability to consider all options? Do some demographic or choice set features (such as the number of plans) increase the likelihood that consumers are attentive? How much better off might consumers be if they were fully informed about the relevant choices? We hope that future work will explore these questions in more detail.

Jason Abaluck, *Yale School of Management & NBER*

Abi Adams, *University of Oxford & Institute for Fiscal Studies*

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.
- Aguiar, V. H. and R. Serrano (2017). Slutsky matrix norms: The size, classification, and comparative statics of bounded rationality. *Journal of Economic Theory* 172, 163–201.
- Barseghyan, L., C. Maura, F. Molinari, and J. Teitelbaum (2018). Heterogeneous consideration sets and preferences.
- Barseghyan, L., F. Molinari, and M. Thirkettle (2018). Discrete choice under risk with limited consideration.
- Ben-Akiva, M. and B. Boccara (1995). Discrete choice models with latent choice sets. *International journal of Research in Marketing* 12(1), 9–24.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.
- 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.
- Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica* 82(5), 1749–1797.
- Bierlaire, M., D. Bolduc, and D. McFadden (2008). The estimation of generalized extreme value models from choice-based samples. *Transportation Research Part B: Methodological* 42(4), 381–394.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and consumer choice. *Journal of Political Economy* 121(5), 803–843.
- Bruins, M., J. A. Duffy, M. P. Keane, and A. A. Smith Jr (2015). Generalized indirect inference for discrete choice models. *arXiv preprint arXiv:1507.06115*.
- Caplin, A., M. Dean, and J. Leahy (2016). Rational inattention, optimal consideration sets and stochastic choice. Technical report.
- Cattaneo, M., X. Ma, Y. Masatlioglu, and E. Suleymanov (2018). A random attention model.
- 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.
- Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2017). Inferring cognitive heterogeneity from aggregate choices.
- Davis, P. and P. Schiraldi (2014). The flexible coefficient multinomial logit (fc-mnl) model of demand for differentiated products. *The RAND Journal of Economics* 45(1), 32–63.
- Draganska, M. and D. Klapper (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research* 48(4), 653–669.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics* 129(4), 1661–1710.
- Gaynor, M., C. Propper, and S. Seiler (2016). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review*.
- Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica* 76(5), 1017–1074.
- Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of applied econometrics* 8(S1).
- Handel, B. R. and J. T. Kolstad (2015). Health insurance for “humans”: Information frictions, plan choice, and consumer welfare. *The American Economic Review* 105(8), 2449–2500.
- Hauser, J. R. and B. Wernerfelt (1990). An evaluation cost model of consideration sets. *Journal of consumer research* 16(4), 393–408.
- Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation¹. *Econometrica* 73(3), 669–738.
- Heiss, F., A. Leive, D. McFadden, and J. Winter (2013). Plan selection in medicare part d: Evidence from administrative data. *Journal of Health Economics* 32(6), 1325–1344.
- Heiss, F., D. McFadden, J. Winter, A. Wupperman, and B. Zhou (2016). Inattention and switching costs as sources of inertia in medicare part d. Working Paper.
- Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5(1), 123–144.

- Ho, K., J. Hogan, and F. Scott Morton (2015). The impact of consumer inattention on insurer pricing in the medicare part d program. Technical report, National Bureau of Economic Research.
- Honka, E. (2014). Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics* 45(4), 847–884.
- Honka, E. and P. Chintagunta (2016). Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science* 36(1), 21–42.
- 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.
- Kawaguchi, K., K. Uetake, and Y. Watanabe (2016). Identifying consumer attention: A product-availability approach.
- Keane, M. and A. A. Smith (2003). Generalized indirect inference for discrete choice models. *Yale University*.
- Kőszegi, B. and A. Szeidl (2012). A model of focusing in economic choice. *The Quarterly journal of economics* 128(1), 53–104.
- 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.
- Lu, Z. (2016). Estimating multinomial choice models with unobserved choice sets.
- Manski, C. F. (1977). The structure of random utility models. *Theory and decision* 8(3), 229–254.
- Manski, C. F. and S. R. Lerman (1977). The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society*, 1977–1988.
- Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica* 82(3), 1153–1176.
- Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012). Revealed attention. *The American Economic Review* 102(5), 2183–2205.

- Matejka, F. and A. McKay (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review* 105(1), 272–298.
- Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The computer journal* 7(4), 308–313.
- Shocker, A. D., M. Ben-Akiva, B. Boccara, and P. Nedungadi (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters* 2(3), 181–197.
- Simon, H. A. (1971). Designing organizations for an information-rich world.
- Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8(S1).
- Swait, J. and M. Ben-Akiva (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological* 21(2), 91–102.
- Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies* 70(1), 147–165.
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization* 1(1), 39–60.
- Treisman, A. M. and G. Gelade (1980). A feature-integration theory of attention. *Cognitive psychology* 12(1), 97–136.
- Van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research* 47(1), 63–74.

8 Tables in Main Text

Table 1: Experimental Data Estimation Results

| | Conditional | ASC Model | | Conditional on |
|-------------------|----------------------|----------------------|----------------------|----------------------|
| | Logit | MLE | Indirect Inf. | Consideration |
| <i>Utility:</i> | | | | |
| Price (dollars) | -0.054*** (0.003) | -0.196*** (0.028) | -0.1284** (0.048) | -0.173*** (0.004) |
| Product 1 | -1.411*** (0.054) | 1.465*** (0.539) | 0.5806 (0.361) | 0.368*** (0.069) |
| Product 2 | -1.955*** (0.069) | -0.065 (0.478) | -0.483* (0.283) | -0.497*** (0.080) |
| Product 3 | -1.627*** (0.059) | 0.625 (0.476) | 0.452 (0.295) | 0.093 (0.073) |
| Product 4 | -1.640*** (0.060) | 0.629 (0.466) | -0.007 (0.302) | 0.088 (0.073) |
| Product 5 | -1.447*** (0.056) | 0.707 (0.478) | 0.165 (0.269) | 0.306*** (0.070) |
| Product 6 | -0.435*** (0.039) | -0.737*** (0.121) | -0.475*** (0.135) | -0.581*** (0.045) |
| Product 7 | -0.855*** (0.045) | -1.280*** (0.141) | -0.875*** (0.155) | -1.075*** (0.051) |
| Product 8 | -0.662*** (0.041) | -1.185*** (0.137) | -0.811*** (0.138) | -0.909*** (0.048) |
| Product 9 | -0.316*** (0.038) | -0.561*** (0.118) | -0.430*** (0.161) | -0.405*** (0.044) |
| <i>Attention:</i> | | | | |
| Price (dollars) | | 0.137*** (0.017) | 0.141*** (0.025) | 0.15 |
| Product 1 | | -2.872*** (0.177) | -2.910*** (0.236) | -2.5 |
| Product 2 | | -2.674*** (0.288) | -2.311*** (0.257) | -2.5 |
| Product 3 | | -2.695*** (0.209) | -2.674*** (0.238) | -2.5 |
| Product 4 | | -2.704*** (0.205) | -2.687*** (0.267) | -2.5 |
| Product 5 | | -2.592*** (0.204) | -2.581*** (0.245) | -2.5 |
| Product 6 | | 0.152 (0.192) | 0.390 (0.249) | 0 |
| Product 7 | | 0.123 (0.292) | 0.137 (0.281) | 0 |
| Product 8 | | 0.258 (0.230) | -0.200 (0.259) | 0 |
| Product 9 | | 0.103 (0.176) | -0.129 (0.253) | 0 |

Notes: Table reports coefficient estimates from conditional logit and the ASC model. 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 "conditional on consideration" 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 2: Sensitivity of Switching to Default and Rival Plan Characteristics

| | Sensitivity to | |
|------------------------------|---------------------|------------------------|
| | Default | Average of Rival Plans |
| Annual Premium (1000s) | 0.852*** (0.031) | -0.287*** (0.100) |
| Out of Pocket Costs (1000s) | 0.352*** (0.050) | -0.310*** (0.052) |
| Variance of Costs (millions) | 0.182*** (0.041) | -0.238*** (0.041) |
| Drug Deductible (1000s) | 0.851*** (0.094) | 0.204 (0.333) |
| Donut Hole Coverage | -0.030 (0.033) | 0.193 (0.125) |
| Avg. Cons. Cost Sharing % | -0.110 (0.071) | 0.235 (0.300) |
| # of Top 100 Drugs | 0.013*** (0.003) | -0.020** (0.010) |
| Normalized Quality | 0.004 (0.005) | 0.008 (0.022) |

Notes: The table reports coefficients from a panel data regression of an indicator for whether individual i switched at time t on attributes of the default plan, average attributes of rival plans, and individual and time fixed effects. The first column reports the coefficients on the characteristics of the plan they were enrolled in the year prior (the default plan); the second column reports the coefficients on the average value of the characteristics of rival plans. Standard errors are in parentheses. *** significance at the 1% level, ** at 5%, and * at 10%.

Table 3: Part D Data: Conditional Logit and Attentive Logit Estimates

| | 2007 | | 2008 | | 2009 | |
|---------------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | Clogit | Alogit | Clogit | Alogit | Clogit | Alogit |
| <i>Utility:</i> | | | | | | |
| Annual Premium (hundreds) | -0.415*** (0.012) | -0.091*** (0.029) | -0.596*** (0.013) | -1.074*** (0.026) | -0.599*** (0.015) | -1.245*** (0.027) |
| Annual Out of Pocket Costs (hundreds) | -0.417*** (0.020) | -0.661*** (0.028) | -0.691*** (0.029) | -0.923*** (0.047) | -0.433*** (0.034) | -0.484*** (0.054) |
| Variance of Costs (millions) | -2.131*** (0.178) | -3.359*** (0.248) | -1.809*** (0.299) | -2.351*** (0.448) | -2.056*** (0.326) | -0.702 (0.526) |
| Deductible (hundreds) | -0.208*** (0.024) | -0.355*** (0.032) | -0.737*** (0.027) | -0.792*** (0.037) | -0.231*** (0.030) | -0.590*** (0.043) |
| Donut Hole Coverage | -0.178*** (0.055) | 0.505*** (0.074) | -0.263*** (0.065) | -0.798*** (0.120) | 1.335*** (0.083) | 1.917*** (0.142) |
| Average Consumer Cost Sharing % | 0.704** (0.280) | -0.071 (0.376) | -2.002*** (0.333) | -4.274*** (0.450) | 0.798** (0.358) | -1.898*** (0.541) |
| # of Top 100 Drugs in Formulary | 0.641*** (0.040) | 1.078*** (0.071) | 0.749*** (0.046) | 0.826*** (0.057) | -0.060*** (0.008) | 0.022* (0.013) |
| Normalized Quality Rating | 0.087*** (0.017) | 0.319*** (0.025) | 0.299*** (0.018) | 0.688*** (0.028) | 0.564*** (0.017) | 0.659*** (0.026) |
| Prior Year Plan | 5.930*** (0.025) | 0 (-) | 6.380*** (0.034) | 3.370*** (0.122) | 6.525*** (0.038) | 2.410*** (0.208) |
| <i>Attention:</i> | | | | | | |
| Annual Premium (hundreds) | | 0.240*** (0.016) | | 0.364*** (0.023) | | 0.068** (0.027) |
| Annual Out of Pocket Costs (hundreds) | | 0.141*** (0.038) | | 0.186*** (0.051) | | -0.029 (0.064) |
| Variance of Costs (millions) | | 2.037*** (0.315) | | -0.113 (0.455) | | 1.777*** (0.589) |
| Deductible (hundreds) | | 0.373*** (0.045) | | 0.182*** (0.053) | | 0.075 (0.065) |
| Donut Hole Coverage | | 0.829*** (0.082) | | -1.364*** (0.128) | | -0.268* (0.142) |
| Average Consumer Cost Sharing % | | 1.321** (0.538) | | -5.493*** (0.693) | | 0.060 (0.733) |
| # of Top 100 Drugs in Formulary | | -0.211*** (0.065) | | 0.429*** (0.102) | | 0.099*** (0.021) |
| Normalized Quality Rating | | 0.002 (0.024) | | 0.034 (0.036) | | -0.600*** (0.032) |

Notes: “Clogit” refers to the conditional logit model; “Alogit” refers to the attentive logit model (the DSC model). Estimates are the coefficients in the utility and attention equations (not marginal effects). The coefficients in the attention equation are the coefficients on the listed characteristics of the default good (demeaned). Standard errors are in parentheses. The attentive model also includes year fixed effects. *** denotes significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level. Standard errors in parentheses.

Table 4: Utility Coefficients for Overidentification Test

| | Coef. | Interactions. | | | |
|----------------------------------|----------------------|-------------------|---------------------|---------------------|----------------------|
| | | I(Age 70–79) | I(Age ≥ 80) | I(Non-White) | Experience |
| Δ Annual Premium (hundreds) | -0.046*** (0.007) | | | | |
| Δ Out of Pocket Costs (hundreds) | 0.007 (0.006) | | | | |
| Δ Drug Deductible (hundreds) | -0.075*** (0.013) | | | | |
| Annual Premium (hundreds) | | 0.043 (0.032) | 0.025 (0.038) | 0.040 (0.076) | -0.086*** (0.031) |
| Out of Pocket Costs (hundreds) | | 0.014 (0.044) | -0.011 (0.052) | 0.073 (0.106) | 0.047 (0.062) |
| Variance of Costs (millions) | | -0.104 (0.431) | -0.914* (0.496) | -0.752 (1.071) | -0.108 (0.586) |
| Drug Deductible (hundreds) | | 0.039 (0.045) | 0.180*** (0.052) | 0.322*** (0.104) | -0.090* (0.049) |
| Donut Hole Coverage | | -0.070 (0.122) | -0.162 (0.146) | 0.162 (0.260) | 0.198 (0.154) |
| Avg. Cons. Cost Sharing % | | -0.152 (0.553) | -0.949 (0.648) | 2.822** (1.394) | -0.580 (0.600) |
| # of Top 100 Drugs | | -0.024 (0.022) | -0.050** (0.025) | -0.032 (0.055) | -0.007 (0.015) |
| Normalized Quality | | 0.020 (0.034) | 0.175*** (0.039) | 0.112 (0.074) | 0.046 (0.030) |

Notes: The table reports the coefficients in the utility equation from estimation of the DSC model in the text with additional terms added to test the identifying restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016). These additional terms include the change in premiums, deductibles and out of pocket cost over time, as well as interactions of each of the included covariates with age, race and experience fixed effects. The first three rows report the coefficients in the utility equation of the changes over time, and the remaining rows report the coefficients for the interaction of the row variable and the column variable (each attribute interacted with age, race and experience fixed effects). All specifications also include direct effects for each of the listed attributes. The corresponding attention coefficients are reported in Appendix Table 9. Standard errors are in parentheses. *** significance at the 1% level, ** at 5%, and * at 10%.

9 Figures in Main Text

Figure 1: Lab Experiment: Sample Product Selection Screen



Collegiate Pacific Banner
("Yale University Lux et
Veritas")
\$8.00



**Embroidered Towel From
Team Golf**
\$20.00



Mug w/ Thumb Piece
\$11.00



LXG Power Bank
\$12.00



**Moleskin Large Notebook
with Debossed Wordmark,
Unruled**
\$23.00

(You must wait 10 seconds before clicking next to make sure you consider all options)

Next

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

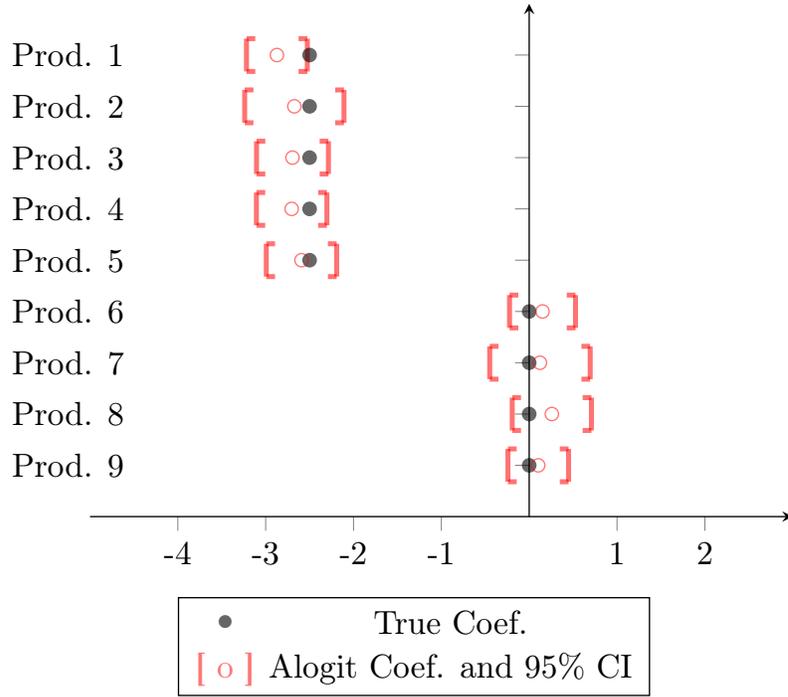
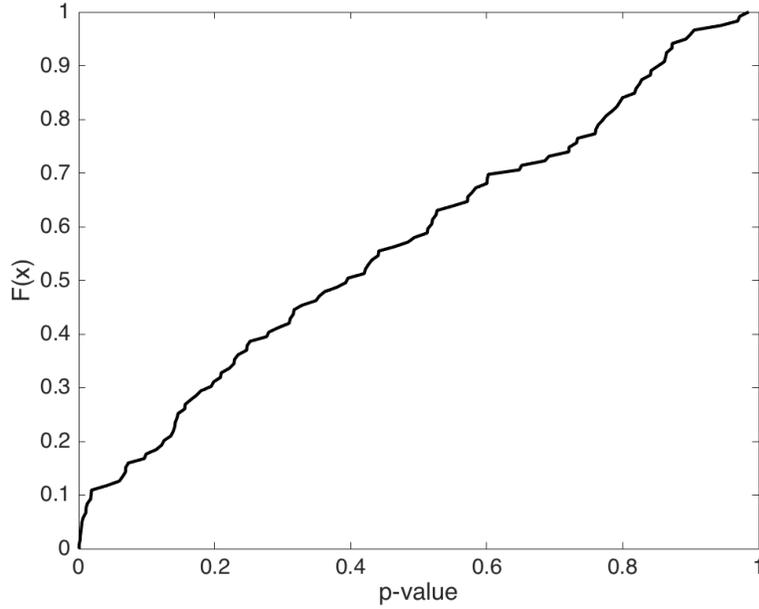
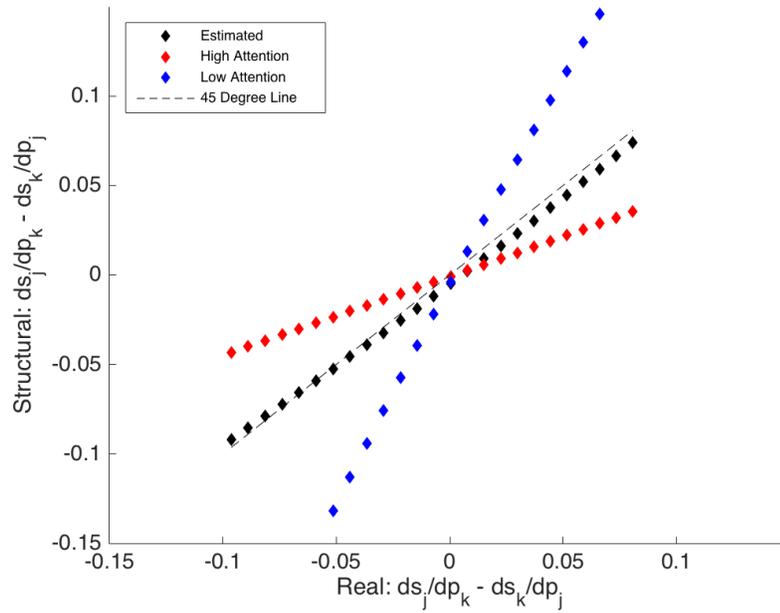


Figure 3: Distribution of p-values in test of $\hat{\theta} - \tilde{\theta} = 0$



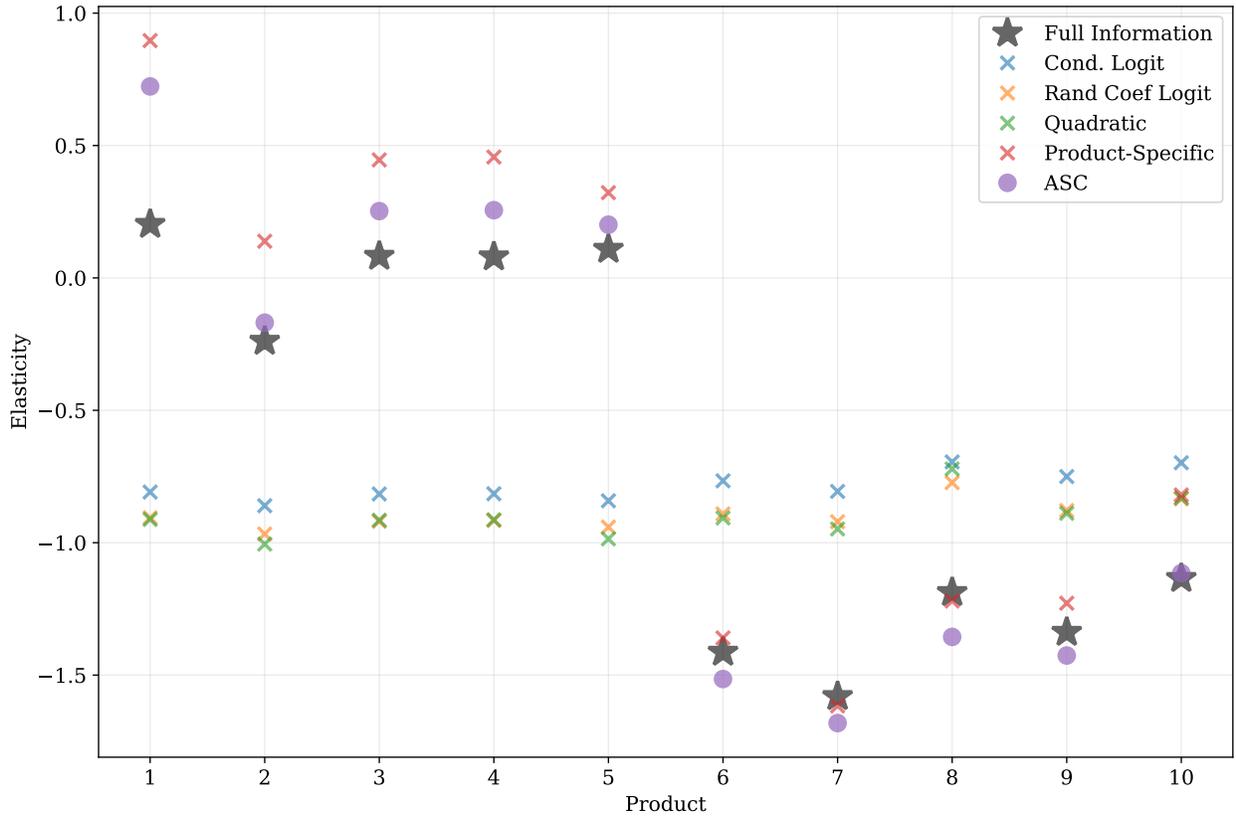
Notes: Figure shows the cumulative distribution of p-values of tests of equality of the coefficients of the auxiliary model when estimated on the real data and data simulated from our consideration model at the estimated parameters.

Figure 4: Matching Cross-Derivative Asymmetries



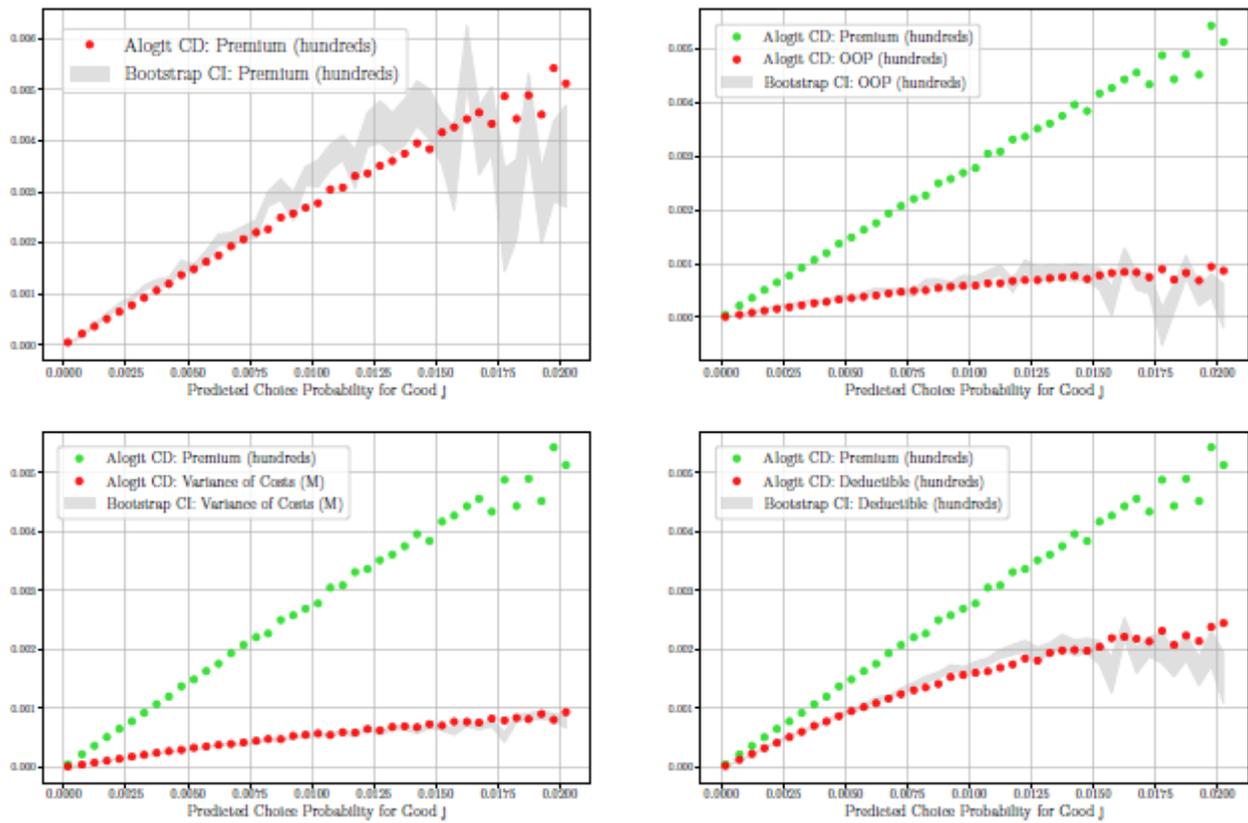
Notes: Figure shows a scatter plot of structural cross derivative asymmetries from the ASC model against those estimated in the auxiliary model between good 1 and good 6 as the price of good 6 varies from -6 to 6 around its average price. The black diamonds show the cross-derivative asymmetries at the estimated parameter values, while the blue and red dots show the cross-derivative asymmetries if the constant in the consideration equation is perturbed to decrease or increase the consideration probability respectively. Average consideration probabilities are 0.88 for the high attention case and 0.26 for the low attention case compared to an average consideration probability of 0.45 in the real experiment.

Figure 5: Experimental Data: Own-Price Elasticities by Good



Notes: Figure shows estimated price elasticities by product. The “full-information” model estimates a conditional logit model given the known consideration sets and computes the resulting reduced form price elasticities using the known relationship between consideration probability and price. The “random-coefficients” specification estimates price elasticities using only choices from all 10 goods in a random-coefficients logit model where each individual has a separate price coefficient. The conditional logit, quadratic and product-specific logit models respectively estimates conditional logit models with linear, quadratic, and alternative-specific price coefficients and with product-specific price coefficients.

Figure 6: Empirical vs. Model Predicted Cross-derivatives



A Model & Identification Proof

A.1 Results to Complement Section 3

COROLLARY 1. *Symmetry of Cross Derivatives:*

$$\frac{\partial s_{jm}^*}{\partial x_{j'm}^1} = \frac{\partial s_{j'm}^*}{\partial x_{jm}^1} \quad (\text{A.1})$$

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

$$\frac{\partial s_{jm}^*}{\partial x_{j'm}^1} = \int -\beta_i \int \dots \left[\int_{-\infty}^{v_{ijm}+e-v_{ijm}} \right] \dots \left[\int_{-\infty}^{v_{ijm}+e-v_{ij'm}} \right] \dots \int_{-\infty}^{v_{ijm}+e-v_{ijm}} f(z_0, \dots, e, \dots, v_{ijm}+e-v_{ij'm}, \dots, z_J) dz_J \dots [dz'_j] \dots [dz_j] \dots dz_0 de dF(\beta_i) \quad (\text{A.2})$$

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

$$\begin{aligned} \frac{\partial s_{jm}^*}{\partial x_{j'm}^1} &= \int -\beta_i \int \int_{-\infty}^{v_{ij'm}+t-v_{i0m}} \dots \left[\int_{-\infty}^{v_{ij'm}+t-v_{ij'm}} \right] \dots \left[\int_{-\infty}^{v_{ij'm}+t-v_{ijm}} \right] \dots \int_{-\infty}^{v_{ij'm}+t-v_{iJm}} \\ &\quad f(z_0, \dots, v_{ij'm}+t-v_{ijm}, \dots, t, \dots, z_J) dz_J \dots [dz'_j] \dots [dz_j] \dots dz_0 dt dF(\beta_i) \\ &= \frac{\partial s_{j'm}^*}{\partial x_{jm}^1} \end{aligned} \quad (\text{A.3})$$

COROLLARY 2. *Absence of Nominal Illusion:*

$$s_j^*(x_m^1, x_m^2) = Pr \left(v_{ijm} + \epsilon_{ijm} = \max_{j' \in \{0, \dots, J\}} v_{ij'm} + \epsilon_{ij'm} \right) \quad (\text{A.4})$$

$$= Pr \left(\beta_i x_{jm}^1 + w_{ij}(x_{jm}^2) + \epsilon_{ijm} = \max_{j' \in \{0, \dots, J\}} \beta_i x_{j'm}^1 + w_{ij}(x_{j'm}^2) + \epsilon_{ij'm} \right) \quad (\text{A.5})$$

$$= Pr \left(\beta_i x_{jm}^1 + \beta_i \delta + w_{ij}(x_{jm}^2) + \epsilon_{ijm} = \max_{j' \in \{0, \dots, J\}} \beta_i x_{j'm}^1 + \beta_i \delta + w_{ij}(x_{j'm}^2) + \epsilon_{ij'm} \right) \quad (\text{A.6})$$

$$= Pr \left(\beta_i (x_{jm}^1 + \delta) + w_{ij}(x_{jm}^2) + \epsilon_{ijm} = \max_{j' \in \{0, \dots, J\}} \beta_i (x_{j'm}^1 + \delta) + w_{ij}(x_{j'm}^2) + \epsilon_{ij'm} \right) \quad (\text{A.7})$$

$$= s_j^*(x_m^1 + \delta, x_m^2) \quad (\text{A.8})$$

PROOF OF LEMMA 1. With a slight abuse of notation, let the set of consideration sets containing

good j and j' be given as:

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

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

$$\frac{\partial s_{jm}}{\partial x_{j'm}^1} - \frac{\partial s_{j'm}}{\partial x_{jm}^1} = \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_c}{\partial x_{j'm}^1} s_{jm}^*(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{c'}}{\partial x_{ij}^1} s_{j'm}^*(c') + \sum_{c' \in \mathbb{P}(j, j')} \pi_{c'} \left(\frac{\partial s_{jm}^*(c')}{\partial x_{j'm}^1} - \frac{\partial s_{j'm}^*(c')}{\partial x_{jm}^1} \right) \quad (\text{A.10})$$

$$= \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_c}{\partial x_{j'm}^1} s_{jm}^*(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{c'}}{\partial x_{jm}^1} s_{j'm}^*(c') \quad (\text{A.11})$$

Thus, non-zero cross-derivative asymmetries imply:

$$\sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_c}{\partial x_{j'm}^1} s_{jm}^*(c) \neq \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{c'}}{\partial x_{jm}^1} s_{j'm}^*(c') \quad (\text{A.12})$$

$$\text{either } \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_c}{\partial x_{j'm}^1} s_{jm}^*(c) \neq 0 \quad \text{and/or} \quad \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{c'}}{\partial x_{jm}^1} s_{j'm}^*(c') \neq 0 \quad (\text{A.13})$$

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

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

$$s_j(x_{jm}^1 + \delta, x_m^2) = \sum_{c \in \mathbb{P}(j)} \pi_c(x_m^1 + \delta, x_m^2) Pr \left(v_{ijm} + \epsilon_{ijm} = \max_{j' \in c} v_{ij'm} + \epsilon_{ij'm} \right) \quad (\text{A.14})$$

If $s_j(x_m^1, x_m^2) \neq s_j(x_m^1 + \delta, x_m^2)$, this implies that for at least one consideration set c

$$\pi_c(x_m^1 + \delta, x_m^2) \neq \pi_c(x_m^1, x_m^2), \quad (\text{A.15})$$

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

ASSUMPTION 6. (RANK CONDITION) The matrix $D'_m D_m$ 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:

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

$$J \geq 2 \quad (\text{A.17})$$

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

$$s_{jm}(\mathcal{J}) \neq s_{j'm}(\mathcal{J}) \quad (\text{A.18})$$

$$\frac{s_{lm}(\mathcal{J}) - s_{lm}(\mathcal{J}/j)}{s_{j'm}(\mathcal{J}) - s_{j'm}(\mathcal{J}/j)} \neq \frac{s_{lm}(\mathcal{J}) - s_{lm}(\mathcal{J}/j')}{s_{jm}(\mathcal{J}) - s_{jm}(\mathcal{J}/j')} \quad (\text{A.19})$$

$$s_{jm}(\mathcal{J}) - s_{jm}(\mathcal{J}/j') \neq 0 \quad (\text{A.20})$$

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

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

$$\begin{bmatrix} -(s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/1)) & 0 & s_{1m}(\mathcal{J}) \\ 0 & -(s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/2)) & s_{2m}(\mathcal{J}) \\ -(s_{2m}(\mathcal{J}) - s_{2m}(\mathcal{J}/1)) & (s_{1m}(\mathcal{J}) - s_{1m}(\mathcal{J}/2)) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \log(\phi_{1m})}{\partial x_{1m}} \\ \frac{\partial \log(\phi_{2m})}{\partial x_{2m}} \\ \frac{\partial \log(\mu_m)}{\partial x_{0m}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_{1m}}{\partial x_{0m}^1} - \frac{\partial s_{0m}}{\partial x_{1m}^1} \\ \frac{\partial s_{2m}}{\partial x_{0m}^1} - \frac{\partial s_{0m}}{\partial x_{2m}^1} \\ \frac{\partial s_{1m}}{\partial x_{2m}^1} - \frac{\partial s_{2m}}{\partial x_{1m}^1} \end{bmatrix} \quad (\text{A.21})$$

The determinant of D_m is:

$$\begin{aligned} \det(D_m) &= s_{2m}(\mathcal{J}) (s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/1)) (s_{1m}(\mathcal{J}) - s_{1m}(\mathcal{J}/2)) \\ &\quad - s_{1m}(\mathcal{J}) (s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/2)) (s_{2m}(\mathcal{J}) - s_{2m}(\mathcal{J}/1)) \end{aligned} \quad (\text{A.22})$$

When D_i is singular:

$$\frac{1}{s_{1m}(\mathcal{J})} \left(\frac{s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/1)}{s_{2m}(\mathcal{J}) - s_{2m}(\mathcal{J}/1)} \right) = \frac{1}{s_{2m}(\mathcal{J})} \left(\frac{s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/2)}{s_{1m}(\mathcal{J}) - s_{1m}(\mathcal{J}/2)} \right) \quad (\text{A.23})$$

ASSUMPTION 8. (RANK CONDITION) $\Pi(\delta)$ is full rank.

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

$$\frac{\phi_j(x_{jm}^1 + \delta_i)}{1 - \phi_j(x_{jm}^1 + \delta_i)} \neq \frac{\phi_j(x_{jm}^1 + \delta_{i'})}{1 - \phi_j(x_{jm}^1 + \delta_{i'})} \quad (\text{A.24})$$

$$\phi_j(x_{jm}^1 + \delta_i) \neq \phi_{j'}(x_{j'm}^1 + \delta_i) \quad \text{at, at least one } i = 1, \dots, N \quad (\text{A.25})$$

for $j, j' > 0$.

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

$$\Pi(\delta) = \begin{bmatrix} \mu(\delta_1)\phi_1(\delta_1)(1 - \phi_2(\delta_1)) & \mu(\delta_1)\phi_1(\delta_1)\phi_2(\delta_1) & 0 & 0 \\ \mu(\delta_2)\phi_1(\delta_2)(1 - \phi_2(\delta_2)) & \mu(\delta_2)\phi_1(\delta_2)\phi_2(\delta_2) & 0 & 0 \\ 0 & 0 & \mu(\delta_1)\phi_2(\delta_1)(1 - \phi_1(\delta_1)) & \mu(\delta_1)\phi_1(\delta_1)\phi_2(\delta_1) \\ 0 & 0 & \mu(\delta_2)\phi_2(\delta_2)(1 - \phi_1(\delta_2)) & \mu(\delta_2)\phi_1(\delta_2)\phi_2(\delta_2) \end{bmatrix} \quad (\text{A.26})$$

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

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

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

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

$$\frac{1 - \phi_2(\delta_1)}{\phi_2(\delta_1)} = \frac{1 - \phi_2(\delta_2)}{\phi_2(\delta_2)} \quad (\text{A.29})$$

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

$$\frac{1 - \phi_1(\delta_1)}{\phi_1(\delta_1)} = \frac{1 - \phi_1(\delta_2)}{\phi_1(\delta_2)} \quad (\text{A.30})$$

When $J > 2$, we require that $\phi_j(x_{jm}^1 + \delta_i) \neq \phi_{j'}(x_{j'm}^1 + \delta_i)$ 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:

$$s_{jm} = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_l(x_{0m}, x_{lm}) \prod_{l' \notin c} (1 - \phi_{l'}(x_{0m}, x_{lm})) \quad (\text{A.31})$$

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

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

$$\frac{\partial s_{jm}}{\partial x_{0m}^1} - \frac{\partial s_{0m}}{\partial x_{jm}^1} = \frac{\partial \log(\phi_{jm})}{\partial x_{0m}^1} s_{jm}(\mathcal{J}) + \sum_{j' \neq \{j, 0\}} \frac{\partial \log(\phi_{j'm})}{\partial x_{0m}^1} (s_{jm}(\mathcal{J}) - s_{jm}(\mathcal{J}/j')) - \frac{\partial \log(\phi_{jm})}{\partial x_{jm}^1} (s_{0m}(\mathcal{J}) - s_{0m}(\mathcal{J}/j)) \quad (\text{A.32})$$

Thus there are now $2J$ derivatives of log consideration probabilities to identify: $\partial \log(\phi_{jm})/\partial x_{jm}^1$ and $\partial \log(\phi_{jm})/\partial x_{0m}^1$ 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):

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

$$J \geq 3 \quad (\text{A.34})$$

In this model, we cannot allow $\phi_{0m}(x_{0m}) \leq 1$ and the rank condition still hold. This is because we will only ever have J independent cross derivatives involving the default good but there will be $J+1$ changes in consideration probabilities with respect to the default good to identify. Other than this restriction, the rest of the proof in Section 3 goes through without modification.

A.3 ASC Identification with an ‘Outside’ Default Good

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

changes (Assumption 6). We now require:

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

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

All cross derivative differences take the form given by Equation and the rest of the identification proof of Theorem 1 continues as in Section 3.

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:

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

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

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

$$\frac{\partial s_{i0}}{\partial x_{ij}^1} - \frac{\partial s_{ij}}{\partial x_{i0}^1} = \frac{\partial \log(\phi_{ij})}{\partial x_{ij}^1} (s_{i0}(\mathcal{J}) - s_{i0}(\mathcal{J}/j)) - \frac{\partial \log(\phi_{i0})}{\partial x_{i0}^1} (s_{ij}(\mathcal{J}) - s_{ij}(\mathcal{J}/0)) \tag{A.39}$$

This expression might seem somewhat odd given that 'leave-zero-out' variation is required. How natural this assumption is might vary across contexts. If default goods are randomly assigned in the population, this variation (or permitting the market share of good-0 to go to zero) might be plausible. If Assumption 7a holds, then the proof of identification follows as in Section 3 with the above modification to cross derivative differences involving the default.

A.5 Identification of Features in General Consideration Set Model

Our proof of constructive point identification relies on the structure imposed by the ASC and DSC frameworks. However, features of a more general model of consideration sets can still be identified from cross-derivative asymmetries. This remains the case with correlation between the unobservables driving consideration probabilities, e.g. correlation in expected random utility errors. To illustrate,

let $g_{jm} = x_{jm}\gamma$ and assume that the impact of characteristics on attention probabilities comes via the indices g_{jm} . The general expression for cross-derivative differences (with respect to attribute k) in consideration set models then takes the form:

$$\frac{\partial s_{jm}}{\partial x_{j'm}^k} - \frac{\partial s_{j'm}}{\partial x_{jm}^k} = \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_c(g_{0m}, \dots, g_{Jm})}{\partial x_{j'm}^k} s_{jm}^*(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{c'}(g_{0m}, \dots, g_{Jm})}{\partial x_{jm}^k} s_{j'm}^*(c') \quad (\text{A.40})$$

$$= \gamma_k \sum_{c \in \mathbb{P}(j)} \frac{\partial \pi_c(g_{0m}, \dots, g_{Jm})}{\partial g_{j'm}} s_{jm}^*(c) - \sum_{c' \in \mathbb{P}(j')} \frac{\partial \pi_{c'}(g_{0m}, \dots, g_{Jm})}{\partial g_{jm}} s_{j'm}^*(c') \quad (\text{A.41})$$

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

$$\frac{\frac{\partial s_{jm}}{\partial x_{j'm}^k} - \frac{\partial s_{j'm}}{\partial x_{jm}^k}}{\frac{\partial s_{jm}}{\partial x_{j'm}^{k'}} - \frac{\partial s_{j'm}}{\partial x_{jm}^{k'}}} = \frac{\gamma_k}{\gamma_{k'}} \quad (\text{A.42})$$

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

B Estimation

B.1 Maximum Likelihood Estimation

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

$$s_{jm}(x_m; \beta) = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_l(x_{lm}; \beta) \prod_{l' \notin c} (1 - \phi_{l'}(x_{l'm}; \beta)) s_j^*(x_m | C, \beta) \quad (\text{B.1})$$

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

$$\log \mathcal{L}(\beta) = \sum_{i=1}^N \sum_{j=0}^J y_{ij} \log(s_{jm}(\beta)) \quad (\text{B.2})$$

We note that a major computational issue arises with larger choice sets - there are 2^J possible

consideration sets to sum over to construct choice probabilities. To deal with this problem, we advocate that researchers follow the simulated likelihood approach outlined in Goeree (2008) and we refer to readers to <https://sites.google.com/view/alogit/home> for further practical details.

B.2 Indirect Inference

In Section 5, we estimate the ASC model by indirect inference, picking our structural parameters to match the coefficients of a flexible auxiliary model that allows for cross-derivative asymmetries. We define a flexible logit model as:

$$\tilde{u}_{ijm} = x_{ijm}\theta_j + \sum_{j'} \sum_k \sum_{k'} \theta_{j,j'}^k x_{ijm}^k x_{ij'm}^k + e_{ijm} \quad (\text{B.3})$$

with e_{ijm} distributed i.i.d Type 1 Extreme Value. This implies choice probabilities:

$$\tilde{f}_{jm} = \frac{\exp\left(x_{ijm}\theta_j + \sum_{j'} \sum_k \theta_{j,j'}^k x_{ijm}^k x_{ij'm}^k\right)}{\sum_{l=0}^J \exp\left(x_{ilm}\theta_l + \sum_{l'} \sum_k \theta_{l',l}^k x_{ilm}^k x_{il'm}^k\right)} \quad (\text{B.4})$$

See below for a formal justification for this specification.

We generate $M = 3$ sets of structural errors and estimate the auxiliary model on data simulated from our structural model given a particular guess of the structural parameters

$$\tilde{\theta}(\beta) = M^{-1} \sum \tilde{\theta}^m(\beta) \quad (\text{B.5})$$

$$\tilde{\theta}^m(\beta) = \arg \min_{\theta} \sum_{i=1}^N \sum_{j=0}^J y_{ij} \log(f_{jm}(\beta)) \quad (\text{B.6})$$

We pick β to minimize the difference between the auxiliary parameters estimated on the real data and data simulated from our consideration set model. Formally, $\hat{\beta}$ solves:

$$\hat{\beta} = \arg \min_{\beta} Q(\beta) \quad (\text{B.7})$$

$$Q(\beta) = \left(\hat{\theta} - \tilde{\theta}(\beta)\right)' W \left(\hat{\theta} - \tilde{\theta}(\beta)\right) \quad (\text{B.8})$$

We choose the weight matrix as the inverse of the variance-covariance matrix of the auxiliary parameters estimated on the real data: $W = \Sigma_{\theta}^{-1}$.

Some computational difficulties arise from the fact that our choice variable is discrete. Therefore, as it stands our objective function is not a smooth function of our structural parameters as small changes in β result in discrete changes in our simulated data and thus auxiliary parameters. This

renders standard gradient-based optimization methods unsuitable. We thus estimate the parameters using the Nelder-Mead simplex approach (Nelder and Mead 1965).

Choice of Auxiliary Model To motivate our choice of auxiliary model, note that the consideration set models we consider can be written as full consideration models in which utility depends on own and rival goods characteristics. We will derive an explicit expression of this form with logit errors.

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

$$u_{ijm} = \beta_i x_{jm}^1 + w_{ij}(x_{jm}^2) + \epsilon_{ijm} \quad (\text{B.9})$$

$$= v_{ijm} + \epsilon_{ijm} \quad (\text{B.10})$$

where $v_{ijm} = \beta_i x_{jm}^1 + w_{ij}(x_{jm}^2)$.

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

$$s_{1m} = \phi_{1m} s_1^*(x_m) \quad (\text{B.11})$$

where $s_1^*(x_m)$ is the probability of choosing good 1 conditional on paying attention and $x_m = [x_m^1, x_m^2]$.

With i.i.d. Type 1 Extreme Value errors, this model is equivalent to a full-consideration model with preferences specified as:

$$\tilde{u}_{ijm} = v_{ijm} + \psi_{i,j=1,m} + \epsilon_{ijm} \quad (\text{B.12})$$

where $\psi_{i,j=1,m} = \psi_{i1m}$ for plan 1 and is 0 otherwise, where ψ_{i1m} is given by:

$$\psi_{i1m} = \ln \left(\frac{\phi_{1m}(x_{1m}) \exp(v_{i0m})}{(1 - \phi_{1m}(x_{1m})) \exp(v_{i1m}) + \exp(v_{i0m})} \right) \quad (\text{B.13})$$

This follows since:

$$s_{1m} = \frac{\exp(v_{i1m} + \psi_{i,j=1,m})}{\exp(v_{i1m} + \psi_{i,j=1,m}) + \exp(v_{i0m})} = \phi_{1m}(x_{1m}) \frac{\exp(v_{i1m})}{\exp(v_{i1m}) + \exp(v_{i0m})} \quad (\text{B.14})$$

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

equations:

$$\psi_{ijm} = \ln \left(\frac{\phi_{jm} \sum_{k \neq j} \exp(v_{ikm} + \psi_{ikm})}{(1 - \phi_{jm}) \exp(v_{ijm}) + \sum_{k \neq j} \exp(v_{ikm} + \psi_{ikm})} \right) \quad (\text{B.15})$$

We showed above that this holds for the case where $J = 2$. Let s_{jm}^a denote the probability of choosing good j conditional on paying full attention to good j , i.e. $s_{jm}^a = s_{jm}(x_m | \phi_{jm} = 1)$. In the two-good case, $s_{1m}^a = s_1^*$ but more generally:

$$s_{jm}^a = \sum_{C \in \mathbb{P}(j)} \prod_{l \in C, l \neq j} \phi_{lm} \prod_{l' \notin C} (1 - \phi_{l'm}) s_{jm}^*(C) \quad (\text{B.16})$$

Next, consider adding a J th plan to which you might be inattentive:

$$s_{Jm} = \phi_{Jm} s_{Jm}^a \quad (\text{B.17})$$

By the inductive hypothesis, we have:

$$s_{Jm}^a = \frac{\exp(v_{iJm})}{\exp(v_{iJm}) + \sum_{k \neq J} \exp(v_{ikm} + \psi_{ikm})} \quad (\text{B.18})$$

Therefore,

$$s_{Jm} = \phi_{Jm} \frac{\exp(v_{iJm})}{\exp(v_{iJm}) + \sum_{k \neq J} \exp(v_{ikm} + \psi_{ikm})} \quad (\text{B.19})$$

It is straightforward to confirm by plugging into the logit formulas that these choice probabilities result from full-consideration utility maximization given that the J th good has utility given by:

$$u_{iJm} = v_{iJm} + \psi_{iJm} + \epsilon_{iJm} \quad (\text{B.20})$$

where:

$$\psi_{ijm} = \ln \left(\frac{\phi_{jm} \sum_{k \neq j} \exp(v_{ikm} + \psi_{ikm})}{(1 - \phi_{jm}) \exp(v_{ijm}) + \sum_{k \neq j} \exp(v_{ikm} + \psi_{ikm})} \right) \quad (\text{B.21})$$

Thus, if this representation holds for a choice set with $J - 1$ plans, it holds for a choice set with J plans, and the proof is complete for the ASC model. The auxiliary equation used in the text in which u_{ijm} depends on all quadratic functions of own and rival attributes can be derived as a 2nd order Taylor-expansion of ψ_{ijm} with respect to the attributes of rival goods around the point where all of these attributes are 0 so that $\psi_{ijm} = 0$.

Next, consider the DSC model.

$$\begin{aligned} s_{0m} &= (1 - \mu_m) + \mu_m s_{0m}^*(\mathcal{J}) \\ s_{jm} &= \mu_m s_{jm}^*(\mathcal{J}) \quad \text{for } j > 0 \end{aligned} \tag{B.22}$$

where s_{jm}^* are the choice probabilities which result from maximizing:

$$u_{ijm} = v_{ijm} + \xi_{j=d} + \epsilon_{ijm} \tag{B.23}$$

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

$$u_{ijm} = v_{ijm} + \xi_{j=d} + \psi_{i,j=d,m} + \epsilon_{ijm} \tag{B.24}$$

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

$$\psi_{im} = \ln \left(\frac{1 + (1 - \mu_m) \sum_{k \neq d} \exp(v_{ikm} - v_{idm} - \xi)}{\mu_m} \right) \tag{B.25}$$

C Additional Tables and Figures for Empirical Work

This Section provides additional tables and figures that illustrate the empirical results of Sections 5 and 6.

Lab Experiment Table 5 shows the products used in the experiment and their list prices. A sample product selection screen is shown in Figure 1.

Part D Analysis Summary statistics from our data after all sample selection restrictions are imposed are reported in Table 6 in Appendix C. We report the mean and standard deviation of a variety of characteristics for all plans and also for chosen plans.

Figure C.1 gives the predicted cross derivative difference between default and non-default goods for included plan characteristics for four variables; the charts for all variables are in Appendix C. We graph both the estimated cross derivatives from Equation D.7 and the cross-derivatives implied by the DSC model ($\gamma_k(1 - \mu_{i0})\hat{s}_{ij}$) against the predicted market share of plan j , \hat{s}_{ij} . To capture the uncertainty in the estimated cross-derivatives, we bootstrap estimation of Equation D.7 and graph

Table 5: 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.

Table 6: Part D Data: Summary Statistics

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

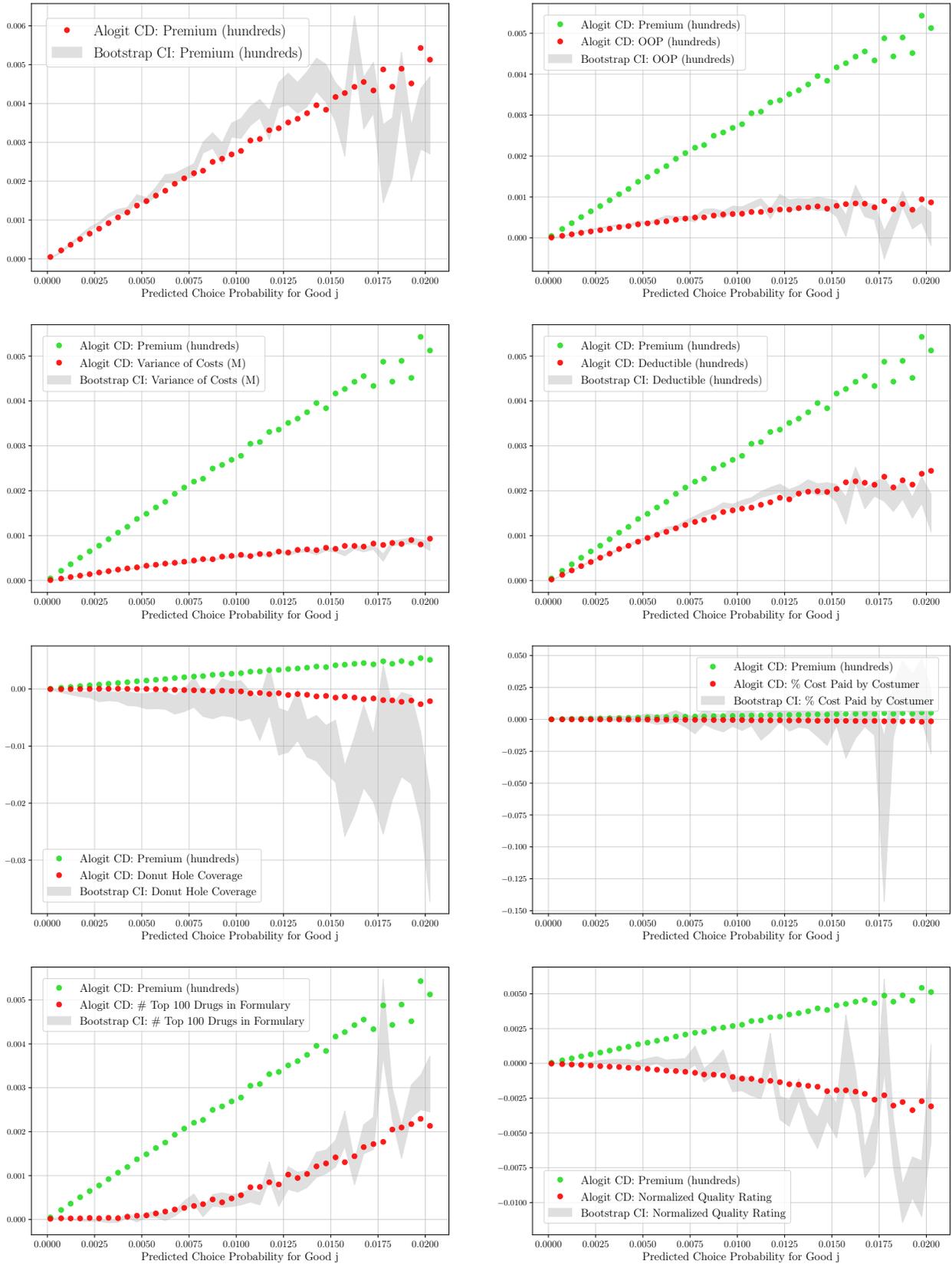
Notes: Table reports means and standard deviations (in parenthesis) of each variable for the beneficiaries in our final sample. The sample consists of an observation for each (year, beneficiary, plan).

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 D.7. 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 C.1: Empirical vs. Model Predicted Cross-derivatives



D Testing Consistency of Cross Derivative Asymmetry Patterns in the DSC Model

Given the individual level variation in the Medicare Part D data, we here drop reference to variation in product attributes across markets. The DSC model models choice probabilities as:

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

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

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

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

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

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

where $\psi_{i,j=d}$ is the a term that reflects the impact of imperfect consideration that varies as a function of own and rival characteristics. The test we propose is to first estimate the DSC model to recover $\hat{\beta}$ and $\hat{\gamma}$, and thus $\hat{\psi}_{ij}$, to form:

$$\hat{v}_{ij} = x_{ij}\hat{\beta} + \hat{\psi}_{i,j=d} \tag{D.5}$$

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

$$\frac{\partial s_{ij}}{\partial x_{idk}} - \frac{\partial s_{id}}{\partial x_{ijk}} = \hat{\gamma}_k (1 - \hat{\mu}_i) \hat{s}_{ij} \tag{D.6}$$

To determine whether the DSC model is sufficiently flexible to be able to capture the patterns in empirical cross-derivatives, we next estimate the following model with a rich set of interaction

terms:

$$\tilde{u}_{ij} = x_{ij}\hat{\beta} + \hat{\psi}_{i,j=d} + \sum_k \sum_{k'} x_{idk}x_{ijk'}\alpha_{k,k'} + \epsilon_{ij} \text{ for } j \neq d \quad (\text{D.7})$$

with predicted market shares \tilde{s}_{ij} and predicted latent shares, \tilde{s}_{ij}^* .

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

$$\frac{\partial s_{ij}}{\partial x_{idk}} - \frac{\partial s_{id}}{\partial x_{ijk}} = \frac{\hat{\beta}_k(\hat{s}_{id} - \hat{s}_{id}^*)\tilde{s}_{id}((1 - \hat{s}_{id})\tilde{s}_{ij} - (1 - \tilde{s}_{id})\hat{s}_{ij}) + \hat{\gamma}_k(1 - \hat{\mu}_i)(1 - \hat{s}_{id})\tilde{s}_{id}\tilde{s}_{ij}}{\hat{s}_{id}(1 - \hat{s}_{id})} + \tilde{s}_{ij} \left[\sum_{k'} (x_{ijk'} - \tilde{x}_{ik'})\alpha_{k,k'} + \tilde{s}_{id} \sum_{k'} x_{idk'}\alpha_{k'k} \right] \quad (\text{D.8})$$

where $\tilde{x}_{ik} = \sum_{j \neq d} \tilde{s}_{ij}x_{ijk}$. Note that when all $\alpha_{k,k'} = 0$, we have: $\tilde{s}_{ij} = \hat{s}_{ij}$ and there is no difference in estimated cross-derivative differences at the first and second stages. Thus, if there are no significant differences between these cross-derivative difference estimates, we conclude that the DSC model fits the data well. In practice, we estimate the difference in Equations D.6 and D.8 by quantile of \tilde{s}_{ij} .

E Robustness

In this section, we report several robustness checks for the empirical specifications in Sections 5 and 6.

ASC Robustness Table 7 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 1 in the text.

DSC Robustness Table 8 reports coefficient estimates from the DSC model with brand fixed effects added. Table 9 reports the coefficients in the attention equation for the additional terms added to the DSC model to test the overidentifying restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016).

Table 7: Experimental Data Estimation Results

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

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 8: Part D Results w/ Brand Fixed Effects

| | Clogit | Alogit |
|---------------------------------------|----------------------|----------------------|
| <i>Utility:</i> | | |
| Annual Premium (hundreds) | -0.575 (0.009)*** | -0.930 (0.018)*** |
| Annual Out of Pocket Costs (hundreds) | -0.470 (0.018)*** | -0.643 (0.026)*** |
| Variance of Costs (millions) | -1.683 (0.160)*** | -2.054 (0.241)*** |
| Deductible (hundreds) | -0.434 (0.018)*** | -0.446 (0.025)*** |
| Donut Hole Coverage | 0.648 (0.045)*** | 0.745 (0.072)*** |
| Average Consumer Cost Sharing % | -1.081 (0.234)*** | -2.167 (0.355)*** |
| # of Top 100 Drugs in Formulary | -0.006 (0.010) | 0.045 (0.015)*** |
| Normalized Quality Rating | 0.202 (0.024)*** | 0.465 (0.040)*** |
| Prior Year Plan | 5.635 (0.020)*** | 2.016 (0.110)*** |
| <i>Attention:</i> | | |
| Annual Premium (hundreds) | | 0.137 (0.022)*** |
| Annual Out of Pocket Costs (hundreds) | | 0.121 (0.032)*** |
| Variance of Costs (millions) | | 1.277 (0.269)*** |
| Deductible (hundreds) | | 0.143 (0.039)*** |
| Donut Hole Coverage | | -0.348 (0.095)*** |
| Average Consumer Cost Sharing % | | 0.813 (0.461)* |
| # of Top 100 Drugs in Formulary | | 0.032 (0.024) |
| Normalized Quality Rating | | 0.060 (0.048) |

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 coefficients in the attention equation are the coefficients on the listed characteristics of the default good (demeaned). Standard errors are in parentheses. The attentive model also includes a constant. *** denotes significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level. Standard errors in parentheses.

Table 9: Attention Coefficients for Overidentification Test

| | Coef. | Interactions. | | | |
|----------------------------------|----------------------|----------------------|----------------------|---------------------|----------------------|
| | | I(Age 70–79) | I(Age ≥ 80) | I(Non-White) | Experience |
| Δ Annual Premium (hundreds) | 0.026 (0.021) | | | | |
| Δ Out of Pocket Costs (hundreds) | 0.064 (0.014)*** | | | | |
| Δ Drug Deductible (hundreds) | -0.041 (0.015)*** | | | | |
| Annual Premium (hundreds) | | -0.009 (0.028) | -0.188*** (0.030) | -0.129** (0.063) | -0.825*** (0.316) |
| Out of Pocket Costs (hundreds) | | -0.111 (0.069) | -0.212*** (0.075) | -0.184 (0.159) | 0.151** (0.070) |
| Variance of Costs (millions) | | 0.987 (0.605) | 0.643 (0.663) | -2.543 (1.600) | -0.559 (0.648) |
| Drug Deductible (hundreds) | | 0.164** (0.074) | 0.147* (0.082) | -0.015 (0.161) | -0.243*** (0.068) |
| Donut Hole Coverage | | 0.211 (0.152) | 1.021*** (0.165) | 0.552* (0.305) | 0.467*** (0.174) |
| Avg. Cons. Cost Sharing % | | 1.742** (0.856) | 2.152** (0.961) | -0.055 (1.912) | -2.527*** (0.778) |
| # of Top 100 Drugs | | -0.123*** (0.040) | -0.073 (0.048) | -0.049 (0.088) | 0.002 (0.028) |
| Normalized Quality | | -0.023 (0.037) | 0.089** (0.042) | 0.276*** (0.081) | -0.064** (0.029) |

Notes: The table reports the coefficients in the attention equation from estimation of the DSC model in the text with additional terms added to test the identifying restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016). These additional terms include the change in premiums, deductibles and out of pocket cost over time, as well as interactions of each of the included covariates with age, race and experience fixed effects. The first three rows report the coefficients in the attention equation of the changes over time, and the remaining rows report the coefficients on the plan attributes interacted with age, race and experience fixed effects. All specifications also include year fixed effects and direct effects for each of the listed attributes. Standard errors are in parentheses. *** denotes significance at the 1% level, ** at 5%, and * at 10%.