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

Jason Abaluck and Abi Adams-Prassl

*

February 12, 2021

Abstract

Consideration set models generalize discrete choice models by relaxing the assumption that consumers consider all available options. Determining which options were considered has previously required either survey data or restrictions on how attributes impact consideration or utility. We provide an alternative route. 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. We show that consideration probabilities are constructively identified from the resulting asymmetries. We validate our approach in a lab experiment where consideration sets are known and then apply our framework to study a "smart default" policy in Medicare Part D, wherein consumers are automatically reassigned to lower cost prescription drug plans with the option of opting out. Full consideration models imply such a policy will be ineffective because consumers will opt out to avoid switching costs. Allowing for inattention, we find that defaulting all consumers to lower cost options produces negligible welfare benefits on average, but defaulting only consumers who would save at least $300 produces large benefits.

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

*Thanks to Leila Bengali, Mauricio Caceres and Anna Papp for excellent research assistance and to Dan Ackerberg, Joe Altonji, Dan Benjamin, Doug Bernheim, Steve Berry, Richard Blundell, Judy Chevalier, Ian Crawford, Francis Di Traglia, Aureo de Paula, Jonathan Feinstein, Jeremy Fox, Xavier Gabaix, Jonathan Gruber, Phil Haile, Hamish Low, Erzo Luttmer, Paola Manzini, Marco Mariotti, Olivia Mitchell, Francesca Molinari, Fiona Scott Morton, Barry Nalebuff, Jeremias Prassl, 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, LSE, 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. We would like to thank the Editor and four anonymous referees for their help in improving the paper. Any remaining errors are rationalizable with sufficiently flexible preferences.

price increase but remain unresponsive if rivals lower prices? Normatively, whether people choose the same products year-after-year because they like those options or because they do not know what else exists has first-order consequences for welfare. If we can measure preferences conditional on consideration, we can assess the benefits of policies that make consumers aware of more options or redirect consumers to products they would choose if they considered them. Such policies are ubiquitous, ranging from defaulting people into lower cost insurance plans to populating online shopping carts with items that people might like.

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 2018), or because consumers face (unobserved) constraints on what options can be chosen (Gaynor, Propper, and Seiler 2016).[1] In contrast to tests of rationality, such as checking whether consumers make dominated choices, consideration set models allow us to simulate how consumers would choose if they were informed about relevant options.

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 usually rely either on auxiliary data on what goods are considered or on additional exclusion restrictions for point identification of the structural functions of interest.

[1] Given 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.

These exclusion restrictions are often questionable and can be in tension with economic theory; excluding prices from consideration, for example, can be inconsistent with simple models of rational inattention, and it is unclear that advertising only impacts choices via informing consumers about which goods exist (Goeree 2008; Van Nierop, Bronnenberg, Paap, Wedel, and Franses 2010). Yet, 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, Coughlin, Molinari, and Teitelbaum 2018).

Our approach builds on a recent literature in behavioral decision theory on limited consideration models (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014; Cattaneo, Ma, Masatlioglu, and Suleymanov 2020). We prove that the restrictions on choice probabilities that are already imposed in most settings are sufficient for point identification of both 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 analog 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. 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, in the class of models we describe, one can ask how consumers would choose with full consideration, and one can do so with no additional data beyond what is required to estimate conventional discrete choice models.

Our identification result builds on the insight that imperfect consideration breaks 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 are inattentive and choose the default option unless that 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 (Moshkin and Shachar 2002). While the link between imperfect attention and Slutsky asymmetry has been discussed in the theoretical literature, notably in Gabaix (2014), and noted as a source of identifying variation in Moshkin and Shachar (2002), this

approach has not yet been developed in the generality we consider.[2] Our framework implies that attempts to model consideration sets such as fixed effects in utility for products on different shelves or interactions between prices and such fixed effects can still yield misspecified models because they do not relax the symmetry assumption.

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.[3] We consider two estimation approaches: indirect inference and maximum likelihood. Maximum likelihood estimation makes less explicit the link between estimation and identification but can be computationally more feasible. To estimate the model by indirect inference, we specify a flexible auxiliary model that permits a general pattern of asymmetries, and then estimate the parameters of our consideration set model to fit them.

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 considered, matching Slutsky asymmetries enables us to accurately recover the probabilities that each good was available as well as recovering the preference parameters that we would estimate conditional on knowing which items were available. Conventional models with a comparable number of parameters misspecify own- and cross-price elasticities relative to the elasticities computed using data on which items were actually available. We formally test whether our consideration set model can generate the asymmetries captured by our flexible auxiliary model, finding that our framework cannot be rejected while ad-hoc generalizations of full-information models (e.g. allowing for good-specific price effects in a standard conditional logit) cannot explain the reduced-form patterns.

We apply our framework to analyze prescription drug insurance choices in Medicare Part D. We allow for consumers to both be completely "asleep" (and simply choose their default plan) and for consumers to attend to only a subset of options even in periods where they actively search. In this market, more than 90% of beneficiaries are inertial. This has led to proposals for a "smart default" policy, in which consumers are automatically switched into lower-cost plans but can opt out (Handel and Kolstad 2018). Evaluating this policy from both positive and normative perspectives requires us to disentangle the degree to which inertia in plan choice is driven by limited consideration or utility-relevant switching costs. If beneficiaries are inertial because it is costly for them to acclimate to a new plan, then most beneficiaries will opt out of the smart default. In fact, models with full attention

[2] 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.

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

imply that defaults have no impact at all on choices,[4] and models with some inattention but high acclimation costs imply that inattentive consumers are impacted but made worse off by having to pay these costs. Alternatively, if beneficiaries are inertial largely due to limited consideration and have low acclimation costs, they may be made better off by being defaulted to a lower cost plan. When we allow for limited consideration, we find a high degree of inattention and positive welfare effects of smart defaults, especially if we only switch consumers with cost savings at least as great as estimated acclimation costs.

The rest of this paper proceeds as follows. Section 2 situates our approach within the existing literature. Section 3 lays out our general model and identification results. Section 4 validates our approach in the lab where we observe consideration and develops an indirect inference estimator in which structural parameters are chosen to match cross-derivative asymmetries in the data. Section 5 harnesses our framework to estimate consumer preferences and limited consideration in Medicare Part D, and conducts a welfare analysis of a smart default policy in this setting. Section 6 concludes.

2 Related Literature

Exclusion Restrictions Identification of consideration set models is typically achieved by restricting which variables can influence consideration and utility. In general, point identification of all structural functions requires one to exclude a set of variables that affect consideration from utility and a set of variables that affect utility from consideration.[5] Our identification result does not rely on the existence of variables that influence consideration but not utility and vice versa. This provides a route for researchers to test generally which factors are important for consideration and utility, enabling one to, for example, distinguish between models of naive versus rational consideration. This contrasts with empirical strategies that exclude prices from consideration, and thus cannot allow consideration to be driven by the expected benefits of search (Kawaguchi, Uetake, and Watanabe 2016; Goeree 2008).

Auxiliary Data A second strand of literature identifies consideration set models using auxiliary data on which goods were considered. Conlon and Mortimer (2013) assume that consideration sets

[4] In Appendix C, we examine the robustness of our results to allowing for "paperwork costs", i.e. costs to choose a plan different from the default regardless of whether you have enrolled in that plan already. With paperwork costs, defaults may be sticky in full-consideration models. We find that our welfare conclusions are robust to the existence of paperwork costs as these costs are small relative to acclimation costs.

[5] For example, Goeree (2008), Hortaçsu, Madanizadeh, and Puller (2015), Gaynor, Propper, and Seiler (2016), and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) proceed in this way. For formal results, see the literature on the identification of mixture models (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 an ASC type model.

are known in some periods, Draganska and Klapper (2011) and Honka and Chintagunta (2016) use survey data on what products are and are not considered when choosing, and Reutskaja, Nagel, Camerer, and Rangel (2011) use eye-tracking methods to follow what options individuals consider. However, there are many scenarios where such auxiliary data does not exist but limited consideration is a first order concern. Further, many process tracking procedures measure attentional inputs but not consideration itself. For example, we may observe the rank of products in search or, perhaps, eye tracking software. As noted by Gabaix (2017), it is important to treat these as correlates of consideration rather than a direct measure of attention itself, i.e. to add such variables as determinants of the (unobserved) consideration probability. Our framework enables researchers to do this.

Theoretical Restrictions We are able to relax the assumptions usually required for identification of consideration set models by exploiting the restrictions implied by standard assumptions made in discrete choice analysis. There is a growing body of literature in behavioral decision theory that highlights the identifying power of theoretical restrictions in consideration set models (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014; Cattaneo, Ma, Masatlioglu, and Suleymanov 2018). These papers show that changes in choice probabilities that result from changes in the set of available products (an exogenous potential change to a consumer's consideration set) place restrictions on the set of preferences and consideration sets that can rationalize the data. 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 choice from every possible non-degenerate subset of feasible alternatives, while model primitives in Masatlioglu, Nakajima, and Ozbay (2012) and Cattaneo, Ma, Masatlioglu, and Suleymanov (2018) remain only partially identified in general. These insights have only been directly harnessed in experimental work in which it is possible to generate such variation (Aguiar, Boccardi, Kashaev, and Kim 2018).

Kawaguchi, Uetake, and Watanabe (2016) make the weaker assumption of "leave-one-out" variation in product availability to identify a consideration set model in order to study optimal product recommendations. Their condition relates the percentage change in product demand when a single product is unavailable to consideration probabilities using an identification at infinity argument made possible by excluding price from consideration. Our results for the ASC and hybrid models rely on variation equivalent to Kawaguchi, Uetake, and Watanabe (2016) but without the need for additional exclusion restrictions.

Our identification result harnesses the identifying power of deviations from Slutsky symmetry.[6]

[6] Davis and Schiraldi (2014) provide generalizations of multinomial logit models that permit asymmetries, but they

In a version of the DSC model that we consider in this paper, Moshkin and Shachar (2002) show that switching probabilities are more sensitive to changes in the characteristics of default plans than to non-default plans. In this paper, we prove that this variation is sufficient for identification of consideration probabilities given the assumptions made in many discrete settings and show that similar insights extend to a much richer class of models than that focused on by Moshkin and Shachar (2002).

Partial Identification While our results do not rely on additional exclusion restrictions, we do work within the structure imposed by popular models of consideration set formation. Permitting preference heterogeneity over alternatives in the population, Barseghyan, Coughlin, Molinari, and Teitelbaum (2018) leave the process generating consideration sets completely unrestricted and allow for dependence between unobservables driving consideration and utility. This approach accommodates a wider set of limited attention models than our approach, at the cost of losing point identification of the structural functions of interest. Our contribution is to show that in a framework which encompasses many specifications currently estimated in the applied economics and marketing literature, all structural functions of interest are point identified from choice probabilities and cross-price elasticities.

Alternative Assumptions on Preferences A final strand of the literature restricts the nature of preferences and preference heterogeneity for the purposes of identification. Crawford, Griffith, and Iaria (2016) show that consideration set heterogeneity can be characterized as an individual-specific fixed effect in panel data when preferences are logit. Assuming that choice sets are either stable over time (with panel data) or across individuals (with cross sectional data), preferences can be recovered from choice probabilities. In a setting where individuals only have capacity to consider a certain number of alternatives, Dardanoni, Manzini, Mariotti, and Tyson (2020) show that consideration probabilities can be identified in a setting with homogeneous preferences from a single cross section of aggregate choice shares.

In this paper, we do not impose a particular functional form on the nature of preference heterogeneity in the population. Thus, our results do not rely on the logit functional form and encompass all of the standard functional form assumptions made on preference heterogeneity. In contrast to Crawford, Griffith, and Iaria (2016), our identification result relies on the insight that changes in product characteristics alter the probability that a consumer pays attention to a particular set of

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.

products and thus unobserved choice sets can vary over time and across markets, and our result does not require panel data.

3 Model & Identification

In this section, we show that assumptions commonly made in discrete choice models are sufficient for identification in several standard models of imperfect consideration. Our central insight is that violations of Slutsky symmetry constructively identify the probability that consumers consider various subsets of products.

3.1 Basic Framework

We consider an individual i who makes a discrete choice among $J+1$ products, $\mathcal{J} = \{0, 1, ..., J\}$, with $J \geq 1$. Each product j is associated with a price, p_j. The price vector $\mathbf{p} = [p_0, ..., p_J]$ is supported on $\mathbb{R}^{J+1}_{++}$. Our framework naturally incorporates additional characteristics ($\mathbf{x}_j$), consumer microdata ($\mathbf{z}_i$), and interactions between consumer and product characteristics. However, variation in these additional characteristics is not required for our identification result and thus we will suppress the dependence of choice on $\mathbf{x}_j$ and $\mathbf{z}_i$ in what follows. Our identification argument focuses on price variation, although it extends to variation in any attribute satisfying the assumptions we state below.

The (unobserved) set of goods that a consumer considers is called the consideration set. We first present a general consideration set model to describe the relationship between asymmetries and imperfect consideration before presenting the DSC, ASC, and hybrid models. 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:

$$\mathbb{P}(j) = \{C : \{0, j\} \subseteq C \in \mathcal{P}(\mathcal{J})\} \tag{3.1}$$

In all of the models that we investigate, observed choice probabilities take the following form:

$$s_j(\mathbf{p}) = \sum_{C \in \mathbb{P}(j)} \pi_C(\mathbf{p}) s_j^\star(\mathbf{p}|C) \tag{3.2}$$

where $s_j \equiv s_j(\mathbf{p})$ is the observed probability of good j being bought given market prices $\mathbf{p}$, $\pi_C(\mathbf{p})$ gives the probability that the set of goods C is considered given observable characteristics, and $s_j^\star(\mathbf{p}|C)$ gives the probability that good j is chosen from the consideration set C. As $\pi_C(\mathbf{p})$ and

$s_j^\star(\mathbf{p}|C)$ represent proper probabilities, we have:

$$\sum_{C \in \mathcal{P}(\mathcal{J})} \pi_C(\mathbf{p}) = 1 \quad , \quad \sum_{j \in C} s_j^\star(\mathbf{p}|C) = 1 \tag{3.3}$$

In this paper, the structural objects of interest are the consideration set probabilities, $\pi_C(\mathbf{p})$, and the unobserved latent choice probabilities, $s_j^\star(\mathbf{p}|C)$. We do not directly address the identification of preference parameters given knowledge of $s_j^\star(\mathbf{p}|C)$ nor the identification of, for example, search costs given consideration probabilities in any generality. 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.

Baseline Theory Assumptions We assume that choice probabilities satisfy the standard Daly-Zachary conditions *within a consideration set* (Daly and Zachary 1978), notably cross-derivative symmetry and an absence of nominal illusion.[7]

Assumption 1. *Daly-Zachary Conditions:* unobserved latent choice probabilities, $s_j^\star(\mathbf{p}|C)$, satisfy the following conditions everywhere on $\mathbb{R}_{++}^{J+1}$:

1. Properties: $s_j^\star(\mathbf{p}|C) \geq 0$, $\sum_{j \in C} s_j^\star(\mathbf{p}|C) = 1$, and

$$\frac{\partial^J s_j^\star(\mathbf{p}|C)}{\partial p_0 ... \partial p_{j-1} \partial p_{j+1} ... \partial p_J}$$

   exists finitely, is ≥ 0, and is continuous.[8]

2. Symmetry: cross-price derivatives are symmetric:

$$\frac{\partial s_j^\star(\mathbf{p}|C)}{\partial p_{j'}} = \frac{\partial s_{j'}^\star(\mathbf{p}|C)}{\partial p_j}$$

3. Absence of Nominal Illusion:

$$s_j^\star(\mathbf{p} + \delta|C) = s_j^\star(\mathbf{p}|C)$$

[7] See Anderson, De Palma, and Thisse (1992) and Koning and Ridder (2003) for further discussion of these conditions.

[8] Intuitively, the derivative condition states that if all goods except j get more expensive, then the probability of choosing j should not decrease.

Standard additive random utility models (ARUM) imply choice probabilities that satisfy these conditions, including the nested and mixed logit models. For example, the following indirect utility function is consistent with Assumption 1:

$$u_j = v(p_j) + \epsilon_j \tag{3.4}$$

$$= \alpha_j - \beta p_j + \epsilon_j \tag{3.5}$$

where $\epsilon_j \perp p_{j'}$ for all $j, j' \in \mathcal{J}$ and the joint distribution of ϵ is absolutely continuous and non-defective.[9]

Baseline Data Assumptions In discussing identification, we treat the probability of selecting good j conditional on observables $\mathbf{p}$, $s_j(\mathbf{p})$, and all cross derivatives as known for all possible prices $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$. This is standard when addressing nonparametric identification of structural functions (Berry and Haile 2016).

Assumption 2. *Population Market Shares, Own- and Cross-Price Derivatives Observed at* $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$*:* the observables consist of the variables:

$$\left\{ s_j(\mathbf{p}), \partial s_j(\mathbf{p}) / \partial p_{j'} \right\}_{j,j' \in \mathcal{J}} \tag{3.6}$$

Loosely, we are considering a scenario in which we have enough markets (across which prices vary) and individuals within these markets such that choice probabilities and their derivatives conditional on observables can be nonparametrically estimated. However, we do not know the extent to which observed choice probabilities reflect consideration versus preferences. In practice, one rarely has enough data to nonparametrically estimate $s_j(\mathbf{p})$; 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)). Prior work on semiparametric identification of multinomial choice models without consideration sets has assumed large price support (Lewbel 2000; Matzkin 2007). This assumption therefore provides a natural benchmark for exploring identification under ideal conditions. We discuss in the text and Appendix A where limited price variation will suffice.

[9] The key restriction that consumers value price equally across choices is substantive, although it is often theoretically well-motivated.

Consideration Set Framework Given Assumption 1, in our baseline model only one mechanism is available to generate cross-derivative asymmetries: imperfect consideration.[10] Theorem 1 makes this point formally.

THEOREM 1. ASYMMETRIES & NOMINAL ILLUSION IMPLY IMPERFECT CONSIDERATION. Let observed choice probabilities have the following structure:

$$s_j(\mathbf{p}) = \sum_{C \in \mathbb{P}(j)} \pi_C(\mathbf{p}) s_j^\star(\mathbf{p}|C) \tag{3.7}$$

Given Assumption 1 and assuming that $\pi_C(\mathbf{p})$ is differentiable with respect to $\mathbf{p}$, if

$$\frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} \neq \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} \tag{3.8}$$

or

$$s_j(\mathbf{p}) \neq s_j(\mathbf{p} + \delta) \tag{3.9}$$

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

To make progress towards point identification of the structural functions of interest, we must place some additional restrictions on consideration set probabilities. If $\pi_C(\mathbf{p})$ are allowed to vary arbitrarily, then point identification of the underlying structural functions is not possible without additional information on what consumers considered (Manzini and Mariotti 2014).[11]

We derive identification results for the two most common consideration set models found in the applied literature, the 'Default Specific Consideration' (DSC) model and the 'Alternative Specific Consideration' (ASC) model, and a hybrid model that combines the two approaches.[12] The DSC

[10] Not all models of inattention generate cross-price asymmetries. For example, Matejka and McKay (2014) show that when actions are homogeneous a priori and exchangable in the decision maker's prior, and the information strategy is time invariant, a rational inattention model provides a foundation for the mulitnomial logit (which yields symmetric cross-derivatives). Models in which consideration is independent of product characteristics would also have observed choice probabilities that satisfy the Daly-Zachary conditions, provided that the latent choice probabilities satisfy those conditions.

[11] Several 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, Coughlin, Molinari, and Teitelbaum (2018).

[12] Compared to some fully microfounded models of inattention, the DSC and ASC models can permit more general patterns of behavior. For example, a rational inattention model imposes that product attributes should impact attention in proportion to their value, but this need not be the case. However, our agnosticism means that we cannot identify how out of sample variation will impact attention and our approach alone does not facilitate the identification of search costs without further structure. One could use the search probabilities recovered by our model to help

model assumes the existence of an inside default good and allows the probability of considering all alternative options to vary only as a function of the characteristics of that default (Moshkin and Shachar 2002; Ho, Hogan, and Scott Morton 2015; Heiss, McFadden, Winter, Wupperman, and Zhou 2016; Hortaçsu, Madanizadeh, and Puller 2015).[13] The ASC model permits each good to have an independent probability of being considered that depends on the characteristics of that good (Goeree 2008; Manzini and Mariotti 2014; 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).[14]

The ASC and DSC models impose substantive restrictions on the data (even when combined into a hybrid model). First, both models impose that the unobservable determinants of attention and utility are uncorrelated. This restriction can be relaxed but not without additional instruments (see Appendix A.7). Second, neither model allows for the possibility of correlated unobservable shocks to attention probabilities. Third, the models require at least one restriction on how attributes of goods are allowed to perturb attention probabilities for rival goods. These restrictions are not without loss and their plausibility must be assessed in a context-specific way. 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 cross-derivative asymmetries.

3.2 The Default Specific Model

Under the DSC model, the market shares of the default (good 0) and non-default goods take the form:

$$\begin{aligned} s_0(\mathbf{p}) &= (1-\mu(p_0)) + \mu(p_0)s_0^\star(\mathbf{p}|\mathcal{J}) \\ s_j(\mathbf{p}) &= \mu(p_0)s_j^\star(\mathbf{p}|\mathcal{J}) \quad \text{for } j>0 \end{aligned} \tag{3.10}$$

Where the differentiable function $\mu(p_0)$ gives the probability of considering all available products, while $s_j^\star(\mathbf{p}|\mathcal{J})$ gives choice probability for good $j \in \mathcal{J}$ conditional on considering all products.

Note that for simplicity, in the main text we assume:

identify a structural model which would identify these costs. In Section 5, we highlight these points and evaluate the robustness of our normative evaluation to alternative assumptions about these unknown values.

[13]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.

[14]The ASC model is also supported by direct empirical evidence: Aguiar, Boccardi, Kashaev, and Kim (2018) conduct a lab experiment in which they observe choices from every possible subset of products and can thus recover flexible consideration set probabilities, finding that consideration patterns in the data can be rationalized by the ASC model with independent choice probabilities.

1. A homogeneous default good; this is 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.

2. That $\mu(\cdot)$ is a function of the characteristics of the default good only; this is to follow the existing literature. We show in Appendix A that our results extend without complication to the case where $\mu(\cdot)$ is also function of the characteristics of any strict subset of non-default goods.[15]

Identifying Changes in Consideration Probabilities The key to our identification argument is that full consideration implies symmetric cross-price derivatives in standard discrete choice models (Slutsky symmetry). However, with imperfect consideration, cross derivative asymmetries arise. Differentiating Equation 3.10 and using the fact that the market shares conditional on full consideration satisfy symmetry, we obtain:

$$\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} = \mu_0 \frac{\partial s_j^\star(\mathbf{p}|\mathcal{J})}{\partial p_0} + \frac{\partial \mu_0}{\partial p_0} s_j^\star(\mathbf{p}|\mathcal{J}) - \mu_0 \frac{\partial s_0^\star(\mathbf{p}|\mathcal{J})}{\partial p_j} \tag{3.11}$$

$$= \frac{\partial \mu_0}{\partial p_0} s_j^\star(\mathbf{p}|\mathcal{J}) \tag{3.12}$$

$$= \frac{\partial \log(\mu_0)}{\partial p_0} s_j(\mathbf{p}) \tag{3.13}$$

where $\mu_0 \equiv \mu(p_0)$ and the last line uses the fact that $s_j^\star(\mathbf{p}|\mathcal{J}) = s_j(\mathbf{p})/\mu_0$. Thus, changes in the probability of full consideration are directly identified from data on choice probabilities:

$$\boxed{\frac{\partial \log(\mu_0)}{\partial p_0} = \frac{1}{s_j(\mathbf{p})} \left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} \right]} \tag{3.14}$$

Intuitively, if the price of the default plan perturbs consideration 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 marketing and health insurance literature by Ho, Hogan, and Scott Morton (2015) and Moshkin and Shachar (2002).

Theorem 2. Identification of Changes in DSC Consideration Probabilities. Given Assumptions 1 and 2, then $\partial \log(\mu_0)/\partial p_0$ is constructively identified.

[15] See Section A.6.

Identifying the Level of Consideration In recovering the derivative of the log consideration probability, μ_0 is identified up to a constant by integrating over the support of p_0:

$$\log\left(\mu(\infty)\right) - \log\left(\mu(\tilde{p}_0)\right) = \int_{\tilde{p}_0}^{\infty} \frac{1}{s_j(\mathbf{p})} \left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j}\right] \, \mathrm{d}p_0 \tag{3.15}$$

Identifying the level of consideration (and thus latent market shares) requires an additional assumption to pin down the constant of integration. Assuming that consumers are prompted to consider good j when p_0 reaches an extreme value enables the level of consideration to be identified (i.e. $\log\left(\mu(\infty)\right) = 0$ in Equation 3.15). This assumption is analogous to those made in the literature on nonparametric identification of multinomial discrete choice models (Berry and Haile (2009), Lewbel (2000)), treatment effects (Heckman and Vytlacil (2005), Lewbel (2007), Magnac and Maurin (2007)), the identification of binary games and entry models (Tamer (2003), Fox, Hsu, and Yang (2012), Lewbel and Tang (2015)), and the use of special regressors more generally. This assumption is testable in our setting by checking that cross derivative differences are symmetric at high default prices.

Assumption DSC. As $p_0 \to \infty$, $\mu(p_0) \to 1$.

Theorem 3. Identification of $\mu(p_0)$ in the DSC Model. Given Assumption 1, 2, and DSC, then consideration probabilities are constructively identified as:

$$\boxed{\mu(\tilde{p}_0) = \exp\left(-\int_{\tilde{p}_0}^{\infty} \frac{1}{s_j(\mathbf{p})} \left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j}\right] dp_0\right)} \tag{3.16}$$

Note that nonparametric identification of $\mu(p_0)$ requires substantially less observed price variation than that implied by Assumption 2.[16] Furthermore, commonly employed functional form assumptions on consideration probabilities substantially reduce the amount of price variation required to identify the level of consideration,[17] even when no further parametric assumptions are placed on preferences. For example, let consideration take the following form:[18]

$$\mu(p_0) = \frac{\exp\left(\gamma_0 + \gamma_p p_0\right)}{1 + \exp\left(\gamma_0 + \gamma_p p_0\right)} \tag{3.17}$$

[16] We require that there exists for each $p_0 \in \mathbb{R}_{++}$ a good j and price vector $\mathbf{p}_{\neg 0} = \{p_k\}_{k \in \mathcal{J}/0}$ such that $\{s_j(p_0, \mathbf{p}_{\neg 0}), \partial s_j(p_0, \mathbf{p}_{\neg 0})/\partial p_0, \partial s_0(p_0, \mathbf{p}_{\neg 0})/\partial p_j\}$ exist and are observed.

[17] For examples of papers using this functional form assumption seee Heiss, McFadden, Winter, Wupperman, and Zhou (2016) for the DSC model and Goeree (2008) for the ASC model.

[18] This is the assumption we will be making in our empirical applications.

In this scenario, all that is required is for there to exist at least two levels of the default price at which market shares and cross derivatives are observed.

THEOREM 4. IDENTIFICATION OF $\mu(p_0)$ WITH LOGIT CONSIDERATION IN THE DSC MODEL. Given Assumption 1 and two strictly positive price vectors $\{\mathbf{p}^a, \mathbf{p}^b\}$ with $p_0^a \neq p_0^b$ at which market shares and cross-price derivatives are observed, then $[\gamma_0, \gamma_p]$ are identified where

$$\boxed{\mu(p_0) = \frac{\exp\left(\gamma_0 + \gamma_p p_0\right)}{1 + \exp\left(\gamma_0 + \gamma_p p_0\right)}} \tag{3.18}$$

Proof in Appendix A.

3.3 The Alternative Specific Model

Under the ASC approach, consideration set probabilities take the form:

$$\pi_C(\mathbf{p}) = \prod_{j \in C} \phi_j(p_j) \prod_{j' \notin C} \left(1 - \phi_{j'}(p_{j'})\right) \tag{3.19}$$

where the probability of good j being considered, $\phi_j \equiv \phi_j(p_j)$, is a differentiable function of own characteristics only and $\phi_j(p_0) = 1$ for all p_0. Observed market shares then take the form:

$$s_j(\mathbf{p}) = \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l(p_l) \prod_{l' \notin C} \left(1 - \phi_{l'}(p_{l'})\right) s_j^\star(\mathbf{p}|C) \tag{3.20}$$

Even in this more complicated model, changes in consideration probabilities can be expressed as a function of observable differences in cross-derivatives and market shares. This is despite the fact that 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 the theoretical framework.

Identifying Consideration Probabilities Even in this richer setting, changes in consideration probabilities can be expressed as a linear function of observables. Let $\bar{\mathbf{p}}_j$ give the price vector $\mathbf{p}$ under which p_j is approaching ∞ with $s_j(\bar{\mathbf{p}}_j) = 0$.[19] As shown fully in Appendix A, one can express cross derivative differences between default and non-default products as:

$$\frac{\partial s_0(\mathbf{p})}{\partial p_j} - \frac{\partial s_j(\mathbf{p})}{\partial p_0} = \frac{\partial \log(\phi_j)}{\partial p_j} \left(s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j)\right) \tag{3.21}$$

[19] Please see Appendix A for a formal discussion of $\bar{\mathbf{p}}_j$.

While cross derivative differences for $j, j' \neq 0$ are:

$$\frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} - \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} = \frac{\partial \log(\phi_{j'})}{\partial p_{j'}} \left(s_j(\mathbf{p}) - s_j(\bar{\mathbf{p}}_{j'})\right) - \frac{\partial \log(\phi_j)}{\partial p_j} \left(s_{j'}(\mathbf{p}) - s_{j'}(\bar{\mathbf{p}}_j)\right) \tag{3.22}$$

Equations 3.21 and 3.22 relate unobservable changes in consideration probabilities to observed cross derivative differences and market shares.

For simplicity, in the main text we provide the just-identified conditions for an inside-default. In Appendix A, we give identification results based on the full system of cross derivative differences defined by Equation 3.21 and 3.22, which are also suitable for scenarios where the default is the outside good. Analogous arguments to those made with respect to the DSC model can be employed to prove identification of the level of $\phi_j(\cdot)$. Without assuming a particular functional form assumption for $\phi_j(\cdot)$, one must continue to rely on large support assumptions for nonparametric identification; namely, that consumers are prompted to consider a product with probability one at extreme values of the covariates (Assumption ASC.i).[20] These data requirements are, however, reduced when one assumes a parametric form for consideration probabilities. Assumption ASC.ii imposes that there is some substitution to good-0 when the price of non-default goods is high; a weak assumption that is easily tested.

Assumption ASC.i: As $p_j \to \infty$, $\phi_j(p_j) \to 1$.

Assumption ASC.ii: $s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j) \neq 0$ at all $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$.

Theorem 5. Identification of $\phi(p_j)$ in the ASC Model. Given Assumption 1, 2, ASC.i, and ASC.ii, then $\phi_j(p)$ for $j = 1, ..., J$ are identified at all $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$:

$$\begin{aligned} \frac{\partial \log(\phi_j)}{\partial p_j} &= \frac{\frac{\partial s_0(\mathbf{p})}{\partial p_j} - \frac{\partial s_j(\mathbf{p})}{\partial p_0}}{s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}})} \\ \phi_j(\tilde{p}_j) &= \exp\left(- \int_{\tilde{p}_j}^{\infty} \frac{\frac{\partial s_0(\mathbf{p})}{\partial p_j} - \frac{\partial s_j(\mathbf{p})}{\partial p_0}}{s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j)} \, \mathrm{d}p_j \right) \end{aligned} \tag{3.23}$$

[20] Note that this is to pin down the constant of integration. An alternative assumption that locates the level of consideration might be more natural in some applications (e.g. consider a good if the level of an attribute falls to a particularly low level).

3.4 The Hybrid Consideration Set Model

The assumptions made to identify consideration probabilities in the ASC model are also sufficient to identify a hybrid model that combines the ASC and DSC models. Combining the models seems natural in many applied settings. For example, with many options, it seems implausible that consumers would consider all goods in the feasible choice set upon "waking up". However, conditions for the identification of this combined hybrid framework have not been previously investigated nor has the framework been harnessed for applied research until this paper.

Under the hybrid consideration set model, the market shares of the default and non-default goods take the form:

$$\begin{aligned} s_0(\mathbf{p}) &= (1-\mu(p_0)) \quad + \quad \mu(p_0)\textstyle\sum_{C\in\mathbb{P}(0)}\prod_{l\in C}\phi_l(p_l)\prod_{l'\notin C}(1-\phi_{l'}(p_{l'}))\, s_0^\star(\mathbf{p}|C) \\ s_j(\mathbf{p}) &= \qquad\qquad\qquad\qquad \mu(p_0)\textstyle\sum_{C\in\mathbb{P}(j)}\prod_{l\in C}\phi_l(p_l)\prod_{l'\notin C}(1-\phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C) \quad \text{for } j>0 \end{aligned} \tag{3.24}$$

where $\phi_0(p_0) = 1$ for all $p_0 \in \mathbb{R}_{++}$. Restricting $\phi_j(p_j) = 1$ for all $j > 0$ gives the DSC model. Restricting $\mu(p_0) = 1$ gives the ASC model.

In this model, cross-derivative differences between non-default goods have the same structure as in the ASC model (Equation 3.22). However, cross-derivatives involving the default good take a slightly different form because of the impact of the default good on the probability of waking up:

$$\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} = \frac{\partial \log(\mu_0)}{\partial p_0} s_j(\mathbf{p}) - \frac{\partial \log(\phi_j)}{\partial p_j}\left(s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j)\right) \tag{3.25}$$

Let the system of equations defined by Equations 3.22 and 3.25 be expressed as:

$$c(\mathbf{p}) = D(\mathbf{p})A(\mathbf{p}) \tag{3.26}$$

where $c(\mathbf{p})$ is a $\frac{1}{2}J(J+1)$-vector of cross derivative differences at prices $\mathbf{p}$, $D(\mathbf{p})$ is the coefficient matrix of market share differences, and $A(\mathbf{p}) = [\partial \log \mu_0/\partial p_0, \partial \log \phi_1/\partial p_1, ..., \partial \log \phi_J/\partial p_J]$ is the $J+1$-vector of log consideration probability derivatives.[21] As there are typically more than $J+1$ cross-derivative differences, it is convenient to work with the system:[22]

$$D'(\mathbf{p})c(\mathbf{p}) = D(\mathbf{p})'D(\mathbf{p})A(\mathbf{p}) \tag{3.27}$$

[21] See Appendix A for illustrations of the structure of these matrices.
[22] Alternative weighting matricies, W_m, can be used: $D_m' W_m D_m$.

If $D'(\mathbf{p})D(\mathbf{p})$ is full rank, there is a unique solution to this system and changes in consideration probabilities are uniquely identified from choice data. Appendix A discusses the restrictions on structural functions required for this rank condition to hold. Intuitively, goods' being imperfect substitutes and being considered with strictly positive probability at all price vectors will suffice. A strength of our approach is that the rank condition is testable given market share data. When it fails, changes in consideration probabilities can be recovered at price vectors where the assumption holds.

Assumption Hybrid. *Rank Condition*: The matrix $D'(\mathbf{p})D(\mathbf{p})$ is full rank at all $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$.

Theorem 6. (Identification of Consideration Probabilities in the Hybrid Consideration Set Model) Given Assumption 1, 2, DSC, ASC.i, ASC.ii, and Hybrid, then $\mu(p_0)$ and $\phi_j(p_j)$ are identified at all $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$. Proof in Appendix A.

3.5 Identifying Latent Market Shares

We have focused our attention so far on the identification of consideration probabilities. In the DSC model, identification of latent market shares is trivial once $\mu(p_0)$ is known as $s^\star_j(\mathbf{p}) = \mu(p_0)/s_j(\mathbf{p})$. In the ASC and Hybrid frameworks, however, matters are more complicated given that there exist 2^J independent latent choice probabilities for any given price vector.[23]

In Appendix A, we show how the restrictions deriving from "nominal illusion" facilitate the identification of the 2^J independent latent choice probabilities in the ASC and hybrid models, $s^\star_j(\mathbf{p}|C)$. To provide intuition for our result, imagine a rise in all prices by some amount δ such that relative prices remain unchanged. Given the Daly-Zachary conditions, this price shift can change consideration probabilities but does *not* alter latent choice probabilities conditional on consideration. Thus, observed market shares can vary even though nominal illusion would suggest invariance of choices to price changes that do not alter relative prices. Hastings and Shapiro (2013) establish this behavioral pattern for gasoline choice. In Appendix A, we show how this variation is sufficient to identify latent choice probabilities.

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

3.6 Overidentification

Given our assumptions, imperfect consideration is the only mechanism giving rise to asymmetric cross-derivatives. Relaxing our background assumptions might, however, give rise to alternative sources of asymmetry that our framework could incorrectly attribute to inattention. We note that our model is over-identified, providing the potential to test the validity of the consideration set model outlined in this paper, and that the asymmetries predicted by our framework have a particular structure.[24] With $J > 2$, the derivative of the log of consideration probabilities are over-identified; intuitively there are more cross-price derivative differences than consideration probabilities. In the hybrid model, for example, there are

$$\underbrace{\frac{1}{2}J(J+1)}_{\#\text{ Independent Cross Deriv. Diffs}} - \overbrace{(J+1)}^{\#\ \phi_j\ \text{Derivatives}} \tag{3.28}$$

overidentifying restrictions for changes in consideration probabilities.[25]

4 Validation & Estimation

In this section, we discuss estimation of our model and validate our approach. To do so, we must observe "true" consideration probabilities; something which is very rare in observational data. We thus validate the practical relevance of our identification result in the lab. We ask consumers to make choices from *known* subsets of 10 goods that are generated according to the ASC model. We ask whether we can recover the (known) consideration probabilities as well as preferences conditional on consideration using information only on observed choices. This test goes beyond a simulation exercise by showing that we can use our model to recover consideration probabilities in a setting where our functional form assumptions on latent choice probabilities (i.e. preferences) are not guaranteed to hold.

4.1 Set Up

To experimentally validate our approach, we conducted a discrete-choice consumption experiment with 149 Yale students.[26] We selected 10 goods sold at the Yale Bookstore with list prices ranging

[24] This also facilitates testing against spurious asymmetries that might arise due to data issues such as measurement error.

[25] Similar reasoning shows that latent market shares are over-identified when there are more shifts in prices than latent market shares.

[26] There were 150 participants in total, but one participant's data was not recorded properly because they refreshed the browser several times during the experiment which interfered with data capture– this participant is dropped from

from \$19.98-\$24.98. These goods and their list prices are shown in Table 7 in Appendix B.5. 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). A sample product selection screen is shown in Figure 1. Consumers were shown images of all the products in the displayed subset set along with the (randomly chosen) prices and asked to select their preferred option. 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.[27] In total, we ran the experiment with 149 participants, resulting in 7,450 choices.[28]

We treat the subset of products that appear on a respondent's screen as the consideration set.[29] The probability that each good appeared on the screen was fixed by us in advance – this probability varied across goods and with prices such that goods were more likely to be considered (i.e. appear on a respondent's screen) if they had a higher price (perhaps mimicking the behavior of a retailer who places their highest margin products where they are most likely to be noticed).[30] We specified the probability that good j was in a participant i's round r consideration set as:[31]

$$\phi_{ijr} = Pr\left(\gamma_j + p_{irj}\gamma_p - \eta_{irj} > 0\right) \tag{4.1}$$

$$= \frac{\exp(\gamma_j + p_{irj}\gamma_p)}{1 + \exp(\gamma_j + p_{irj}\gamma_p)} \tag{4.2}$$

where η_{irj} is distributed logistic, p_{irj} gives the product's price, and γ_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. Our key question is whether we can accurately recover these parameters by harnessing asymmetries in cross-price responses.

Figure 1 Here

the final analysis.

[27] Participants had to chose one item and thus could not simply take the \$25 payment. While this does not introduce a bias to our results as participants ranking over the remaining options are unaffected, we do not think that many of our participants would have simply taken the cash if given the chance. First, no participants chose the lowest priced item in every round. Second, prices in the experiment were typically lower than the list price creating arbitrage opportunities.

[28] 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 8 reports results using only this subset of users who passed this test and shows that results are qualitatively unchanged.

[29] 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.

[30] 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.

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

4.2 Estimation

Existing applications of consideration set models are typically estimated by maximum likelihood (Goeree 2008).[32] For validation purposes, the principle downside of this approach is the lack of transparency regarding what variation is driving our results. Are the estimated consideration probabilities driven by the asymmetries in the choice probabilities or by parametric assumptions made in specifying the model? To address this question directly, we estimate the model by indirect inference in addition to the more conventional maximum likelihood approach.[33]

Indirect Inference 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 (Smith 1993; Gourieroux, Monfort, and Renault 1993). Following Keane and Smith (2003), we define a flexible auxiliary model characterized by the parameter vector θ. Our identification proof points to the importance of specifying an auxiliary model that permits asymmetric cross elasticities.[34] We specify a flexible logit model, where we begin with a conventional logit model with good-specific price coefficients and then also add additional interaction terms between the prices of alternative goods to capture flexibly asymmetries. That is, we specify the reduced-form auxiliary model for choice probabilities, $\widetilde{s}_{irj}$, as follows:

$$\widetilde{s}_{irj} = \frac{\exp\left(\widetilde{u}_{irj}\right)}{\sum_k \exp\left(\widetilde{u}_{irk}\right)} \tag{4.3}$$

$$\widetilde{u}_{irj} = \theta_j + \theta_j^0 p_{irj} + \sum_{j'} \theta_{jj'} p_{irj} p_{irj'} \tag{4.4}$$

In our experimental setting with 10 goods, this specification gives rise to an auxiliary model with 119 parameters.

In our auxiliary model, the set of $\theta_{jj'}$ parameters capture asymmetric cross-derivatives. This

[32] Please see our stata command "alogit" for estimation by maximum likelihood.

[33] Our result is constructive, so nonparametric estimation of choice probabilities is possible in principle; in practice, the curse of dimensionality renders this approach infeasible in most applied settings of interest. A further consideration for applied researchers is that of endogeneity. We do not consider this in detail in this paper given our focus on the identification issues arising from imperfect consideration alone. Goeree (2008) considers estimation of the ASC model in the presence of price endogeneity. More generally, our results show that if instruments can be used to identify the structural derivatives of choice probabilities with respect to product attributes conditional on any unobserved correlate of product attributes, then one can identify consideration probabilities. Estimation of these structural derivatives is non-trivial and we intend to address it in future work.

[34] 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.

specification has several desirable properties. We show in Appendix B that consideration set models 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.[35] 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 zero (which yields the logit choice probabilities).[36] Additionally, this specification nests the conventional logit model, which yields a symmetric substitution matrix, as a special case.

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.[37]

We estimate the auxiliary model using our experimental data to obtain parameter estimates $\hat{\theta}$:

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(y; p, \theta) \tag{4.5}$$

where

$$\mathcal{L}(y; p, \theta) = \sum_i \sum_r \sum_j 1(y_{ir} = j) \log \widetilde{s}_{irj}(p_{ir}; \theta) \tag{4.6}$$

where $y_{ir} \in \mathcal{J}$ records which option consumer i bought in round r of the experiment and y gives this choice variable stocked across rounds and individuals, and p gives the vector of prices, p_{irj}, stacked across individuals, rounds, and options.

Given prices p and structural parameters ψ, we use our consideration set model to simulate M statistically independent simulated data sets, $\{\widetilde{y}^m(\psi)\}^{m=1,\ldots,M}$, by redrawing the structural error terms from their parametric distributions. In our empirical applications, we assume that the additive random error term in preferences, ϵ_{ijr}, is distributed iid Type 1 Extreme Value. While this is

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

[36] Note 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.

[37] 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 constants (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 (variable) attributes of rival goods.

(intentionally) a restrictive model of preferences, we ask if we are able to recover the process generating consideration sets even when preferences are modelled in this simplistic way. The auxiliary model is then estimated on each of the M simulated data sets to obtain a set of estimated parameter vectors $\widetilde{\theta}^m(\psi)$. Formally, $\widetilde{\theta}^m(\psi)$ solves:

$$\widetilde{\theta}^m(\psi) = \arg\max_{\theta} \mathcal{L}(\widetilde{y}^m(\psi); p, \theta) \tag{4.7}$$

where $\mathcal{L}(\cdot)$ is defined as in Equation 4.6.

Indirect inference generates an estimate $\widehat{\psi}$ 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 $\widetilde{\theta}(\psi) = M^{-1} \sum \widetilde{\theta}^m(\psi)$. Formally, $\widehat{\psi}$ solves:

$$\widehat{\psi} = \arg\min_{\psi} \left(\hat{\theta} - \widetilde{\theta}(\psi)\right)' W \left(\hat{\theta} - \widetilde{\theta}(\psi)\right) \tag{4.8}$$

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 ψ. As the sample size grows large, $\hat{\theta}$ and $\widetilde{\theta}(\psi)$ both converge to the same "pseudo true" value, θ_0, underlying the consistency of the approach (Gourieroux, Monfort, and Renault 1993).

4.3 Results

Table 1 shows the results of our validation exercise. In column 4, we give the 'true' consideration and preference coefficients. The true consideration coefficients are known with certainty given that they were specified by us (Equation 4.2). In the case of preference parameters, we take conditional logit parameters estimated by maximum likelihood using information on the *actual* choice sets that consumers faced as the relevant comparator, i.e. we use all information on what respondents saw.[38]

Columns 1-3 of Table 1 report parameter estimates that only use information about the product consumers actually chose and not information about the specific subset of 10 goods they could choose from in each instance. In Appendix B.5, we also analyze the ability of the models to capture price elasticities. First, consider the maximum likelihood preference parameters shown in the top panel of Table 1. The conditional logit model assuming a considered choice set of all 10 goods gives

[38] We take a simple conditional-on-consideration conditional logit specification as the benchmark throughout; this is a natural generalization of the specification estimated in Column (1), which uses no information on what items were shown to respondents. Appendix B.2 shows that this model fits the choice data well within consideration sets and serves as a good benchmark for comparison.

a price effect of -0.05, less than a third of the true value. 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. Further, the conditional logit fixed effects are systematically biased because they conflate attention and utility; products that are rarely in the considered choice set are assumed to be low utility.

In contrast, our consideration set models are able accurately to recover the process generating consideration and imply much more elastic price responses. We give results estimated by both maximum likelihood and our indirect inference strategy. Both sets of results imply confidence intervals on the consideration price effect that include the true value. The consideration set models also recover preference 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.

Using our indirect inference strategy, we cannot reject the null hypothesis that the ASC model can explain the patterns reflected in the auxiliary model (p-value = 0.4182).[39] The restriction that consideration is independent of price is also rejected at all conventional confidence levels as is a full-consideration model with good-specific price coefficients (which thus allows for asymmetric cross-derivatives, although not in the manner predicted by limited consideration models). See Appendix B.5 for more details. These results provide strong evidence that our approach can discriminate between consideration and preferences using real choice data.

Table 1 here

Figure 2 here

5 Limited Consideration & Smart Defaults in Medicare Part D

Disentangling the degree to which choice behavior reflects limited consideration versus preferences is important for market and policy design, especially so in the case of insurance. A large literature

[39] Appendix B gives the formal details of the goodness-of-fit test. In summary, the minimized value of the objective function is distributed chi-squared with degrees of freedom equal to the difference in the number of parameters in the auxiliary versus structural model. In Appendix B.5, we also report that an alternative full-consideration model with good-specific price parameters (which can generate cross-derivative asymmetries, although ones which have a different structure to a limited consideration model) is rejected at all conventional significance levels (p-value =0.0000; $\chi^2 = 2,279$).

finds that consumers choosing insurance plans fail to minimize costs and make systematic errors (Abaluck and Gruber 2011; Heiss, Leive, McFadden, and Winter 2013; Bhargava, Loewenstein, and Sydnor 2015; Abaluck and Gruber 2016). Consumers are also highly inertial (Handel 2013); in Medicare Part D, where elderly consumers choose prescription drug insurance, over 90% of returning consumers choose the same plan as the previous year.

In response to such behavior, Handel and Kolstad (2018) propose a "smart default" policy, under which consumers would be automatically assigned to lower cost plans but given the option of switching back. Whether such a policy will make consumers better off depends on whether the associated cost savings outweigh the "acclimation costs" of learning to navigate a new plan.[40] Acclimation costs include costs such as scheduling deliveries for mail-order drugs or learning which of several chemically equivalent drugs are covered by any given plan. A smart default policy has not yet been implemented. In this paper, we harness our framework to estimate limited consideration in Medicare Part D plan choice, allowing us to predict how consumers would respond to the policy and to normatively evaluate the results.[41]

Contribution of Our Approach We model limited consideration using the hybrid version of our model. This allows consumers to exhibit two types of inattention: they can be "asleep", in which case they remain enrolled in the plan they chose last year (the DSC model), or they can be "awake" and actively choose but attend to only a subset of options (the ASC model). This has important consequences for the evaluation of a smart default policy. In our model, full consideration implies that smart defaults will have no impact on behavior and acclimation costs must be huge to rationalize observed inertia.[42] Allowing some consumers to be asleep means consumers will stick with the smart default, and cost savings may or may not outweigh acclimation costs. Allowing consumers to consider only a subset of options when awake lowers the acclimation costs necessary to rationalize why a disproportionate share of awake individuals choose the default. Thus, ignoring inattention of either type will cause us to misstate acclimation costs and misstate how sticky smart defaults will be.

[40] In Appendix C, we consider two alternative drivers of inertia. The first is "spurious state dependence", wherein inertia arises because chosen plans are desirable for unobserved reasons. The second is "paperwork costs", where consumers have a cost to choosing any plan that is not the default regardless of whether they have previously chosen that plan.

[41] Low-income subsidy beneficiaries have been defaulted into plans with low premiums, although these beneficiaries face substantially less differentiation than regular beneficiaries due to subsidized cost-sharing. In Appendix C, we use variation from this defaulting among low-income subsidy beneficiaries to aid identification.

[42] In Appendix C, we examine the robustness of our results to allowing for paperwork costs, i.e. costs in arranging to switch plans. This provides a route for a smart default policy to change choice probabilities and cause consumers to stick to the smart default, even in full-consideration models. We find that our welfare conclusions are robust to the existence of paperwork costs as these costs are small relative to acclimation costs.

The most directly relevant work in the existing literature is Heiss, McFadden, Winter, Wupperman, and Zhou (2016), which also attempts to decompose inertia into inattention and utility relevant factors. Our analysis goes beyond this work in three ways. First, we allow for consumers to attend to only a subset of plans conditional on being awake. Second, we relax assumptions concerning which characteristics are utility-relevant and which are attention-relevant. We show in Appendix C.3 that exclusion restrictions relied upon in prior work are often rejected.[43] Finally, we consider the implications of our estimates for the (partial-equilibrium) welfare effect of a proposed "smart default" policy.

5.1 Context & Data

Medicare Part D plans provide prescription drug insurance to elderly beneficiaries in the United States. The program was created in 2006 in response to increased spending on pharmaceuticals creating large out of pocket costs for elderly Medicare recipients who at the time had no prescription drug coverage. Our analysis focuses on stand-alone prescription drug insurance plans (PDP plans) and we do not consider plans that provide broader medical insurance ("Medicare Advantage"). Our main analysis sample consists of 100,000 randomly chosen non-dual beneficiaries enrolled in stand-alone plans in 2008-2009. We restrict the sample to beneficiaries for which we observed a prior year plan (dropping 17.5% of beneficiaries). To manage the computational burden of estimating alternative-specific attention parameters, we also restrict our sample only to include plans whose market share is at least 1.5% of plans available in the state in question. This restriction causes us to drop an additional 4% of beneficiary-years, leaving us with a maximum of 17 plans per choice set and 79,286 beneficiaries.

Table 2 gives the key summary statistics for our sample. The main plan attributes that we observe are annual premiums, deductibles, indicators for whether plans provide coverage in the Part D "donut hole", the number of the top 100 drugs included in the formulary, and a quality rating.[44] Additionally, we follow Abaluck and Gruber (2016) to construct what a beneficiary would pay out-of-pocket for their claims in all alternative plans in their choice set.[45] The average beneficiary in our

[43] In Appendix C.3, we explicitly test the alternative restrictions imposed by Heiss, McFadden, Winter, Wupperman, and Zhou (2016) to identify attention probabilities in this setting. These restrictions include assuming that changes in plan attributes do not matter for utility conditional on levels and that demographics such as age impact attention probabilities but not preferences. We statistically reject these assumptions, although we find that models that impose them produce similar attention probabilities (in all cases, cross-derivative asymmetries are contributing to the identification of these probabilities).

[44] The donut hole is a gap in coverage included in Part D to lower the fiscal cost of the program. When beneficiaries exceed an "initial coverage limit" (shifting over time between $2,000 and $4,000), they must then pay the full cost for the next several thousand dollars in drug costs until reaching the catastrophic coverage threshold (which also varies by year). Some plans offer additional coverage in the donut hole in exchange for higher premiums. The quality rating is based on customer service, member complaints, and "member experience" with the drug plan.

[45] To account for uncertainty, we match each individual to 2,000 beneficiaries in the same decile of expenditures in

sample pays \$839 in annual out of pocket costs and \$415 in plan premiums each year. On average, after our sample restrictions, beneficiaries face a choice of 11.9 plans in a given year. Switching between plans is rare in our sample: 93.7% stick with the same plan that they observed purchasing the previous year (the default plan). This is in line with previous studies of Medicare Part D choice behavior (Abaluck and Gruber 2016; Heiss, McFadden, and Winter 2010). Seven out of ten Medicare beneficiaries enrolled in these plans during all four annual open enrollment periods from 2006 to 2010 did not voluntarily switch plans in any of these periods (Hoadley, Hargrave, Summer, Cubanski, and Neuman 2013).

Table 2 here

Reduced form evidence Before giving our structural estimates, we first provide reduced form evidence that some inertia is driven by inattention. We treat the plan that an individual chose in the previous year as the default plan. Following Section 3, inattention and utility-driven switching costs are separately identified by asymmetries in how the decision to remain inertial depends on default vs. rival plan characteristics.[46] To test for such asymmetries, 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).

$$y_{it} = x_{idt}\alpha_d + (x_{idt} - \bar{x}_{ijt})\alpha_x + \delta_i + \delta_t + e_{it} \tag{5.1}$$

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. We consider share weighted averages (using the choices of new beneficiaries to construct shares), as well as unweighted averages among the three lowest cost plans and setting $\bar{x}_{ijt}$ directly equal to attributes of the lowest cost plan. We report the α_d coefficients, which test whether default attributes are weighted more heavily in switching decisions than rival attributes.

The results are shown in Table 3. In all specifications, switching decisions are significantly more sensitive to default premiums and deductibles than to rival attributes (for premiums, the sensitivity to the default is almost three times that of rival attributes). This asymmetry is consistent with the findings of Ho, Hogan, and Scott Morton (2015), who find that consumers do not respond to changes in premiums of the lowest cost plan, the lowest cost plan within the same brand, nor the average of the five lowest cost brands. These patterns of variation are consistent with a model where

the previous year and then compute the mean and variance of out of pocket costs in each plan for all such beneficiaries. More details of this procedure can be found in Abaluck and Gruber (2016).

[46] See also Moshkin and Shachar (2002).

many consumers do not actively search each period but are 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. A notable exception to this pattern is that the coefficients in Table 3 suggest that consumers are more sensitive to donut hole coverage of *rival* plans. In our hybrid model, this pattern can be rationalized if consumers are especially likely to attend to attributes of rival plans that offer donut hole coverage.

Table 3 here

5.2 Choice Model

To quantify the importance of limited consideration versus utility in rationalizing plan choice, we estimate the hybrid model (Section 3.4). Recall that in the hybrid model, consumers are either asleep and choose the default good or, if the default good becomes sufficiently unsuitable, they "wake up" and make an active choice. Conditional on waking up, however, consumers attend only to some of the available options, with the probability of attending to each option depending on the attributes of that option. Following the existing literature, we assume that consumers compare plans only in the current year when making a choice.[47]

As discussed in detail in Section 3, the probability of selecting option j, $s_j(\cdot)$ is expressed as:

$$\begin{aligned} s_0(\mathbf{x}) &= \mu(\mathbf{x}_0) \sum_{C \in \mathbb{P}(0)} \prod_{l \in C} \phi_l(\mathbf{x}_l) \prod_{l' \notin C} (1 - \phi_{l'}(\mathbf{x}_{l'}))\, s_0^\star(\mathbf{x}|C) + (1 - \mu(\mathbf{x}_0)) \\ s_j(\mathbf{x}) &= \mu(\mathbf{x}_0) \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l(\mathbf{x}_l) \prod_{l' \notin C} (1 - \phi_{l'}(\mathbf{x}_{l'}))\, s_j^\star(\mathbf{x}|C) \quad \text{for } j > 0 \end{aligned} \tag{5.2}$$

where $\mu(\mathbf{x}_0)$ gives the probability of being awake (a function of the attributes of the default good), $\phi_j(\mathbf{x}_j)$ denotes the probability of attending to good j,[48] and $s_j^\star(\mathbf{x}|C)$ denotes the probability of choosing j from choice set C.

There are, therefore, three sets of parameters to identify and estimate: those that index the probability of waking up, $\mu(\cdot)$; those that index the probability of paying attention to good j conditional on waking up, $\phi_j(\cdot)$; and those that index a consumer's utility function, which determine

[47] This assumption could be rationalized by assuming that consumers model plans as being static overtime, that consumers fail to forecast their inertia, or that consumers are myopic in their plan choices (Dalton, Gowrisankaran, and Town (2015) and Abaluck, Gruber, and Swanson (2018) both estimate that Medicare Part D consumers are highly myopic in their drug choices).

[48] $\phi_0(\mathbf{x}_0) = 1$ for all $\mathbf{x}_0 \in \chi$.

latent choice shares $s_j^\star(\cdot)$. We assume logistic functional forms for attention probabilties:

$$\phi_j(\mathbf{x}_j) = \frac{\exp(\mathbf{x}_j\gamma)}{1+\exp(\mathbf{x}_j\gamma)} \tag{5.3}$$

$$\mu(\mathbf{x}_0) = \frac{\exp(\mathbf{x}_0\alpha)}{1+\exp(\mathbf{x}_0\alpha)} \tag{5.4}$$

We assume that the (positive) utility of individual i from choosing plan j at time t is given by:

$$u_{ijt} \quad = \quad \mathbf{x}_{ijt}\beta + \xi \cdot StatusQuo_{ijt} + \epsilon_{ijt} \tag{5.5}$$

where ϵ_{ijt} is distributed i.i.d Type 1 extreme value, $\mathbf{x}_{ijt}\beta$ gives the utility arising from plan characteristics $\mathbf{x}_{ijt}$, and ξ denotes utility-relevant switching costs (acclimation costs) that consumers must pay if they choose any plan other than the status quo from the previous year. Following the earlier literature, we allow consumers to make "errors" by being overly responsive to some plan attributes. Specifically, we allow for separate coefficients on premiums and out of pocket costs (although both are in dollar units),[49] and we allow financial plan characteristics to matter for (positive) utility even conditional on their individualized consequences for consumer costs (a rational consumer should only care about deductibles insofar as they impact the distribution of out of pocket costs).

5.3 Structural Results

Table 4 gives our structural choice model results. We present results for both the hybrid model and also a conditional logit model that does not allow for inattention. We start by discussing the results for the standard conditional logit model (Column 1 of Table 4). The stylized facts from Abaluck and Gruber (2011) and Abaluck and Gruber (2016) are apparent. Consumers weigh premiums more heavily than out of pocket costs, and consumers appear responsive to plan attributes such as deductibles even after we control for the financial impact of those deductibles via out of pocket costs. The fact that consumers are overwhelmingly likely to choose the default plan implies acclimation costs of $1,224 under the conditional logit model.

Next, we estimate the hybrid model on the same sample. After adjusting for limited attention, this model implies acclimation costs of $287; less than a quarter of the size of those estimated assuming full consideration. In the conditional logit model, we estimate that consumers are risk-loving. However, allowing for limited consideration, they appear risk-averse. We also generally find that consumers are more responsive to plan attributes conditional on consideration than is implied

[49] Appendix C of Abaluck and Gruber (2009) shows how this pattern can be derived from a model where some consumers are imperfectly informed about out of pocket costs.

by the conditional logit model.

Our consideration coefficients imply that, on average, consumers consider only 1.97 available plans each year. This is driven both by consumers not actively searching each period and thus choosing their default option with probability one (μ) and by them only considering a subset of available plans if they do search (ϕ). Consider first the estimates of the impact of plan attributes on μ, the probability of being awake. These estimates imply that the majority of consumers do not actively search each period: only 15% of consumers consider a plan other than the default option. The impact of default characteristics on the probability of waking up are intuitive. Consumers are more likely to wake up and consider alternative plans when premiums, out of pocket costs, cost variance, or deductibles increase and when there is a decrease in quality ratings or the number of top 100 drugs are covered. Conditional on waking up, we find that consumers attend to 7.16 plans in their choice set on average. We find that consumers are more likely to attend to options with lower premiums and out of pocket costs, and those with higher quality ratings, more of the top 100 drugs on their formularies, and greater donut hole coverage. The sensitivity of consideration probabilities to donut hole coverage is especially notable – if a rival plan has donut hole coverage, consumers are about ten times more likely to consider it.

Table 4 here

5.4 Welfare Analysis of Smart Defaults

Motivated partly by the considerable inertia in insurance plan choice patterns described above, Handel and Kolstad (2018) propose a 'smart default' policy in which an individual is switched into the lowest cost plan available in each year provided that their monetary gain from such a switch exceeds some threshold.[50] Under this proposal, all enrollees would retain the ability to opt out of their new default and either switch back to their original plan or instead to choose any of the alternative plans available. This policy has not been implemented. Therefore, we must rely on existing variation in the data and structural methods to predict how consumers would respond to the policy and to normatively evaluate the results.

Normative Utility To evaluate the smart default policy, we must take a stand on what is relevant for normative utility. Following Abaluck and Gruber (2011) and Heiss, McFadden, and Winter (2007),[51] we assume that, apart from switching costs (discussed below), normative utility depends

[50] In theory, we could also consider a policy where consumers are only switched if their utility gain exceeds some threshold. Our key results go through under this modification, although the monetary proposal has more practical relevance due to its transparency.

[51] This specification is also supported by our finding of no spurious state dependence in Appendix C.

only on total cost, risk protection and observable quality measures. We denote this utility by v_{ij} (we suppress the subscript t, although plan attributes vary over time as well).[52] In other words, normative utility is given by (the negative) of expected out of costs, plus the dollar-equivalent risk protection and the dollar-equivalent plan quality rating (where, in each case, the dollar-equivalent measures are computed by normalizing by the coefficient on premiums).

In our baseline results, we assume that utility-relevant switching costs are all "acclimation costs". In other words, these are costs that must be paid when a beneficiary enrolls in a plan with which they do not previously have experience.[53] In Appendix C, we consider several other possible drivers of inertia and discuss in more detail how our analysis fits with alternative attempts to decompose inertia into different mechanisms. One alternative story is that inertia may be driven by the "paperwork costs" required to fill-out and send in the forms required to switch plans (Luco 2019). Paperwork costs have different counterfactual implications than acclimation costs: paperwork costs make it costly to return to the original plan after being defaulted away, while acclimation costs make it costly to remain in a new plan even if it otherwise saves money. With positive paperwork costs, a smart default policy can effect choice behavior in full-consideration models as it will now be costly to opt out of the policy. Another possibility is that inertia arises from "spurious state dependence"; i.e. something unobservably good about chosen plans. To separately identify these factors, our analysis in the appendix exploits additional variation from the random assignment of a subset of beneficiaries (low-income subsidy beneficiaries) to alternative plans. This analysis suggests that both paperwork costs and spurious state dependence (at the brand level) play a limited role in explaining observed inertia.[54]

Welfare Change Let consumer i's original plan, or old default, be denoted by the subscript o. The expected welfare change, $\triangle W_i$, can be expressed as:

$$\begin{aligned} \triangle W_i &= W_i^1 - W_i^0 \\ &= \xi \triangle s_{io} + \sum_j \triangle s_{ij} v_{ij} \end{aligned} \tag{5.6}$$

[52] Formally, $v_{ij} = \pi_j + \mu_{ij}^* + \frac{\beta_2}{\beta_0}\sigma_{ij}^2 + \frac{\delta}{\beta_0} q_j$

[53] This is a form of what Heckman (1981) calls "structural state dependence", wherein choices directly impact choice probabilities.

[54] Specifically, when low-income subsidy recipients no longer qualify for low-income subsidies, they must pay paperwork costs but not acclimation costs to return to their original plans. We find that they nonetheless do so at a disproportionate rate, suggesting that paperwork costs are small. Additionally, we find that these beneficiaries are not more likely to return to other plans from the same brand, suggesting that persistent unobserved heterogeneity at the brand level is small as well.

where ξ is the acclimation cost and $\triangle s_{ij} = s^1_{ij} - s^0_{ij}$, where superscripts represent either the current scenario (0) or the counterfactual policy scenario (1).

Defaults change welfare through two channels: first, acclimation costs will be paid by anyone who switches away from the original plan as a result of the new default (we will generally have $\triangle s_{io} < 0$). Second, the policy will change choice probabilities given its aim of inducing people to choose lower cost plans (which might have a higher v_{ij}). Estimating consideration probabilities is required to bring Equation 5.6 to the data for two reasons. First, estimates of ξ will depend on the degree of inattention. Second, to recover Δs_{ij}, we need to simulate the impact of smart defaults on choice probabilities; this in turn will depend on the degree of inattention as well as the structural preference parameters. Inattention will tend to make smart defaults "stickier" in the sense that consumers will not return to their original plans.

The above allows for the smart default policy to change consideration probabilities given that the new default will have different characteristics and, therefore, a different $\mu(\cdot)$. However, it does not allow for the smart default policy to have a direct effect on attention. Defaulting consumers to a different plan might directly wake them up (especially given any outreach campaign that might realistically accompany such a policy). Alternatively, if consumers are rationally inattentive, they may be less likely to pay attention if the smart default is even more suitable (Carroll, Choi, Laibson, Madrian, and Metrick 2009). This matters for positive and normative reasons. Positively, if smart defaults cause people to "wake up", we may see more people revisiting the original plan or making an active choice than we would otherwise predict. Normatively, if smart defaults shift the degree of inattention, we might worry that we are imposing an additional effort cost on some consumers. This additional effort cost is not identified in our model without further assumptions about what drives the decision to pay attention (e.g. it could be identified if we imposed rational inattention). In our results, we thus evaluate smart defaults under a range of assumptions about how attention is directly perturbed, and under a range of values for the effort cost of paying attention.[55]

5.5 Simulation Results

Figure 3 gives the distribution of cost savings achieved under the policy. Note that this ignores the impact of switching costs or other utility-relevant factors. On average, beneficiaries save \$286 from switching to the lowest cost plan in their choice set. However, there are some consumers with substantially greater potential savings; for example, 5% of beneficiaries could save over \$1,000.

[55] This approach is similar to Goldin and Reck (2017), who advocate assuming normative switching costs are a fraction of positive switching costs between 0 and 1. We instead attempt to identify normative switching costs other than effort costs, and consider the robustness of our model to different assumptions about these effort costs.

Figure 3 here

Looking at cost savings alone ignores the impact of acclimation costs associated with switching consumers that lower the utility benefits from moving plans. Further, total cost is not the only relevant characteristic for normative utility (i.e. risk aversion and quality ratings also matter). We simulate welfare under the smart default policy conditional on the structural parameters reported in Table 4. To simulate smart defaults, we switch a consumer to the lowest cost plan available but then allow them to either switch back to their original plan or some other plan. If they choose a plan other than their original plan, they must pay acclimation costs. Consumers might be made worse off by this policy if inattentive consumers are inadvertently driven to pay acclimation costs that outweigh the gains from being enrolled in an otherwise better plan.

Table 5 gives our baseline results. Only 11% of consumers are in the lowest cost plan and thus 89% of beneficiaries are switched by the policy to the lowest cost plan, and among these, 96% stick with the new default. When we use the acclimation costs estimated in the hybrid model, we find this policy has small but positive effects on welfare. This reflects the fact that the utility benefits of the lowest cost plan on average outweigh our estimated switching costs.[56] Note that in the full consideration conditional logit model, this policy would have no effect on choice behavior or welfare. While acclimation costs are large in this model ($1,224), a full-consideration model predicts that all consumers would switch back to their original plan as, absent inattention, defaults have no impact on behavior in our model.

Table 5 here

These results still ignore the possibility that smart defaults might induce people to pay attention, which may itself be costly. While this cost is not identified in our model without additional structure, we consider alternative assumptions about the induced attention probability μ and the cost of paying attention. As the direct effect of the policy on the likelihood of waking up increases, defaulted consumers become more likely to wake up, pay attention costs, and potentially switch to a new plan. While this could in theory lead to a positive effect on welfare, we find that consumers tend to switch to worse plans on average from the perspective of normative utility.

An alternative policy would only reassign those consumers who stand to benefit the most from reassignment. What if we reassign only those beneficiaries for whom the potential cost savings exceed our estimated acclimation costs? The results are shown in Table 6. In this case, the average benefits almost triple and remain positive for any plausible assumptions about attention costs for

[56] The lowest cost plan also often has better non-cost characteristics than a consumer's original plan.

consumers who are switched by the policy. We now reassign far fewer beneficiaries, but the benefits conditional on being reassigned increase substantially to an average of $400-$500 per reassigned beneficiary.

Table 6 here

One important caveat to these results is that we consider only a partial equilibrium analysis: premiums and plan attributes are held constant. In practice, defaulting a large number of beneficiaries into alternative plans would likely cause firms to respond by altering their premiums and coverage characteristics. Decarolis (2015) highlights one way in which such incentives can backfire in a context where the government pays premiums. In the more general context, Ho, Hogan, and Scott Morton (2015) suggests that reducing inertia should enhance competition between plans and lower premiums. Nonetheless, the responsiveness of plan attributes to changes in inertia is not fully understood.

If firms lowered the cost of their plans so that beneficiaries would be defaulted into their product while simultaneously reducing plan desirability on other dimensions such as plan quality, then the welfare benefits of the policy might be diminished. To combat this, one might consider a smart default policy in which beneficiaries are only assigned to plans which also have high plan quality ratings (based on beneficiary feedback). In Appendix Table 11, we consider such an alternative policy. We again only reassign beneficiaries whose potential cost savings exceed acclimation costs, and this time reassigning beneficiaries to the lowest cost plan with a quality rating in the top quartile of available plans. In this case, the welfare benefits are actually slightly smaller than in our baseline simulation (as these new plans are higher cost).

Our analyses reported here do not allow for the possibility that in addition to acclimation costs, consumers face "paperwork costs" that prevent them from switching plans even conditional on paying attention. When paperwork costs are large, even consumers who pay attention may become stuck in unsuitable plans because they do not want to bother to switch. In Appendix C, we consider a generalization of the model here in which consumers face both paperwork and acclimation costs, which we identify using the random assignment of low-income subsidy enrollees to alternative plans (focusing specifically on their choices when they are no longer eligible for such subsidies). The upshot of that analysis is that paperwork costs are negligible, so our conclusions here are unchanged.

6 Conclusion

Discrete choice models with consideration sets relax the strong assumption that consumers consider all of the options available to them before making a choice. In the applied 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 not required for identification. 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 Slustky symmetry, i.e. asymmetries in the matrix of cross-derivatives of choice probabilities with respect to characteristics of rival goods.

To highlight the power of this result, we use our framework to model limited consideration in Medicare Part D, a setting with considerable inertia in choice behavior. Our model allows for two forms of inattention: we allow for consumers to both be "asleep" and simply choose their default option, and also for "awake" consumers to only consider a subset of the available plans. Our results show that, while most inertia is driven by inattention in Part D, there remain non-trivial utility-relevant adjustment costs. We simulate the welfare effect of a "smart default" policy, finding that defaulting consumers into lower cost plans can produce large benefits. This is in contrast to models that assume full consideration, which predict a smart default policy would have no effect on choice behavior and consumer welfare.

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 is 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. Additionally, while we consider consideration at the level of goods, an important question for future work is to characterize the conditions under which choice data suffices to recover inattention at the level of attributes (as in e.g. Bordalo, Gennaioli, and Shleifer (2013, Kőszegi and Szeidl (2012)).[57] 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.

The model we consider provides a general empirical framework for analyzing policies which

[57] Abaluck and Compiani (2020) produces results along these lines.

change defaults across a wide variety of settings such as 401K plans, insurance, school choice, and online consumer purchases among others (Bernheim, Fradkin, and Popov 2015; Carroll, Choi, Laibson, Madrian, and Metrick 2009). Our model implies that defaults change behavior due to inattention. By allowing awake individuals to consider only a subset of options, we avoid overstating utility-relevant switching costs. Rather than imposing specific models of rational inattention, our model estimates consideration probabilities from the data. Without an explicit microfoundation for attention, we cannot predict how attention might shift as we change the underlying context, but we show in our application that our welfare conclusions are robust to alternative assumptions about such shifts.

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? 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 increase the likelihood that consumers are attentive? How much better off might consumers be if we switched them to alternative options? We hope that future work will explore these questions in more detail in other policy-relevant scenarios.

References

Abaluck, J. and G. Compiani (2020). A method to estimate discrete choice models that is robust to consumer search. Technical report, National Bureau of Economic Research.

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.

Abaluck, J., J. Gruber, and A. Swanson (2018). Prescription drug use under medicare part d: A linear model of nonlinear budget sets. *Journal of public economics 164*, 106–138.

Abaluck, J. T. and J. Gruber (2009, February). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. Working Paper 14759, National Bureau of Economic Research.

Aguiar, V. H., M. Boccardi, N. Kashaev, and J. Kim (2018). Does random consideration explain behavior when choice is hard? evidence from a large-scale experiment.

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.

Anderson, S. P., A. De Palma, and J.-F. Thisse (1992). *Discrete choice theory of product differentiation*. MIT press.

Barseghyan, L., M. Coughlin, 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.

Bernheim, B. D., A. Fradkin, and I. Popov (2015). The welfare economics of default options in 401 (k) plans. *American Economic Review 105*(9), 2798–2837.

Berry, S. and P. Haile (2016). Identification in differentiated products markets. *Annual review of Economics 8*, 27–52.

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.

Bhargava, S., G. Loewenstein, and J. Sydnor (2015). Do individuals make sensible health insurance decisions? evidence from a menu with dominated options. Technical report, National Bureau of Economic Research.

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 (2018). Rational inattention, optimal consideration sets and stochastic choice. Technical report.

Carroll, G. D., J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick (2009). Optimal defaults and active decisions. *The quarterly journal of economics 124*(4), 1639–1674.

Cattaneo, M., X. Ma, Y. Masatlioglu, and E. Suleymanov (2018). A random attention model.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020). A random attention model. *Journal of Political Economy 128*(7), 000–000.

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.

Dalton, C. M., G. Gowrisankaran, and R. Town (2015). Myopia and complex dynamic incentives: Evidence from medicare part d. *NBER Working Paper 21104*.

Daly, A. and S. Zachary (1978). Improved multiple choice models. *Determinants of travel choice 335*, 357.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020). Inferring cognitive heterogeneity from aggregate choices. *Econometrica 88*(3), 1269–1296.

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.

Decarolis, F. (2015). Medicare part d: Are insurers gaming the low income subsidy design? *American Economic Review 105*(4), 1547–80.

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.

Duffie, D. and K. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*(4), 929–52.

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

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies 72*(2), 395–427.

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

Gabaix, X. (2017). Behavioral inattention. Technical report, National Bureau of Economic Research.

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

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

Goldin, J. and D. Reck (2017). Optimal defaults with normative ambiguity.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of applied econometrics 8*(S1).

Handel, B. and J. Kolstad (2018). *Getting the Most from Marketplaces: Smart Policies on Health Insurance Choices*.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review 103*(7), 2643–82.

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

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

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, pp. 91–140. University of Chicago Press.

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, and J. Winter (2007, November). Mind the gap! consumer perceptions and choices of medicare part d prescription drug plans. Working Paper 13627, National Bureau of Economic Research.

Heiss, F., D. McFadden, and J. Winter (2010). Mind the gap! consumer perceptions and choices of medicare part d prescription drug plans. In *Research findings in the economics of aging*, pp. 413–481. University of Chicago Press.

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.

Hoadley, J., E. Hargrave, L. Summer, J. Cubanski, and T. Neuman (2013). To switch or not to switch: Are medicare beneficiaries switching drug plans to save money. *Kaiser Family Foundation*.

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*.

Ketcham, J. D., N. V. Kuminoff, and C. A. Powers (2019). Estimating the heterogeneous welfare effects of choice architecture: An application to the medicare prescription drug insurance market. Technical report, National Bureau of Economic Research.

Koning, R. H. and G. Ridder (2003). Discrete choice and stochastic utility maximization. *The Econometrics Journal 6*(1), 1–27.

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.

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

Lu, Z. (2016). Estimating multinomial choice models with unobserved choice sets.

Luco, F. (2019). Switching costs and competition in retirement investment. *American Economic Journal: Microeconomics 11*(2), 26–54.

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

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.

Matzkin, R. L. (2007). Nonparametric identification. *Handbook of econometrics 6*, 5307–5368.

Moshkin, N. V. and R. Shachar (2002). The asymmetric information model of state dependence. *Marketing Science 21*(4), 435–454.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, 1027–1057.

Raval, D. and T. Rosenbaum (2018). Why do previous choices matter for hospital demand? decomposing switching costs from unobserved preferences. *Review of Economics and Statistics 100*(5), 906–915.

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

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.

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.

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.

7 Tables in Main Text

Table 1: Experimental Data Estimation Results

	Conditional Logit	ASC Model		Conditional on Consideration
		MLE	Indirect Inf.	
Utility:				
Price (dollars)	-0.054***	-0.196***	-0.157**	-0.173***
	(0.003)	(0.028)	(0.029)	(0.004)
Product 1	-1.411***	1.465***	0.613	0.368***
	(0.054)	(0.539)	(0.406)	(0.069)
Product 2	-1.955***	-0.065	0.133	-0.497***
	(0.069)	(0.478)	(0.573)	(0.080)
Product 3	-1.627***	0.625	0.344	0.093
	(0.059)	(0.476)	(0.403)	(0.073)
Product 4	-1.640***	0.629	0.443	0.088
	(0.060)	(0.466)	(0.386)	(0.073)
Product 5	-1.447***	0.707	0.934***	0.306***
	(0.056)	(0.478)	(0.354)	(0.070)
Product 6	-0.435***	-0.737***	-0.422	-0.581***
	(0.039)	(0.121)	(0.252)	(0.045)
Product 7	-0.855***	-1.280***	-0.474	-1.075***
	(0.045)	(0.141)	(0.234)	(0.051)
Product 8	-0.662***	-1.185***	-0.711***	-0.909***
	(0.041)	(0.137)	(0.271)	(0.048)
Product 9	-0.316***	-0.561***	-0.706***	-0.405***
	(0.038)	(0.118)	(0.184)	(0.044)
Attention:				
Price (dollars)		0.137***	0.133***	0.15
		(0.017)	(0.027)	
Product 1		-2.872***	-2.636***	-2.5
		(0.177)	(0.359)	
Product 2		-2.674***	-2.914***	-2.5
		(0.288)	(0.544)	
Product 3		-2.695***	-2.599***	-2.5
		(0.209)	(0.360)	
Product 4		-2.704***	-2.703***	-2.5
		(0.205)	(0.389)	
Product 5		-2.592***	-3.042***	-2.5
		(0.204)	(0.389)	
Product 6		0.152	-0.464	0
		(0.192)	(0.367)	
Product 7		0.123	-0.925***	0
		(0.292)	(0.345)	
Product 8		0.258	0.476	0
		(0.230)	(0.425)	
Product 9		0.103	0.779	0
		(0.176)	(0.532)	

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 that conditions on the actual choice set consumers faced. The true attention parameters are set by us in advance. The ASC model also includes a constant in consideration probabilities. *** Denotes significance at the 1% level, ** significance at the 5% level and * significance at the 10% level.

Table 2: Sample Demographics and Plan Characteristics

Beneficiary Characteristics:		
	Mean	Std Dev
Age	75.5	8.6
Female	0.643	–
White	0.944	–
Plan Characteristics of Chosen Plans:		
	Mean	Std Dev
Premiums	$415	$191
Out of Pocket Costs	$839	$713
Deductible	$59	$113
Donut Hole Coverage	0.121	–
Choice Characteristics:		
	Mean	Std Dev
Options in Choice Set	11.9	2.43
Inertial	93.7	–
No. Beneficiary-Plan-Year	2,261,878	
No. Beneficiary Observations	79,286	

Notes: Table shows summary statistics for the demographic characteristics and available insurance plans for our sample of Medicare Part D PDP beneficiaries in 2008 and 2009. "Donut Hole Coverage" refers to either generic or full donut hole coverage.

Table 3: Excess Sensitivity to Default Attributes in Switching Model

	Share Weighted	Lowest 3 Plans	Lowest Cost Plan
Annual Premium (hundreds)	0.0801***	0.0931***	0.0914***
	(0.0305)	(0.0348)	(0.0292)
Annual Out of Pocket Costs (hundreds)	0.0057**	-0.0051	-0.0096**
	(0.0028)	(0.0048)	(0.0046)
Variance of Costs (millions)	-0.0075**	0.0040	-0.0006
	(0.0037)	(0.0058)	(0.0003)
Deductible (hundreds)	0.1203**	0.1242**	0.1219**
	(0.0545)	(0.0501)	(.0556)
Donut Hole Coverage	0.0827**	0.0926**	0.0909**
	(0.0410)	(0.0376)	(0.0396)
Average Consumer Cost Sharing %	0.0186	0.0234	0.0210
	(0.0141)	(0.0161)	(0.0143)
Normalized Quality Rating	0.0024	0.0479	0.0103
	(0.0156)	(0.1663)	(0.0136)
# of Top 100 Drugs in Formulary	0.1669	0.0183	0.1347
	(0.0985)	(0.0157)	(0.1158)

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, as well as the difference between default plan attributes and rival plan attributes for three different models of rival plan attributes. All models include individual and time fixed effects. In all cases, the reported coefficient is the coefficient on the default attribute conditioning on the difference between the default and rival plan attribute (α_d in Equation 5.1). The first column computes rival plan attributes as a share-weighted average of non-default plans where the shares are computed for each (year, state, plan) using the choices of new beneficiaries (who are not included in this regression). The second column uses (unweighted) average attributes among the three lowest cost plans. The third column uses attributes of the lowest cost plan for rival plan attributes. The regression also includes an indicator for plans that are missing the # of top 100 drugs in formulary variable as well as an interaction of variance of costs and an indicator for individuals with no claims. Standard errors in parentheses. *** denotes significance at the 1% level, ** at 5% level, and * at 10%.

Table 4: Conditional Logit & Hybrid Model

		Hybrid Model		
	Conditional Logit	Utility	$\phi(\cdot)$	$\mu(\cdot)$
Annual Premium (hundreds)	-0.4298***	-0.6293***	-1.118***	0.0281**
	(0.0103)	(0.0331)	(0.0731)	(0.0116)
Annual Out of Pocket Costs (hundreds)	-0.1420***	-0.0054	-0.3797***	0.0657***
	(0.0131)	(0.0321)	(0.0697)	(0.0206)
Variance of Costs (millions)	0.1626***	-0.1687***	0.5501***	1.5589***
	(0.0183)	(0.0492)	(0.0981)	(0.0301)
Deductible (hundreds)	-0.4108***	-0.6642***	-0.0531	0.2008***
	(0.021)	(0.0552)	(0.1017)	(0.0483)
Donut Hole Coverage	0.5656***	1.3579***	10.1685***	-0.0889
	(0.0609)	(0.2071)	(0.7432)	(0.1127)
Average Consumer Cost Sharing %	-0.2209***	-1.2484***	1.3849***	-0.3911***
	(0.0242)	(0.0678)	(0.1474)	(0.0663)
# of Top 100 Drugs in Formulary	0.1243***	0.0010	0.4557***	-0.1157***
	(0.0122)	(0.0362)	(0.0483)	(0.0371)
Normalized Quality Rating	0.0399***	-0.0066	0.1976***	-0.0843***
	(0.0021)	(0.0062)	(0.0181)	(0.0043)
Default	5.2599***	1.8052***		
	(0.0224)	(0.1473)		
Constant			1.626***	-1.853***
			(0.2126)	(0.0764)
Switching Cost	$1224	$287		
Mean Attention Probability			67.46%	15.09%

Notes: This table gives estimates from the conditional logit and "hybrid" version of the consideration set framework. Estimates in all models are the coefficients in the utility and attention equations (not marginal effects). The coefficients in the DSC component of the model, $\mu(\cdot)$, are the coefficients on the listed characteristics of the default good. The coefficients in the ASC component of the model, $\phi(\cdot)$, are the coefficients on the listed characteristics of good j on the likelihood of j being considered. The model also includes an indicator for plans that are missing the # of top 100 drugs in formulary variable as well as an interaction of variance of costs and an indicator for individuals with no claims. Standard errors in parentheses. *** denotes significance at the 1% level, ** at 5% level, and * at 10%.

Table 5: Welfare Impact of Smart Default Policy

	Market Shares Conditional on Being Switched			Attention Cost			
	Smart Default	Previous Plan	Other Plan	\$0	\$50	\$100	\$200
Conditional Logit	1.000	0.000	0.000	\$0	\$0	\$0	\$0
Hybrid Model	0.961	0.001	0.048	\$46	\$42	\$37	\$29
Direct Effect on μ							
25%	0.946	0.002	0.053	\$39	\$32	\$25	\$12
50%	0.902	0.003	0.095	\$18	\$2	-\$14	-\$46
75%	0.853	0.005	0.142	-\$3	-\$30	-\$57	-\$111
100%	0.804	0.006	0.189	-\$25	-\$63	-\$101	-\$177
Proportion Switched by Policy: 89%							

Notes: The table shows overall welfare impacts of a smart default policy where consumers are switched to the lowest cost plan available in their market. Each row shows results with different assumptions about the direct effect of the smart default policy on the probability of paying attention, while each column shows alternative assumptions about the cost of paying attention.

Table 6: Welfare Impact of Smart Default Policy – Restricted Reassignment to those with Savings Greater than Switching Costs

	Market Shares Conditional on Being Switched			Attention Cost			
	Smart Default	Previous Plan	Other Plan	$0	$50	$100	$200
Mean Welfare Change: Full Sample							
Hybrid Parameters				$114	$113	$112	$111
Direct Effect on Attention Probability							
25%				$109	$107	$106	$102
50%				$94	$90	$85	$76
75%				$79	$71	$64	$49
100%				$63	$52	$42	$21
Proportion Consumers Switched by Policy: 24%							
Mean Welfare Change: Conditional on Being Switched							
Hybrid Parameters	0.947	0.001	0.052	$479	$476	$472	$465
Direct Effect on Attention Probability							
25%	0.923	0.002	0.076	$458	$451	$444	$429
50%	0.852	0.003	0.145	$395	$377	$358	$320
75%	0.778	0.005	0.217	$330	$299	$267	$205
100%	0.705	0.006	0.289	$264	$220	$177	$89

Notes: The table shows overall welfare impacts of a smart default policy where consumers who will experience a monetary saving of at least ξ are switched to the lowest cost plan available in their market. Each row shows results with different assumptions about the direct effect of the smart default policy on the probability of paying attention, while each column shows alternative assumptions about the cost of paying attention.

8 Figures in Main Text

Figure 1: Lab Experiment: Sample Product Selection Screen

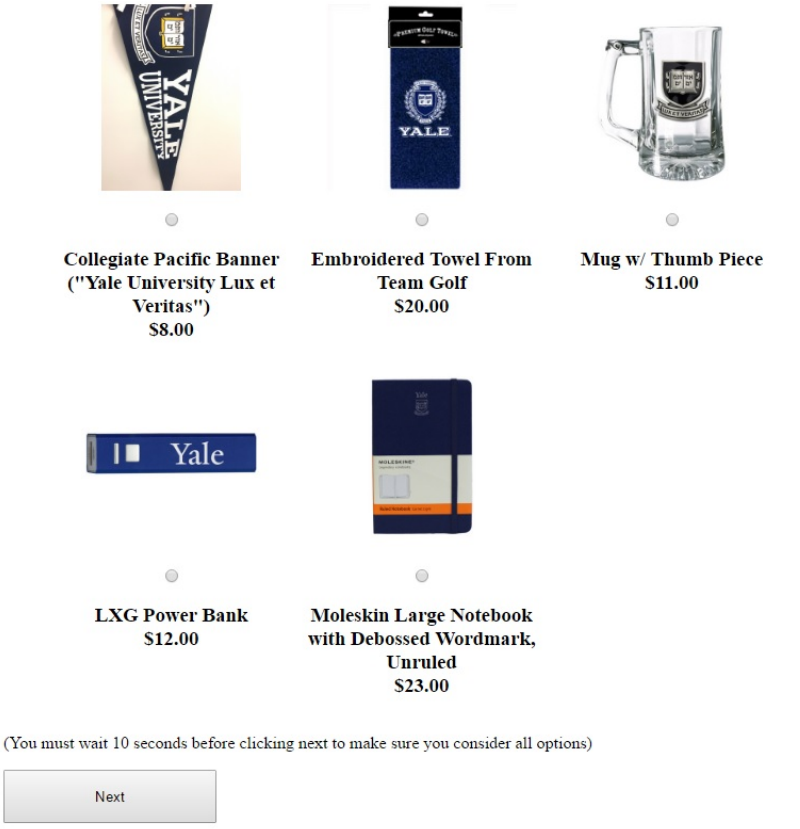


Figure 2: Product Fixed Effects in Attention: Truth vs. ASC Model

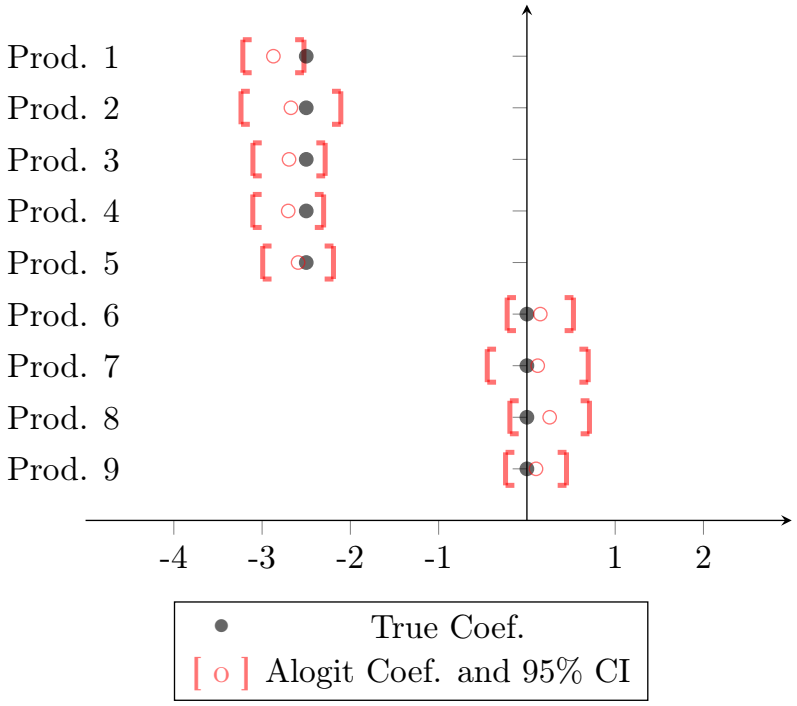
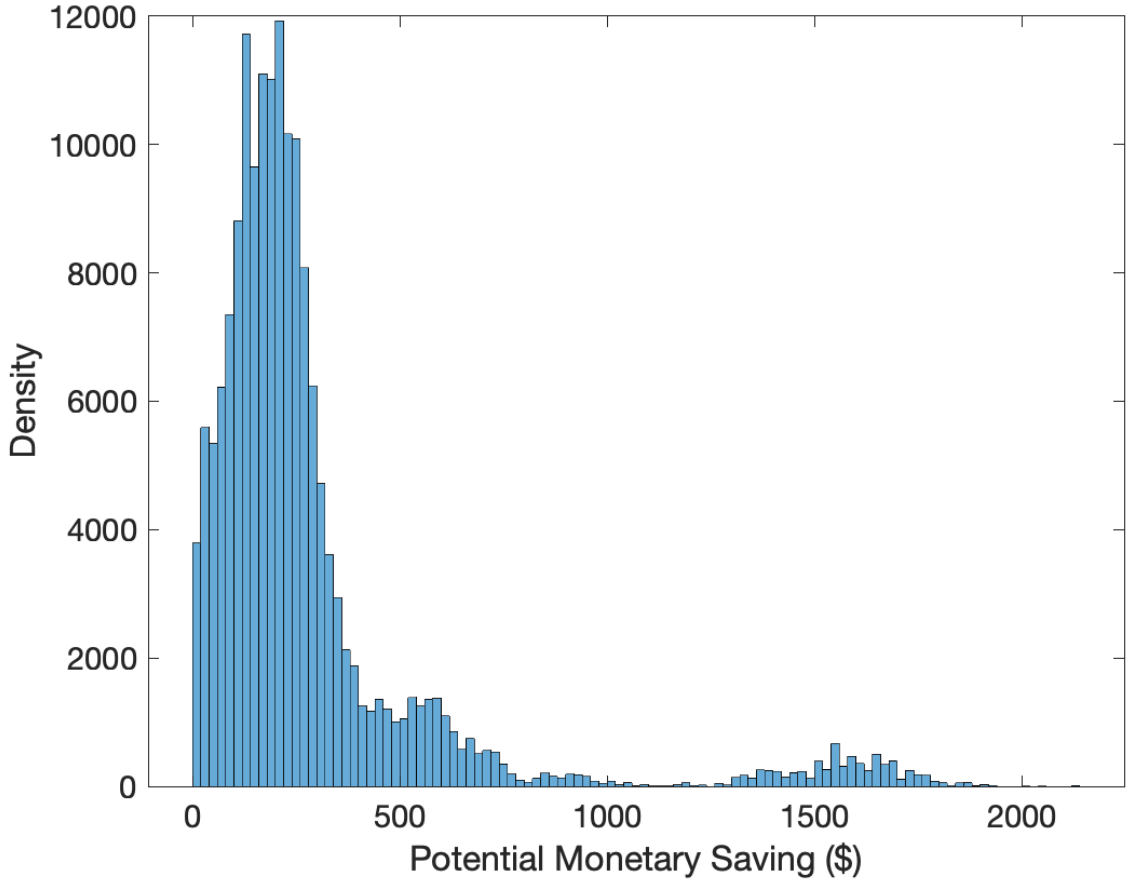


Figure 3: Distribution of Savings



Notes: Figure shows the distribution of potential monetary savings (annual premiums plus estimated out-of-pocket costs) from being switched from a recipient's current plan to the lowest cost plan.

A Model & Identification Proof

A.1 Results to Complement Section 3

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

$$\mathbb{P}(j, j') = \{c : c \in \mathbb{P}(\mathcal{J}) \quad \& \quad j \in c \quad \& \quad j' \in c \quad \& \quad 0 \in c\}, \tag{A.1}$$

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. We here suppress dependence of market shares and consideration probabilities on $\mathbf{p}$ to prevent the working becoming unnecessarily burdensome.

$$\frac{\partial s_j}{\partial p_{j'}} - \frac{\partial s_{j'}}{\partial p_j} = \sum_{C \in \mathbb{P}(j)} \frac{\partial \pi_C}{\partial p_{j'}} s_j^\star(C) - \sum_{C' \in \mathbb{P}(j')} \frac{\partial \pi_{C'}}{\partial p_j} s_{j'}^\star(C') + \sum_{C'' \in \mathbb{P}(j,j')} \pi_{C''} \left(\frac{\partial s_j^\star(C'')}{\partial p_{j'}} - \frac{\partial s_{j'}^\star(C'')}{\partial p_j} \right) \tag{A.2}$$

$$= \sum_{C \in \mathbb{P}(j)} \frac{\partial \pi_C}{\partial p_{j'}} s_j^\star(C) - \sum_{C' \in \mathbb{P}(j')} \frac{\partial \pi_{C'}}{\partial p_j} s_{j'}^\star(C') \tag{A.3}$$

Thus, non-zero cross-derivative asymmetries imply:

$$\sum_{C \in \mathbb{P}(j)} \frac{\partial \pi_C}{\partial p_{j'}} s_j^\star(C) \neq \sum_{C' \in \mathbb{P}(j')} \frac{\partial \pi_{C'}}{\partial p_j} s_{j'}^\star(C') \tag{A.4}$$

$$\text{either} \sum_{C \in \mathbb{P}(j)} \frac{\partial \pi_C}{\partial p_{j'}} s_j^\star(C) \neq 0 \quad \text{and/or} \sum_{C' \in \mathbb{P}(j')} \frac{\partial \pi_{C'}}{\partial p_j} s_{j'}^\star(C') \neq 0 \tag{A.5}$$

Given π_C represent proper probabilities, this is only possible when $\pi_{\mathcal{J}} < 1$.

Similarly, while level shifts in prices 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 $\delta \neq 0$,

$$s_j(\mathbf{p} + \delta) = \sum_{c \in \mathbb{P}(j)} \pi_C(\mathbf{p} + \delta) Pr \left(v_i(p_j) + \epsilon_{ij} = \max_{j' \in C} v_i(p_{j'}) + \epsilon_{ij'} \right) \tag{A.6}$$

If $s_j(\mathbf{p}) \neq s_j(\mathbf{p}+\delta)$, this implies that for at least one consideration set C

$$\pi_C(\mathbf{p}) \neq \pi_C(\mathbf{p}+\delta), \tag{A.7}$$

Given π_C represent proper probabilities, this is only possible when $\pi_{\mathcal{J}} < 1$.

DSC Model

Proof of Theorem 4. From Theorem 2, we have:

$$\frac{\partial \log(\mu_0)}{\partial p_0} = \frac{1}{s_j(\mathbf{p})}\left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j}\right] \tag{A.8}$$

When μ_0 takes the logit form we have:

$$\mu(p_0) = \frac{\exp(\delta+\gamma p_0)}{1+\exp(\delta+\gamma p_0)} \tag{A.9}$$

$$\frac{\partial \log(\mu_0)}{\partial p_0} = \frac{\gamma}{1+\exp(\delta+\gamma p_0)} \tag{A.10}$$

Let us observe demand at two prices of the default good, p_0^a and p_0^b, with $p_0^a < p_0^b$. Without loss of generality, let $\gamma > 0$. Let

$$\frac{1}{s_j(\mathbf{p}^a)}\left[\frac{\partial s_j(\mathbf{p}^a)}{\partial p_0} - \frac{\partial s_0(\mathbf{p}^a)}{\partial p_j}\right] = d^a \tag{A.11}$$

$$\frac{1}{s_j(\mathbf{p}^b)}\left[\frac{\partial s_j(\mathbf{p}^b)}{\partial p_0} - \frac{\partial s_0(\mathbf{p}^b)}{\partial p_j}\right] = d^b \tag{A.12}$$

As $p_0^a < p_0^b$, we have $d^a > d^b$.

We have two expressions, with two unknowns:

$$\frac{\gamma}{1+\exp(\delta+\gamma p_0^a)} = d^a \tag{A.13}$$

$$\frac{\gamma}{1+\exp(\delta+\gamma p_0^b)} = d^b \tag{A.14}$$

Solving Equation A.13 for δ we have:

$$\delta = \log\left(\frac{\gamma - d^a}{d^a}\right) - \gamma p_0^a \tag{A.15}$$

Substituting into Equation A.14 gives

$$\log\left(\frac{\gamma - d^a}{d^a}\right) - \log\left(\frac{\gamma - d^b}{d^b}\right) + \gamma(p_0^b - p_0^a) = 0 \tag{A.16}$$

Let $f(\gamma) = \log\left(\frac{\gamma - d^a}{d^a}\right) - \log\left(\frac{\gamma - d^b}{d^b}\right) + \gamma(p_0^b - p_0^a)$. As $d^a > d^b$, we have that $0 < \gamma - d^a < \gamma - d^b$. Thus,

$$\frac{\partial f(\gamma)}{\partial \gamma} > 0 \tag{A.17}$$

And thus there is a unique value of γ at which $f(\gamma) = 0$.

ASC Model

Derivation of ASC Cross-Derivative Differences: Observed market shares take the form:

$$s_j(\mathbf{p}) = \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C) \tag{A.18}$$

$$= \phi_{j'}(p_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C) + (1 - \phi_{j'}(p_{j'})) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C/j') \tag{A.19}$$

$$= \phi_{j'}(p_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_j^\star(\mathbf{p}|C) - s_j^\star(\mathbf{p}|C/j')) + \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C/j') \tag{A.20}$$

where $\widetilde{\pi}(C) = \prod_{l \in C/j'} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))$.

In markets where j' is not available, market shares can be expressed as:

$$s_j(\mathbf{p}|\mathcal{J}/j') = \sum_{C \in \{\mathbb{P}(j)/\mathbb{P}(j')\}} \prod_{l \in C} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C) \tag{A.21}$$

$$= \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \prod_{l \in C/j'} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C/j') \tag{A.22}$$

$$= \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C/j') \tag{A.23}$$

Given that changes in latent market shares cancel out within a consideration set, cross derivative

differences take the form:

$$\frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} - \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} = \frac{\partial \phi_{j'}}{\partial p_{j'}} \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_j^\star(\mathbf{p}|C) - s_j^\star(\mathbf{p}|C/j')) - \frac{\partial \phi_j}{\partial p_{j'}} \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_{j'}^\star(\mathbf{p}|C) - s_{j'}^\star(\mathbf{p}|C/j)) \tag{A.24}$$

$$= \frac{\partial \phi_{j'}}{\partial p_{j'}} \frac{1}{\phi_{j'}}(s_j(\mathbf{p}) - s_j(\mathbf{p}|\mathcal{J}/j')) - \frac{\partial \phi_j}{\partial p_j} \frac{1}{\phi_j}(s_{j'}(\mathbf{p}) - s_{j'}(\mathbf{p}|\mathcal{J}/j)) \tag{A.25}$$

$$= \frac{\partial \log \phi_{j'}}{\partial p_{j'}}(s_j(\mathbf{p}) - s_j(\mathbf{p}|\mathcal{J}/j')) - \frac{\partial \log \phi_j}{\partial p_j}(s_{j'}(\mathbf{p}) - s_{j'}(\mathbf{p}|\mathcal{J}/j)) \tag{A.26}$$

In scenarios where leave-one-out variation is not observed, we can use the large support assumption on prices to derive an expression for cross derivative differences as a linear function of changes in consideration probabilities. Let $\mathbf{p}'_{j'}$ be the price vector $\mathbf{p}$ with only $p'_{j'} > p_{j'}$. The difference in the market share of good j is:

$$s_j(\mathbf{p}) - s_j(\mathbf{p}'_{j'}) = \phi_{j'}(p_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_j^\star(\mathbf{p}|C) - s_j^\star(\mathbf{p}|C/j')) - \phi_{j'}(p'_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_j^\star(\mathbf{p}'_{j'}|C) - s_j^\star(\mathbf{p}'_{j'}|C/j')) \tag{A.27}$$

As $p'_{j'} \to \infty$, we require $s_j^\star(\mathbf{p}'_{j'}|C) \to s_j^\star(\mathbf{p}'_{j'}|C/j')$ (a property satisfied by all parametric ARUMs). Thus,

$$s_j(\mathbf{p}) - s_j(\bar{\mathbf{p}}_{j'}) = \phi_{j'}(p_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_j^\star(\mathbf{p}|C) - s_j^\star(\mathbf{p}|C/j')) \tag{A.28}$$

$$= s_j(\mathbf{p}) - s_j(\mathbf{p}|\mathcal{J}/j') \tag{A.29}$$

with $\bar{\mathbf{p}}_{j'}$ the price vector $\mathbf{p}$ with only $p_{j'} \approx \infty$.

Hybrid Model

For the rank condition in Theorem 6 (Assumption Hybrid) to hold (or in the ASC model, if one would like to rely on cross-derivatives beyond those involving the inside default), we must have that the number of independent cross-derivative differences is at least as large as the number of derivatives of the log of consideration probabilities:

$$\frac{1}{2}J(J+1) \geq J+1 \tag{A.30}$$

$$J \geq 2 \tag{A.31}$$

Further, all columns of $D(\mathbf{p})$ must be linearly independent. Sufficient conditions for this are:

$$s_j(\mathbf{p}|\mathcal{J}) \neq s_{j'}(\mathbf{p}|\mathcal{J}) \tag{A.32}$$

$$s_j(\mathbf{p}|\mathcal{J}) \neq s_j(\mathbf{p}|\mathcal{J}/j') \tag{A.33}$$

$$\frac{s_l(\mathbf{p}|\mathcal{J}) - s_l(\mathbf{p}|\mathcal{J}/j)}{s_{j'}(\mathbf{p}|\mathcal{J}) - s_{j'}(\mathbf{p}|\mathcal{J}/j)} \neq \frac{s_l(\mathbf{p}|\mathcal{J}) - s_l(\mathbf{p}|\mathcal{J}/j')}{s_j(\mathbf{p}|\mathcal{J}) - s_j(\mathbf{p}|\mathcal{J}/j')} \tag{A.34}$$

for all $j, j', l \in \mathcal{J}$ with $j, j' > 0$. Equation A.34 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:

$$\begin{bmatrix} -(s_0(\mathcal{J}) - s_0(\mathcal{J}/1)) & 0 & s_1(\mathcal{J}) \\ 0 & -(s_0(\mathcal{J}) - s_0(\mathcal{J}/2)) & s_2(\mathcal{J}) \\ -(s_2(\mathcal{J}) - s_2(\mathcal{J}/1)) & (s_1(\mathcal{J}) - s_1(\mathcal{J}/2)) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \log(\phi_1)}{\partial p_1} \\ \frac{\partial \log(\phi_{2m})}{\partial p_2} \\ \frac{\partial \log(\mu_m)}{\partial p_0} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_1}{\partial p_1} - \frac{\partial s_0}{\partial 0_1} \\ \frac{\partial s_2}{\partial p_0} - \frac{\partial s_0}{\partial 0_2} \\ \frac{\partial s_1}{\partial p_2} - \frac{\partial s_2}{\partial 0_1} \end{bmatrix} \tag{A.35}$$

$$D(\mathbf{p})A(\mathbf{p}) = c(\mathbf{p}) \tag{A.36}$$

The determinant of D is:

$$\begin{aligned} \det(D) = {} & s_2(\mathcal{J})\,(s_0(\mathcal{J}) - s_0(\mathcal{J}/1))\,(s_1(\mathcal{J}) - s_1(\mathcal{J}/2)) \\ & - s_{1m}(\mathcal{J})\,(s_0(\mathcal{J}) - s_0(\mathcal{J}/2))\,(s_2(\mathcal{J}) - s_2(\mathcal{J}/1)) \end{aligned} \tag{A.37}$$

When D is singular:

$$\frac{1}{s_1(\mathcal{J})}\left(\frac{s_0(\mathcal{J}) - s_0(\mathcal{J}/1)}{s_2(\mathcal{J}) - s_2(\mathcal{J}/1)}\right) = \frac{1}{s_2(\mathcal{J})}\left(\frac{s_0(\mathcal{J}) - s_0(\mathcal{J}/2)}{s_1(\mathcal{J}) - s_1(\mathcal{J}/2)}\right) \tag{A.38}$$

A.2 Identification of Latent Market Shares

"Nominal illusion" facilitates the identification of the 2^J independent latent choice probabilities in the ASC and Hybrid models, $s_j^\star(\mathbf{p}|C)$. We treat $\phi_j(\cdot)$ as known given the arguments above. Imagine that $N = 2^J$ price shifts are observed (meaning that prices for all goods are perturbed by the same constant). Given the Daly-Zachary conditions, these price shifts alter consideration probabilities

but do *not* alter latent choice probabilities conditional on consideration.[58] Let $k = 1, ..., \kappa$ index the consideration sets of which j is a member. The probabilities of these consideration sets containing j are given as $\pi_{j1}, ..., \pi_{j\kappa}$. For each good $j > 0$,[59] define the matrices:

$$\Pi_j(\mathbf{p}, \delta) = \begin{bmatrix} \pi_{j1}(\delta_1) & \cdots & \pi_{j\kappa}(\delta_1) \\ \vdots & \ddots & \vdots \\ \pi_{j1}(\delta_N) & \cdots & \pi_{j\kappa}(\delta_N) \end{bmatrix} \tag{A.39}$$

$$s_j^\star(\mathbf{p}) = \left[s_j^\star(\mathbf{p}|C_{j1}), ..., s_j^\star(\mathbf{p}|C_{j\kappa})\right] \tag{A.40}$$

$$s_j^\delta(\mathbf{p}) = [s_j(\mathbf{p} + \delta_1), ..., s_j(\mathbf{p} + \delta_N)] \tag{A.41}$$

where

$$\pi_{jC}(\delta) = \mu(p_0) \prod_{l \in C} \phi_l(p_l + \delta) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'} + \delta)) \tag{A.42}$$

with $\phi_0(p_0) = 1$ for all p_0. Unobserved latent choice probabilities are defined as the solution to the following linear system:

$$\Pi_j(\mathbf{p}, \delta) s_j^\star(\mathbf{p}) = s_j^\delta(\mathbf{p}) \tag{A.43}$$

$$s_j^\star(\mathbf{p}) = \Pi_j^{-1}(\mathbf{p}, \delta) s_j^\delta(\mathbf{p}) \tag{A.44}$$

There is a unique solution to this system, and thus all $s_j^\star(\mathbf{p})$ are identified, when $\Pi_j(\mathbf{p}, \delta)$ is known and has full rank for $j = 1, ..., J$.

Assumption Share (Rank Condition) $\Pi_j(\mathbf{p}, \delta)$ is full rank for $j = 1, ..., J$.

To avoid too much repetition, we here examine the conditions for this assumption to hold for the Hybrid case only. Consider the just identified case where $J = 2$ and $N = 2$. Suppressing dependence

[58] We can relax quasi-linearity and allow for income effects if the parametric form of these income effects can be estimated.

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

on product characteristics, the coefficient matrix takes the form:

$$\Pi_1(\delta) = \begin{bmatrix} \mu(\delta_1)\phi_1(\delta_1)(1-\phi_2(\delta_1)) & \mu(\delta_1)\phi_1(\delta_1)\phi_2(\delta_1) \\ \mu(\delta_2)\phi_1(\delta_2)(1-\phi_2(\delta_2)) & \mu(\delta_2)\phi_1(\delta_2)\phi_2(\delta_2) \end{bmatrix} \tag{A.45}$$

$$\Pi_2(\delta) = \begin{bmatrix} \mu(\delta_1)\phi_2(\delta_1)(1-\phi_1(\delta_1)) & \mu(\delta_1)\phi_1(\delta_1)\phi_2(\delta_1) \\ \mu(\delta_2)\phi_2(\delta_2)(1-\phi_1(\delta_2)) & \mu(\delta_2)\phi_1(\delta_2)\phi_2(\delta_2) \end{bmatrix} \tag{A.46}$$

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

$$\frac{1-\phi_2(\delta_1)}{\phi_2(\delta_1)} = \frac{1-\phi_2(\delta_2)}{\phi_2(\delta_2)} \tag{A.47}$$

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

$$\frac{1-\phi_1(\delta_1)}{\phi_1(\delta_1)} = \frac{1-\phi_1(\delta_2)}{\phi_1(\delta_2)} \tag{A.48}$$

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

A.3 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:

$$s_j = \sum_{c\in\mathbb{P}(j)} \prod_{l\in C} \phi_l(p_0, p_l) \prod_{l'\notin C} (1-\phi_{l'}(p_0, p_{l'})) \tag{A.49}$$

with $\phi_0 = 1$ and $\mathbb{P}(j) = \{C : C \in \mathcal{P}(\mathcal{J}) \quad \& \quad j \in C \quad \& \quad 0 \in C\}$.

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

$$\frac{\partial s_j}{\partial p_0} - \frac{\partial s_0}{\partial p_j} = \frac{\partial \log(\phi_j)}{\partial p_0} s_j(\mathbf{p}) + \sum_{j'\neq\{j,0\}} \frac{\partial \log(\phi_{j'})}{\partial p_0} \left(s_j(\mathbf{p}) - s_j(\bar{\mathbf{p}}_{j'}))\right) - \frac{\partial \log(\phi_j)}{\partial p_j} \left(s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j))\right) \tag{A.50}$$

Thus there are now $2J$ derivatives of log consideration probabilities to identify: $\partial \log(\phi_j)/\partial p_j$ and

$\partial \log(\phi_j)/\partial p_0$ 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 Hybrid):

$$\frac{1}{2}J(J+1) \geq 2J \tag{A.51}$$

$$J \geq 3 \tag{A.52}$$

In this model, we cannot allow $\phi_0(p_0) \leq 1$ 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.

A.4 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 that 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 Hybrid). We now require:

$$\frac{1}{2}J(J-1) \geq J \tag{A.53}$$

$$J \geq 3 \tag{A.54}$$

All cross derivative differences take the form given by Equation 3.25 and identification proceeds analogously to Theorem 6.

A.5 Identification of Model 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. To illustrate, let $g_j = x_j\gamma$, where $x_j \in \mathbb{R}^K$ and assume that the impact of characteristics on consideration probabilities comes via the indices g_j. The general expression for cross-derivative differences (with respect to attribute k) in consideration

set models then takes the form:

$$\frac{\partial s_j(\mathbf{x})}{\partial x_{j'}^k} - \frac{\partial s_{j'}(\mathbf{x})}{\partial x_j^k} = \sum_{C \in \mathbb{P}(j)} \frac{\partial \pi_C(g_0, ..., g_J)}{\partial x_{j'}^k} s_j^\star(\mathbf{x}|C) - \sum_{C' \in \mathbb{P}(j')} \frac{\partial \pi_{C'}(g_0, ..., g_J)}{\partial x_j^k} s_{j'}^\star(\mathbf{x}|C') \tag{A.55}$$

$$= \gamma_k \sum_{C \in \mathbb{P}(j)} \frac{\partial \pi_C(g_0, ..., g_J)}{\partial g_{j'}} s_j^\star(\mathbf{x}|C) - \sum_{C' \in \mathbb{P}(j')} \frac{\partial \pi_{C'}(g_0, ..., g_J)}{\partial g_j} s_{j'}^\star(\mathbf{x}|C') \tag{A.56}$$

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

$$\frac{\frac{\partial s_j}{\partial x_{j'}^k} - \frac{\partial s_{j'}}{\partial x_j^k}}{\frac{\partial s_j}{\partial x_{j'}^{k'}} - \frac{\partial s_{j'}}{\partial x_j^{k'}}} = \frac{\gamma_k}{\gamma_{k'}} \tag{A.57}$$

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'.

A.6 DSC Model with Consideration Dependence on Non-Default Goods

The DSC model is also identified when consideration probabilities depend on some strict subset of non-default goods or if the impact of changes of at least two non-default goods on consideration probabilities are restricted to be the same.

In the first scenario, let there exist some $k \in \{1, ..., J\}$ such that $\partial \mu(\mathbf{p})/\partial p_k = 0$. Then one can identify changes in consideration probabilities from cross derivative asymmetries between any other good and good-k:

$$\frac{\partial s_k}{\partial p_j} - \frac{\partial s_j}{\partial p_k} = \frac{\partial \mu(\mathbf{p})}{\partial p_j} s_k^\star(\mathbf{p}|\mathcal{J}) \tag{A.58}$$

$$= \frac{\partial \log \mu(\mathbf{p})}{\partial p_j} s_k(\mathbf{p}|\mathcal{J}) \tag{A.59}$$

In the case where at least two non-default goods, k and k', are restricted to have the same impact on consideration probabilities,

$$\frac{\partial \log \mu(\mathbf{p})}{\partial p_k} = \frac{\partial \log \mu(\mathbf{p})}{\partial p_{k'}} \tag{A.60}$$

then changes in consideration probabilities can be recovered by inverting a system of linear equations.

For example, consider the $J = 2$ case.

$$\begin{bmatrix} -s_1 & (s_0 - 1) \\ -s_2 & (s_0 - 1) \end{bmatrix} \begin{bmatrix} \frac{\partial \log(\mu)}{\partial p_0} \\ \frac{\partial \log(\mu)}{\partial p_{\neg 0}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_0}{\partial p_1} - \frac{\partial s_1}{\partial p_1} \\ \frac{\partial s_0}{\partial p_2} - \frac{\partial s_2}{\partial p_1} \end{bmatrix} \tag{A.61}$$

This system is invertible when $s_1 \neq s_2$.

A.7 Correlated Unobservables in Utility & Consideration

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. However, allowing utility and attention to have correlated unobservables requires additional exclusion restrictions. Specifically, let there be a finite set of types $n = 1, ..., N$ such that:

$$s_{jm} = \sum_{n=1}^{N} \omega^n(w) s_j^n(x_m) \tag{A.62}$$

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:

$$s_{jm}^n = \mu_m^n \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l^n \prod_{l' \notin C} (1 - \phi_{l'}^n)\, s_{jm}^{n\star}(C) \tag{A.63}$$

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 A.62 (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.

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. In the ASC model, the probability of choosing any specific alternative as a function of the parameters $\psi = (\beta, \gamma)$ is given by:

$$s_j(\mathbf{p}; \psi) = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_l(p_l; \gamma) \prod_{l' \notin c} (1 - \phi_{l'}(p_{l'}; \gamma)) s_j^\star(\mathbf{p}|C, \beta) \tag{B.1}$$

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

$$\log \mathcal{L}(\psi) = \sum_{i=1}^{N} \sum_{r=1}^{R} \sum_{j=0}^{J} y_{irj} \log \left(s_j(\mathbf{p}_{ir}; \psi) \right) \tag{B.2}$$

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 and code.

B.2 Indirect Inference

In Section 4, 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 the probability of respondent i in round r choosing option j in our flexible logit model as:

$$\widetilde{s}_{irj} = \frac{\exp\left(\widetilde{u}_{irj}\right)}{\sum_k \exp\left(\widetilde{u}_{irk}\right)} \tag{B.3}$$

$$\widetilde{u}_{irj} = \theta_j + \theta_j^0 p_{irj} + \sum_{j'} \theta_{jj'} p_{irj} p_{irj'} \tag{B.4}$$

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, ψ:

$$\widetilde{\theta}(\psi) = M^{-1} \sum \widetilde{\theta}^m(\psi) \tag{B.5}$$

$$\widetilde{\theta}^m(\psi) = \arg\min_{\theta} \sum_{i=1}^{N} \sum_{r=1}^{R} \sum_{j=0}^{J} \widetilde{y}^m_{irj}(\psi) \log\left(s^m_{irj}(\theta)\right) \tag{B.6}$$

On each round we have the same number of simulated observations as 'real' observations: $y^m_{irj} = 1$ if given the set m of structural errors and structural parameters ψ, option j is selected on the simulated choice for prices $\mathbf{p}_{ir}$.

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

$$\hat{\psi} = \arg\min_{\psi} Q(\psi) \tag{B.7}$$

$$Q(\psi) = \left(\hat{\theta} - \widetilde{\theta}(\psi)\right)' W \left(\hat{\theta} - \widetilde{\theta}(\psi)\right) \tag{B.8}$$

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

B.3 Goodness of Fit

The number of parameters in the auxiliary model exceeds the number of structural consideration and preference parameters. This allows us to conduct a formal overidentification test. Given that we use the optimal weighting matrix $W = \Sigma_{\theta}^{-1}$, then

$$\frac{1}{1+\tau} \left(\hat{\theta} - \widetilde{\theta}(\hat{\psi})\right)' W \left(\hat{\theta} - \widetilde{\theta}(\hat{\psi})\right) \tag{B.9}$$

is asymptotically distributed as chi-squared with degrees of freedom equal to the difference in the number of auxiliary and structural parameters (98 in our case) (French 2005; Duffie and Singleton 1993; Pakes and Pollard 1989). τ gives the ratio between the number of observations to the number of simulated observations.

B.4 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 which you always consider (plan 0) and an alternative, plan 1, which you consider with probability ϕ_1. For ease of notation, we suppress the dependence of preferences and consideration on prices and other characteristics. Let preferences be given by:

$$u_{ij} = v_j + \epsilon_{ij} \tag{B.10}$$

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

$$s_1 \quad = \quad \phi_1 s_1^\star \tag{B.11}$$

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

$$\widetilde{u}_{ij} = v_j + \zeta_{j=1} + \epsilon_{ij} \tag{B.12}$$

where $\zeta_{j=1} = \zeta_1$ for plan 1 and is 0 otherwise, where ζ_1 is given by:

$$\zeta_1 = \ln\left(\frac{\phi_1 \exp(v_0)}{(1-\phi_1)\exp(v_1) + \exp(v_0)}\right) \tag{B.13}$$

This follows since:

$$s_1 = \frac{\exp(v_1 + \zeta_{j=1})}{\exp(v_1 + \zeta_{j=1}) + \exp(v_0)} = \phi_1 \frac{\exp(v_1)}{\exp(v_1) + \exp(v_0)} \tag{B.14}$$

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

$$\zeta_j = \ln\left(\frac{\phi_j \sum_{k\neq j} \exp(v_k + \zeta_k)}{(1-\phi_j)\exp(v_j) + \sum_{k\neq j} \exp(v_k + \zeta_k)}\right) \tag{B.15}$$

We showed above that this holds for the case where $J = 2$. Let s_j^a denote the probability of choosing good j conditional on paying full attention to good j, i.e. $s_j^a = s_j(x|\phi_j = 1)$. In the two-good case, $s_1^a = s_1^\star$ but more generally:

$$s_j^a = \sum_{C\in\mathbb{P}(j)} \prod_{l\in C, l\neq j} \phi_l \prod_{l'\notin C} (1-\phi_{l'})\, s_j^\star(C) \tag{B.16}$$

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

$$s_J = \phi_J s_J^a \tag{B.17}$$

By the inductive hypothesis, we have:

$$s_J^a = \frac{\exp(v_J)}{\exp(v_J) + \sum_{k \neq J} \exp(v_k + \zeta_k)} \tag{B.18}$$

Therefore,

$$s_J = \phi_J \frac{\exp(v_J)}{\exp(v_J) + \sum_{k \neq J} \exp(v_k + \zeta_k)} \tag{B.19}$$

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

$$u_{iJ} = v_J + \zeta_J + \epsilon_{iJ} \tag{B.20}$$

where:

$$\zeta_j = \ln\left(\frac{\phi_j \sum_{k \neq j} \exp(v_k + \zeta_k)}{(1 - \phi_j)\exp(v_j) + \sum_{k \neq j} \exp(v_k + \zeta_k)}\right) \tag{B.21}$$

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_{ij} depends on all quadratic functions of own and rival attributes can be derived as a 2nd order Taylor-expansion of ζ_j with respect to the attributes of rival goods around the point where all of these attributes are 0 so that $\zeta_j = 0$.

Next, consider the DSC model.

$$\begin{aligned} s_0 &= (1 - \mu) + \mu s_0^\star(\mathcal{J}) \\ s_j &= \mu s_j^\star(\mathcal{J}) \quad \text{for } j > 0 \end{aligned} \tag{B.22}$$

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

$$u_{ij} = v_j + \zeta_{j=d} + \epsilon_{ij} \tag{B.23}$$

Let $\zeta_{j=d} = \zeta$ and zero when $j \neq d$. The full-consideration model will be equivalent to the DSC

model with:

$$\zeta = \ln\left(\frac{1 + (1-\mu)\sum_{k\neq d}\exp(v_k - v_d)}{\mu}\right) \tag{B.24}$$

B.5 Additional Experimental Results

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

Table 7: Product Names and Prices

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

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

An additional question of interest that we do not address in the main text 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,[60] 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

[60] "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.

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.

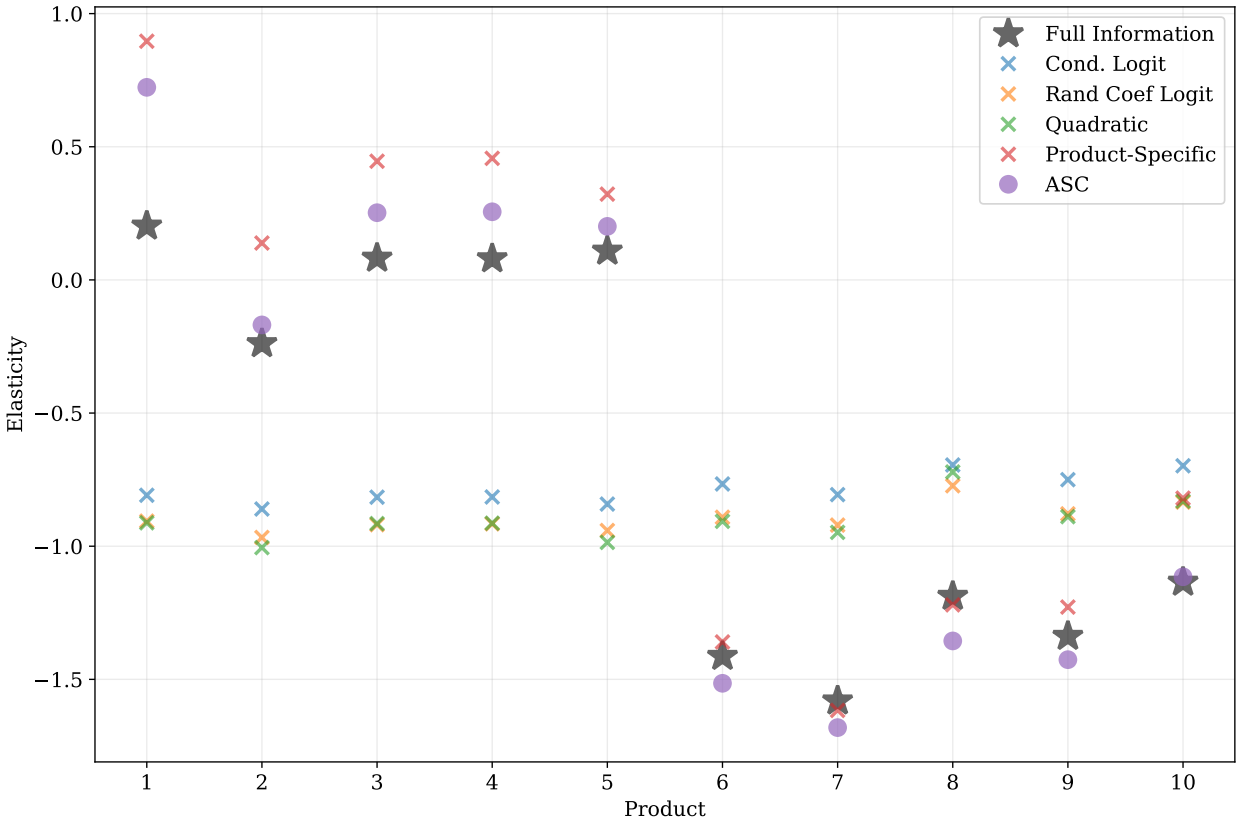
Figure B.1 shows the average own-price elasticities by good in each model. For goods 1-4, true own-price elasticities are positive because a higher price makes a good more likely to be considered. As noted above, this is an intentional feature of the model designed to mimic the fact that in some real world settings, consumers might be more likely to see higher priced items.

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 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. We formally test whether the product-specific model is able to fit the patterns captured by the auxiliary model using the test described in Section B.3. We reject this model at all conventional significance levels (p-value=0.0000; $\chi^2 = 2,279$).

Figure B.2 shows the predicted and observed choice probabilities (averaged over all goods) at different price levels conditional on a good being in the consideration set. While the conditional logit estimated on the full choice data performs poorly given then biased preference coefficients, the conditional-on-consideration conditional logit and ASC models perform well and closely match observed choice probabilities. This also points to the fact that the conditional-on-consideration logit is a good benchmark for comparison.

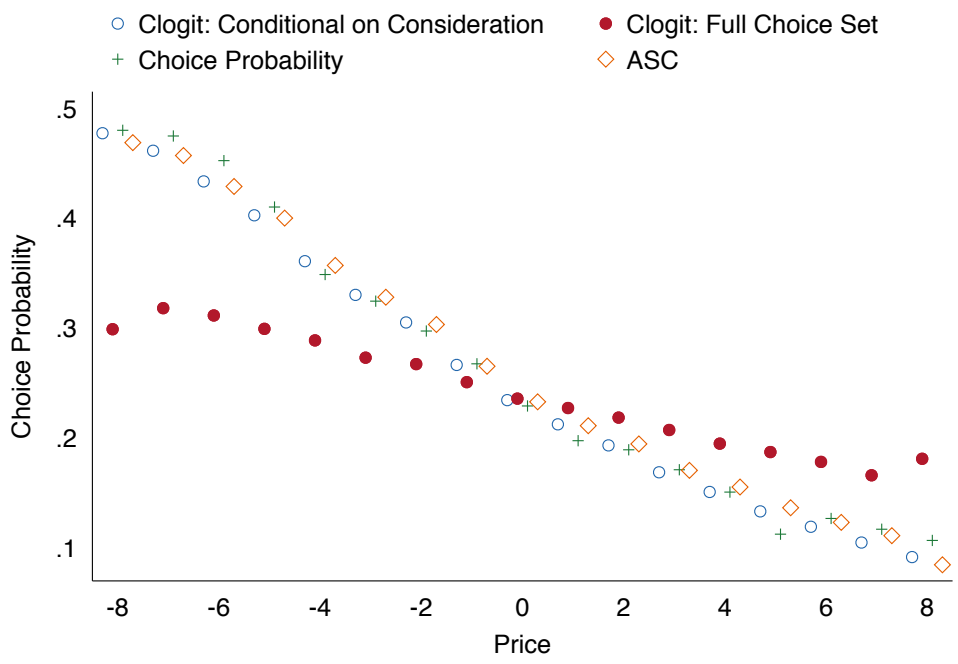
Table 8 reports estimates of the ASC model for the subset of experimental participants who correctly answered the question testing their understanding of the instructions. The results are very comparable to those in the text.

Figure B.1: 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 estimate conditional logit models with linear, quadratic, and alternative-specific price coefficients and with product-specific price coefficients.

Figure B.2: Predicted & Observed Choice Probabilities - Conditional on Consideration Set



Notes: Figure shows predicted choice probabilities for goods in the consideration set by the deviation of price from the average.

Table 8: Experimental Data Estimation Results

	Conditional Logit	ASC	Truth
Utility:			
Price (dollars)	-0.052***	-0.160***	-0.170***
	(0.004)	(0.033)	(0.005)
Product 1	-1.129***	1.561**	0.751***
	(0.084)	(0.769)	(0.109)
Product 2	-1.577***	0.143	-0.026
	(0.101)	(0.661)	(0.119)
Product 3	-1.331***	0.287	0.329***
	(0.091)	(0.582)	(0.111)
Product 4	-1.544***	0.393	0.234*
	(0.099)	(0.701)	(0.12)
Product 5	-1.162***	1.429*	0.664***
	(0.086)	(0.832)	(0.108)
Product 6	0.26***	0.487***	0.327***
	(0.056)	(0.136)	(0.066)
Product 7	-0.675***	-0.996***	-0.898***
	(0.073)	(0.181)	(0.081)
Product 8	-0.615***	-1.067***	-0.875***
	(0.07)	(0.2)	(0.079)
Product 9	-0.215***	-0.168	-0.311***
	(0.063)	(0.157)	(0.072)
Attention:			
Price (dollars)		0.158***	1.5
		(0.029)	
Product 1		-3.302***	-2.5
		(0.399)	
Product 2		-2.855***	-2.5
		(0.484)	
Product 3		-2.629***	-2.5
		(0.392)	
Product 4		-2.97***	-2.5
		(0.439)	
Product 5		-3.344***	-2.5
		(0.395)	
Product 6		-0.326	0
		(0.317)	
Product 7		0.638	0
		(0.795)	
Product 8		0.725	0
		(0.578)	
Product 9		-0.244	0
		(0.325)	

Notes: Table reports coefficient estimates from conditional logit and ASC 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 ASC model also includes a constant. *** Denotes significance at the 1% level, ** significance at the 5% level and * significance at the 10% level.

C Additional Results: Limited Consideration in Medicare Part D

In the main text, we decompose observed inertia into that attributable to limited consideration and to acclimation costs. In this section, we discuss two alternative explanations for inertia: an alternative source of switching cost, "paperwork costs", and spurious state dependence.

C.1 Model and Results with "Paperwork Costs"

We define "paperwork costs" as costs that are paid to choose a different plan from teh default, regardless of whether one has previously enrolled in either plan (Luco 2019). If consumers are reluctant to switch due to paperwork costs, being automatically defaulted into a new alternative can make them better off as these costs can then be avoided. However, inattentive consumers with large acclimation costs may be worse off from such a move as they are forced to pay the costs of navigating a new plan.

We start by modifying the model in the text to allow for both paperwork costs ρ and acclimation costs α rather than a monolithic ξ. Suppose utility conditional on being awake is given by:

$$u_{ijt} = \mathbf{x}_{ijt}\beta + (\alpha + \rho)Default_{ijt} + \epsilon_{ijt} \tag{C.1}$$

where ρ denotes paperwork costs that must be incurred whenever a consumer chooses a plan which is not the current default and α denotes acclimation costs that must be paid whenever a consumer chooses a plan they have not previously chosen (all other parameters are as in the text).

Independent price variation and cross-derivative asymmetries alone do not allow us to separately identify ρ and α — both of these are utility-relevant factors. To separate these, we make use of variation in our context generated by the random reassignment of low-income subsidy (LIS) beneficiaries into new plans (explained below). When these beneficiaries no longer qualify for full premium subsidies, utility is given by:

$$u_{ijt} = x_{ijt}\beta + (\alpha + \rho)Default_{ijt} + \alpha\left(Default_{ij,t-1} \times Reassigned_{ijt}\right) + \epsilon_{ijt} \tag{C.2}$$

If they choose to switch back to their original default, they must pay paperwork costs ρ but not acclimation costs since they already have experience with that plan (and thus their utility "bonus" relative to other plans is the acclimation costs). If they choose any plan other than the original or current default, they must pay both paperwork and acclimation costs.[61]

[61] Note that this variation is not identical to that possible under the smart default policy. In the data, LIS beneficiaries remain enrolled in their assigned plan for a year and thus always face acclimation costs. Under the smart default

Allowing for both paperwork and acclimation costs, the change in welfare from smart defaults can be expressed as:

$$\begin{aligned}\triangle W_i &= W_i^1 - W_i^0 \\ &= \rho\left(s_{id}^1 - s_{io}^0\right) + \alpha\triangle s_{io} + \sum_j \triangle s_{ij} v_{ij}\end{aligned} \tag{C.3}$$

Defaults now change welfare through three channels: the first term captures the impact via paperwork costs. Changing the default increases welfare by avoiding paperwork costs if the old default is chosen at a lesser rate than the new default under the previous regime. The second term captures acclimation costs; these will be paid by anyone who switches to a new plan as a result of the new default. Finally, the third term captures the direct effect of the change of defaults on normative utility not due to inertia (e.g. inducing people to choose lower cost plans).

Data & LIS Beneficiaries To separately identify paperwork and adjustment costs (Equation C.2), we combine our usual sample with a sample of low-income subsidy (LIS) beneficiaries who are randomly assigned to alternative plans, then earn enough that they no longer qualify for subsidies and must make an active choice. In the LIS program, those who qualify for low-income subsidies in Medicare Part D and do not explicitly opt out are randomly reassigned each year to a plan with premiums below the low-income subsidy amount (see Decarolis (2015) for more details of this policy).

Due to heavily subsidized cost-sharing, reassigned LIS beneficiaries experience little cost differentiation across plans in the years that they are eligible for reassignment. However, in subsequent years, individuals might no longer qualify for full premium or copay subsidies and thus face substantial cost variation across candidate plans. In our 20% sample of Medicare Part D data, we observe 2,852 LIS beneficiaries who are randomly defaulted into a plan and who then in subsequent years do not qualify for program thus must pay premiums for the plan in which they enroll.

To estimate our structural model, we use a pooled sample combining the random sample of non-LIS beneficiaries with reassigned enrollees who lost LIS status. We use the variation from reassigned LIS beneficiaries to separately identify acclimation and paperwork costs. We use variation from non-reassigned beneficiaries to identify the degree of inattention. To identify acclimation and paperwork costs given inattention, we must assume that acclimation and paperwork costs from the reassigned sample are the same as in the non-reassigned sample.

policy, individuals would immediately be given the option of enrolling in their original plan. The structural parameters we identify nonetheless allow us to evaluate the smart default policy given our estimates of α and ρ.

Results Table 9 give our structural results. The parameters imply that the majority of switching costs are acclimation costs ($194) rather than paperwork costs ($21). This implies that our assumption in the main text that all switching costs are acclimation costs does not materially impact on our welfare conclusions.

C.2 Spurious state dependence

Using the terminology of Heckman (1981), "structural state dependence" arises when the utility from a given choice depends directly on what choices were made in the past while "spurious state dependence" arises when choices are correlated over time because of some unobserved attribute of choices. Distinguishing between spurious versus structural state dependence is thus important for forecasting what fraction of consumers will return to their original plans if defaulted away.[62]

To separate these, we again use variation from reassigned LIS beneficiaries. We also assume that spurious state dependence, if it is present at all, arises at the level of brands. This is consistent with the existing literature and mechanisms proposed for such spurious dependence (Ketcham, Kuminoff, and Powers 2019).[63] Given this assumption, we can separate spurious state dependence from acclimation costs by asking: are consumers randomly assigned to a plan within the same brand more likely to remain in their new default than consumers randomly assigned to a plan in a different brand? In the data, we observe only a small number of beneficiaries who actively choose a plan (with stakes), then qualify for reassignment due to low-income subsidies and who then earn enough that they actively choose again. Among these few hundred beneficiaries, we see that beneficiaries reassigned to a plan from the same brand are slightly *less* likely to be inertial, although the estimate is imprecise (our regression estimate is 3 percentage points less likely, with a standard error of 8 percentage points). This suggests that spurious state dependence is not the primary driver of inertia in our data. In our model, we thus attempt to decompose the reasons for structural state dependence.

[62]Conventionally, spurious and structural state dependence are separately identified based on whether consumers who are reassigned disproportionately return to their original choice (Raval and Rosenbaum 2018). The conventional rationale is that if consumers return to their original choice, this is due to spurious state dependence (they liked something about that choice). This test is incomplete: if consumers are reassigned to an unsuitable plan and it is costly to become acclimated to that plan (a form of structural state dependence), they may return to their original plan because choosing any other plan would require paying the acclimation costs to learn about that alternative plan. Thus, both spurious state dependence and acclimation costs would suggest that reassigned individuals would disproportionately return to their original plans relative to other plans in the data.

[63]Ketcham, Kuminoff, and Powers (2019) estimate a parametric model which suggests that the portion of utility not explained by brand fixed effects is better accounted for by "optimization error" than "tastes".

Table 9: Hybrid Model including Paperwork Costs

	Utility	$\phi(\cdot)$	$\mu(\cdot)$
Annual Premium (hundreds)	-1.2484***	-0.4581***	0.5944***
	(0.0369)	(0.0481)	(0.037)
Annual Out of Pocket Costs (hundreds)	-0.3507***	0.2137***	0.0152
	(0.0548)	(0.0795)	(0.0286)
Variance of Costs (millions)	0.2924***	-0.284***	-0.0603
	(0.0754)	(0.0887)	(0.0376)
Deductible (hundreds)	-0.9845***	-0.7612***	0.5528***
	(0.0718)	(0.127)	(0.085)
Donut Hole Coverage	1.0365***	0.8174	-2.9507***
	(0.2465)	(0.505)	(0.1607)
Average Consumer Cost Sharing %	-1.1136***	-0.0223	-0.1009
	(0.0763)	(0.159)	(0.0862)
# of Top 100 Drugs in Formulary	0.0111	0.6565***	-0.1337***
	(0.0371)	(0.1054)	(0.0529)
Normalized Quality Rating	-0.0786***	0.3557***	-0.2347***
	(0.0075)	(0.016)	(0.008)
Default	2.6952***		
	(0.1172)		
Old Default × Reassigned	2.4268***	-0.0260	
	(0.8616)	(0.9613)	
Constant		1.0517***	-1.136***
		(0.155)	(0.0729)
Acclimation Cost	$194		
Paperwork Cost	$21		

Notes: This table gives estimates from the "hybrid" version of the consideration set framework with a separate "old default" dummy. Estimates in all models are the coefficients in the utility and attention equations (not marginal effects). The coefficients in the DSC component of the model, $\mu(\cdot)$, are the coefficients on the listed characteristics of the default good. The coefficients in the ASC component of the model, $\phi(\cdot)$, are the coefficients on the listed characteristics of good j on the likelihood of j being considered. The model also includes an indicator for plans that are missing the # of top 100 drugs in formulary variable as well as an interaction of variance of costs and an indicator for individuals with no claims. Standard errors in parentheses. *** denotes significance at the 1% level, ** at 5% level, and * at 10%.

C.3 Testing the Validity of Additional Exclusion Restrictions

As we can allow all characteristics to enter consideration and utility, we are able 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: 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.[64]. The utility coefficients relevant for these tests are reported in Table 10. Several of the exclusion restrictions in Heiss, McFadden, Winter, Wupperman, and Zhou (2016) are rejected. Like Heiss, McFadden, Winter, Wupperman, and Zhou (2016), we find that consumers are more likely to wake up if their plan increases premiums, out of pocket costs or deductibles. However, we find that, conditional on the level of these variables, consumers are also more likely to choose plans that experienced a large increase. Put differently, if premiums are high today, even accounting for inertia, consumers are more likely to choose a plan if it had low premiums yesterday. This might occur if, for example, consumers are more likely to stay asleep for plans which have had good outcomes for them in the past. 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.[65] The remainder of Table 10 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 – younger beneficiaries are more sensitive to premiums, and older and non-white beneficiaries are less sensitive to high deductibles.

Despite this, we find that imposing the above exclusion restrictions in the model has only a small effect on the estimated attention probability, which decreases from 19.1% in the model with

[64] We use the DSC model here rather than the hybrid model following Heiss, McFadden, Winter, Wupperman, and Zhou (2016)

[65] An 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.

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 inHeiss, McFadden, Winter, Wupperman, and Zhou (2016), we find that these are not large enough to qualitatively change their results.

Table 10: Utility Coefficients for Overidentification Test

	Coef.	Interactions			
		I(Age 70 - 79)	I(Age $\geq$ 80)	I(Non-White)	Experience
Change in Annual Premium (hundreds)	0.349***				
	(0.016)				
Change in Out of Pocket Costs (hundreds)	0.052***				
	(0.004)				
Change in Deductible (hundreds)	0.190***				
	(0.018)				
Annual Premium (hundreds)		-0.095***	0.061*	0.378***	-0.035
		(0.022)	(0.024)	(0.045)	(0.024)
Annual Out of Pocket Costs (hundreds)		-0.017	-0.023	0.026	-0.039
		(0.028)	(0.029)	(0.052)	(0.027)
Variance of Costs (millions)		0.143	0.313	-0.486	-1.526
		(0.393)	(0.399)	(0.717)	(0.848)
Deductible (hundreds)		-0.023	0.150***	0.344***	-0.139***
		(0.029)	(0.031)	(0.054)	(0.028)
Donut Hole Coverage		-0.351**	-0.446***	-0.297	0.193
		(0.118)	(0.126)	(0.224)	(0.137)
Average Consumer Cost Sharing %		-0.812*	-0.528	2.055**	0.223
		(0.366)	(0.390)	(0.675)	(0.355)
# of Top 100 Drugs in Formulary		0.037**	0.023	-0.062*	0.018
		(0.014)	(0.015)	(0.026)	(0.014)
Normalized Quality Rating		0.108**	-0.005	-0.307***	0.227***
		(0.032)	(0.035)	(0.065)	(0.032)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes:

Table reports the utility coefficients from the overidentification test where we estimate the DSC model specification reported in column 2 of Table 4 in the text but also include interactions with age, race and experience and allow changes in product attributes to impact both attention and utility. Standard errors in parentheses. *** denotes significance at the 1% level, ** at 5% level, and * at 10%.

Table 11: Welfare Impact of Smart Default Policy - Restricted Reassignment to those with Savings Greater than Switching Costs & Into High Quality Plan

	Market Shares Conditional on Being Switched			Attention Cost			
	Smart Default	Previous Plan	Other Plan	$0	$50	$100	$200
Mean Welfare Change: Full Sample							
Hybrid Parameters	0.947	0.001	0.052	$79	$79	$78	$78
Direct Effect on Attention Probability							
25%				$74	$73	$72	$70
50%				$63	$60	$58	$52
75%				$52	$48	$43	$34
100%				$42	$35	$29	$16
Proportion Consumers Switched: 25%							
Mean Welfare Change: Conditional on Being Switched							
Hybrid Parameters	0.947	0.001	0.052	$530	$529	$528	$526
Direct Effect on Attention Probability							
25%	0.901	0.002	0.097	$496	$490	$483	$471
50%	0.803	0.004	0.194	$424	$406	$387	$350
75%	0.705	0.005	0.290	$352	$321	$290	$228
100%	0.606	0.007	0.387	$281	$237	$194	$107

Notes: The table shows overall welfare impacts of alternative assumption about the probability of paying attention and the cost of paying attention. Each row shows a different assumption about the probability of paying attention, while each column shows alternative assumptions about the cost of paying attention.

C.4 Robustness to Alternative Specifications

Table 11 shows the results of our counterfactual simulation using the model in the text when we only consider reassigning beneficiaries to plans in the top quartile of quality rating. Table 12 shows welfare results for a modified smart default policy under which consumers are only switched to plans in the same brand as the default. To make this a fairer test, we also include a brand inertia dummy in our positive choice model in order to generate this counterfactual. The welfare effects of this policy are negative. First, the majority of individuals are in the lowest cost plan provided by their brand. Second, the reduction in cost possible is typically smaller than acclimation costs for within-brand switches.

Table 12: Welfare Impact of Smart Default Policy: Including Brand Inertia & Only Switching Within Brand

	Attention Cost			
	\$0	\$50	\$100	\$200
Hybrid Model	-\$98	-\$99	-\$100	-\$104
Direct Effect on μ				
25%	-\$98	-\$101	-\$101	-\$110
50%	-\$99	-\$102	-\$104	-\$113
75%	-\$100	-\$103	-\$106	-\$115
100%	-\$102	-\$106	-\$111	-\$119
Proportion Switched: 31%				

Notes: The table shows overall welfare impacts of a smart default policy where individuals are only switched to plans within the same brand as the original choice. Each row shows an alternative assumption about the direct effect of the policy on the probability of paying attention and the cost of paying attention. The estimates are based on a positive model that allows utility to depend on whether a plan is in the same brand as the default.