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

Jason Abaluck and Abi Adams

*

August 2, 2019

Abstract

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

1 Introduction

Discrete choice models generally assume that consumers consider all available options when making their choices. This prevents researchers from asking many questions of interest. What factors lead consumers to become aware of more options? Will inertial consumers 'wake up' in response to a price increase but remain unresponsive if rivals lower prices? Which products will respond well to advertising because they are valued highly be consumers who see them? Normatively, whether people eat the same foods and go to the same stores year after year because they like those options or

*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, Aureo de Paula, Jonathan Feinstein, Jeremy Fox, Xavier Gabaix, Jonathan Gruber, Phil Haile, Erzo Luttmer, Paola Manzini, Marco Mariotti, Olivia Mitchell, Francesca Molinari, Fiona Scott Morton, Barry Nalebuff, 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.

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.

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.[1] 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).[2] 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. These exclusion restrictions are often questionable and in tension with economic theory: for example, does advertising only impact choices via informing consumers about which goods exist?

[1] The behavioral decision theory literature typically does not provide rigorous microfoundations for the formation of consideration sets. As in this paper, much of the prior literature either takes consideration sets as primitives for the consumer or their relationship to observables is modelled in a relatively reduced form way (see Caplin, Dean, and Leahy (2018) for a discussion).

[2] 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.

(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 that highlights the identifying power of theoretical restrictions in the context of consideration set models (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014; Cattaneo, Ma, Masatlioglu, and Suleymanov 2018). We prove that the restrictions on choice probabilities imposed by economic theory are sufficient for point identification of preferences and consideration probabilities in the DSC and ASC frameworks as well as hybrid models combining features of both alternatives. Our method does not require auxiliary information on consideration sets and it allows all observables to impact both consideration and utility. We provide simple closed form expressions for consideration set probabilities in terms of differences in cross-derivatives (the discrete choice analogue of 'Slutsky asymmetries'). Our framework subsumes many of the consideration set models in the applied literature and does not rely on assuming a particular functional form for random utility errors.

Our identification result builds on the insight that imperfect consideration breaks the symmetry between cross-price responses (or more generally, cross-characteristic responses). For example, in a model with a default, symmetry would ordinarily require that switching decisions be equally responsive to an increase in the price of the default good by \$100 or a decrease in the price of all rival goods by \$100. Suppose instead that consumers will be inattentive and choose the default option unless the default good becomes sufficiently unsuitable. Now, switching decisions will be unresponsive to changes in the price of rival goods but more responsive to changes in the price of the default to the degree that these changes perturb attention (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 any generality.[3] Our framework implies that ad hoc attempts to model consideration sets such as fixed effects in utility for products on different shelves or interactions between prices and such fixed effects can still yield misspecified models because they do not relax the symmetry assumption.

Our results imply that the assumptions already made in many discrete choice applications are sufficient to identify frameworks that allow for limited consideration.[4] In cross-sectional data, our

[3] 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.

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

results can be used to identify whether goods are demanded because they are high-utility or because they are more likely to be considered. In panel data, one can evaluate whether inertia reflects utility-relevant factors or inattention. More generally, one can perform behavioral welfare analyses with no additional data beyond what is needed to estimate conventional discrete choice models. We also show that the consideration set framework considered in this paper is generally over-identified, and if instruments are available, one can use our results to test the validity of additional exclusion restrictions.

Our identification proof is constructive and so, in theory, consistent nonparametric estimators could be based upon it. However, in most applications of interest, we advocate estimating parametric generalizations of conventional models. We consider two approaches: indirect inference and maximum likelihood. 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 the asymmetries we see in the data. We also show how to use such flexible auxiliary models to test whether parameters estimated by maximum likelihood explain well the asymmetries observed in reduced form data. Maximum likelihood estimation makes less explicit the link between estimation and identification but may be computationally more feasible.

We validate our approach in a lab experiment in which participants made a series of choices from (known) proper subsets of 10 possible goods. Using only data on choices and ignoring information on what items were available, matching Slutsky asymmetries enables us accurately to recover the probabilities that each good was available as well as recovering the preference parameters that we would estimate conditional on knowing which items were available. Conventional models with a comparable number of parameters misspecify own- and cross-price elasticities relative to the elasticities computed using data on which items were actually available. The average absolute error in cross-elasticities in conventional models is 2.5 times larger than in our consideration set model model or 45.5 percentage points larger as a fraction of the average absolute cross-elasticity.

We then apply the Default Specific Consideration model to analyze "Smart Defaults" for prescription drug insurance choices in Medicare Part D (Handel and Kolstad 2018). In this market, 90% of beneficiaries are inertial. We could default beneficiaries to plans that would be lower cost on average, but whether this makes beneficiaries better off depends on why they were inertial. If beneficiaries are inertial because it is costly for them to acclimate to a new plan, then defaulting them to a new plan may make them worse off by forcing them to pay these acclimation costs, especially if they are inattentive. Alternatively, if beneficiaries are inertial because they are asleep (i.e. they rarely make active choices) or because it is costly to fill out the paperwork necessary to switch plans, they may be made better off by being assigned to a lower cost plan. To separate these mechanisms, we consider the choices of consumers who in past years received low-income subsidies and qualified for random reassignment to alternative plans. We find that these consumers rarely return to their original plans, although they do so at a rate higher than we would expect given observables, implying non-trivial acclimation costs. Models without inattention imply that these acclimation costs roughly offset the benefits of reassignment. In the model with inattention, the dollar-equivalent value of these acclimation costs is smaller as consumers appear more price elastic conditional on paying attention. Defaulting consumers to lower cost plans is then seen to produce large benefits.

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 discusses estimation of our consideration set framework, developing an indirect inference estimator in which structural parameters are chosen to match cross-derivative asymmetries in the data. We then take two special cases of our framework to data. Section 5 validates the practical relevance of our result using data from a lab experiment to estimate the Alternative Specific Consideration model. Section 6 applies the Default Specific model to data from Medicare Part D to separately identify switching costs from consumer inattention. Section 7 concludes.

2 Related Literature

Before turning to a formal treatment of our framework, we situate our contribution within the existing literature on the identification of consideration set models.

Exclusion Restrictions Identification of consideration set models is typically achieved by imposing an additional set of restrictions on 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] Goeree (2008), Hortaçsu, Madanizadeh, and Puller (2015), Gaynor, Propper, and Seiler (2016), and Heiss, McFadden, Winter, Wupperman, and Zhou (2016) proceed in this way. These exclusion restrictions are often questionable and can be in tension with economic theory; excluding prices from consideration, for example, can be in tension with simple models of rational inattention. Yet, agnosticism over what variables impact utility and which impact attention is usually associated with

[5] 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.

only partial identification of the objects of interest (Lu 2016; Barseghyan, Coughlin, Molinari, and Teitelbaum 2018). Our identification result does not rely on the existence of variables that influence consideration but not utility and vice versa.

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 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 attention 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 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 making use of the restrictions on choice probabilities that are implied by economic theory. 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 main result relies 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] In a version of the Default Specific 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 own at the cost of losing point identification of the structural functions of interest. Our contribution in this paper 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 (2017) 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

[6] Davis and Schiraldi (2014) provide generalizations of multinomial logit models that permit asymmetries, but they explicitly note that these models cannot be rationalized by an underlying random utility interpretation and do not attempt to use these asymmetries to identify inattention.

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 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 the standard assumptions made in many discrete choice settings are sufficient to identify commonly employed frameworks that allow for imperfect consideration. Our central insight is that violations of Slutsky symmetry constructively identify the probability that consumers consider various subsets of products. Our results demonstrate the practical identifying power of deviations from the neoclassical model and show that additional exclusion restrictions are not necessary in order for applied researchers to allow for imperfect consideration within the wide class of models considered here. We direct the applied reader to Section 3.4 where the intuition for our results is given in a simple example.

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, with support $\chi \subseteq \mathbb{R}_{++}$. Our framework naturally incorporates additional characteristics (x_j), consumer microdata (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 x_j and z_i in what follows. Our identification argument focuses on price variation, although it of course extends to variation in any attribute satisfying the assumptions we state below.

We allow for individuals to consider an (unobserved) subset of available goods when making their choice. The set of goods that a consumer considers is called the consideration set. At this point, we place no restrictions on consideration set formation except that there exists a default option, good-0, that is always considered. Let $\mathcal{P}(\mathcal{J})$ represent the power set of goods, with any given element of $\mathcal{P}(\mathcal{J})$ indexed by C. The set of consideration sets containing good j is then given as:

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

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 $\mathbf{p} = [p_0, ..., p_J]$, $\pi_C(\mathbf{p})$ 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 Individuals make choices from any given consideration set to maximize their utility. We take a standard additive random utility model (RUM) approach, decomposing individual i's utility from good j, u_{ij}, into a deterministic component that depends on the good's characteristics and an additive random error term.

Assumption 1. *Random Utility Model:* Indirect utility is of the form:

$$u_{ij} = v_i(p_j) + \epsilon_{ij} \tag{3.4}$$

$$= \alpha_j - \beta_i p_j + \epsilon_{ij} \tag{3.5}$$

where $\beta_i \sim F_\beta(\cdot)$ where β_i is independent of $\mathbf{p}$, and in which the distribution of the random error term, $F_\mathcal{J}(\cdot)$ has unknown covariance matrix Ω and satisfies:

1. Independence: $\epsilon_{ij} \perp p_{j'}$.

2. Absolute Continuity: the joint distribution of ϵ is absolutely continuous with respect to

Lebesgue measure on $\mathbb{R}^{J+1}$.

3. Non-Defectiveness: the joint distribution of ϵ is non-defective, i.e. $\lim_{\epsilon \to -\infty} F_{\mathcal{J}}(\cdot) = 0$ and $\lim_{\epsilon \to \infty} F_{\mathcal{J}}(\cdot) = 1$.

While point identification of consideration probabilities requires only additive separability from other characteristics,[7] in the proof in the main text we will make the stronger assumption of quasi-linearity (Equation 3.5) for expositional simplicity. The key restriction that consumers value the separable characteristic equally across choices can be substantive, although it is often theoretically well-motivated.[8]

The probability that option j is chosen given consideration of the set of options C, with $j \in C$, is then given by:

$$s_j^{\star}(\mathbf{p}|C) = Pr\left(\alpha_j - \beta_i p_j + \epsilon_{ij} = \max_{j' \in c} \alpha_{j'} - \beta_i p_{j'} + \epsilon_{ij'}\right) \tag{3.6}$$

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. 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:* the observables consist of the variables:

$$\left\{s_j, p_j, \partial s_j / \partial p_{j'}\right\}_{j,j' \in \mathcal{J}} \tag{3.7}$$

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.[9] 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 (fol-

[7] Albeit with more stringent data requirements.

[8] This restriction will fail when the value of a given attribute depends highly on other attributes of the good: for example, the value of a warranty for a product may depend on the value of the product in the first place. In our canonical case of price, provided income-effects are small or non-existent, the restriction that β_i is the same across goods follows immediately from the structural restriction that consumers have a constant marginal utility of income.

[9] With microdata in which attributes vary across individuals, saying we observe many individuals in each "market" means that we observe many individuals with each possible choice set (where choice sets are characterized by a set of alternatives and attributes for those alternatives). This may be satisfied in a single choice environment with rich individual variation in attributes.

lowing Berry and Haile (2014)).

Assumption 3. *Price Variation*: p_j is continuously distributed for all j. Depending on what is assumed about consideration set formation, there are four assumptions, of differing degrees of strength, that we will impose on the degree of price variation in our proofs to come:

1. Assumption 3a. *Independent Price Variation:* $Var(p_j|p_{j'}) \neq 0$ for all $j \in \mathcal{J}$, $j' \neq j$.
2. Assumption 3b. *Large Support for* p_0: $\exists \{p_j\}_{j \in \mathcal{J}/0}$ such that supp $p_0|p_1, ..., p_J = R_{++}$.
3. Assumption 3c. *Large Support for* p_j: $\exists \{p_{j'}\}_{j' \in \mathcal{J}/j}$ such that supp $p_j|\{p_{j'}\}_{j' \in \mathcal{J}/j} = R_{++}$ for all $j \in \mathcal{J}$.
4. Assumption 3d. *Full Support*: supp $p_j|\{p_{j'}\}_{j' \in \mathcal{J}/j} = R_{++}$ for all $j \in \mathcal{J}$.

Assumption 3d is the strongest of our price variation assumptions, essentially requiring that there is sufficient price variation to move choice probabilities through the entire unit simplex for each good. Equivalent conditions are assumed in prior work on multinomial choice without consideration sets (see the discussion in Barseghyan, Molinari, and Thirkettle (2018) and Lewbel (2000)). This assumption thus provides a natural benchmark for exploring identification under ideal conditions. However, we will also outline results with substantially weaker assumptions to demonstrate what might be identified in scenarios with limited price variation. Assumption 3c generates variation equivalent to the assumption of leave-one-out choice sets made in Kawaguchi, Uetake, and Watanabe (2016), which is sufficient for identification of consideration probabilities in our most complex model.[10]

3.2 Slutsky Asymmetries & Nominal Illusion

Conditional on choosing from a given consideration set, choice probabilities will satisfy the standard Daly-Zachary conditions (Daly and Zachary 1978), notably cross-derivative symmetry and an absence of nominal illusion.[11] Thus, in our baseline model, only one mechanism is available to gener-

[10]But in this paper, we do not make use of any further exclusion restrictions as were required in Kawaguchi, Uetake, and Watanabe (2016).

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

ate cross-derivative asymmetries: imperfect consideration.[12] Theorem 1 makes this point formally.[13]

Theorem 1. Asymmetries & Nominal Illusion Imply Imperfect Consideration. Given Assumption 1 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.

3.3 Consideration Set Framework

To make progress towards point identification of the structural functions of interest, we must place some additional restrictions on consideration set probabilities. If $\pi_C(\mathbf{p})$ 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).[14]

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

[12]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 and individual characteristics would also satisfy the Daly-Zachery conditions.

[13]Later in this section, we describe overidentification tests to determine whether the pattern of asymmetries is consistent with our model of imperfect consideration as opposed to resulting from violations of our underlying assumptions or other behavioral anomalies.

[14]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).

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

Non-experimental applied work has (until now) had to rely on a further set of assumptions on what variables are excluded from utility and from consideration for identification of these models.[15] In this paper we show that these additional exclusion restrictions are unnecessary: the structure already imposed within the random utility framework itself suffices for identification in these models. Thus, our results highlight a new source of identifying variation for researchers who do not have access to auxiliary data on what options were actively considered by consumers.

Further, that we can allow all product characteristics to influence both utility and consideration brings benefits to those wanting to test alternative theories of inattention. Our results allow 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 specifications 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).

Relative to certain fully structural 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. On the other hand, our agnosticism means that we cannot identify how out of sample variation will impact attention (for example, what if we assign individuals to a new default, how many will "wake up" and make an active choice?). Our model alone also cannot identify the costs that individuals must pay to search or pay attention. One could use the search probabilities recovered by our model

[15] Aguiar, Boccardi, Kashaev, and Kim (2018) were able to proceed without such assumptions because their experimental design allowed them to observe choice behavior from every potential subset of feasible options (a "rich domain" assumption).

to help identify a structural model which would identify these costs. Alternatively, one might prefer to remain agnostic. In our empirical example, we highlight these points and evaluate the robustness of our normative evaluation to alternative assumptions about these unknown values.

This being said, the ASC and DSC models do still 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.8). 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.

In the remainder of this section we show that consideration probabilities are identified from cross-derivative asymmetries in the DSC model, ASC model, and a hybrid framework subsuming the ASC and DSC models. If interest lies in scenarios that cannot be nested within this hybrid framework, in Appendix A we show in a more general environment that features of consideration probabilities are identified using the same methods.[16]

3.4 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 $s_0^\star(\mathbf{p}|\mathcal{J})$ gives choice probabilities conditional on considering all products.

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

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.

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

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. This allows us to subsume a large number of frameworks that microfound DSC-type models.

Identifying Changes in Consideration Probabilities The key to our identification argument is that maximizing behavior implies symmetric cross-price derivatives given full consideration (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:

$$\begin{aligned} \frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} &= \mu_0 \frac{\partial s_j^\star(\mathbf{p}|\mathcal{J})}{\partial p_0} + \frac{\partial \mu_0}{\partial p_0} s_j^\star(\mathbf{p}|\mathcal{J}) - \mu_0 \frac{\partial s_0^\star(\mathbf{p}|\mathcal{J})}{\partial p_j} && (3.11) \\ &= \frac{\partial \mu_0}{\partial p_0} s_j^\star(\mathbf{p}|\mathcal{J}) && (3.12) \\ &= \frac{\partial \log(\mu_0)}{\partial p_0} s_j(\mathbf{p}) && (3.13) \end{aligned}$$

where $\mu_0 \equiv \mu(p_0)$. 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 Assumption 1, 2 and 3a, then $\partial \log(\mu_0) / \partial p_0$ is constructively identified.

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

$$\log(\mu_0) = \int \frac{1}{s_j(\mathbf{p})} \left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} \right] \mathrm{d}p_0 + \alpha \tag{3.15}$$

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

Assumption DSC. As $p_0 \to \infty$, $\mu_0(p_0) \to 1$.

Theorem 3. Identification of the DSC Model. Given Assumption 1, 2, 3b, and DSC, then consideration probabilities are constructively identified as:

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

Commonly employed functional form assumptions on consideration substantially reduce the amount of price variation required to identify the level of consideration, even when no further parametric assumptions are placed on preferences. For example, let consideration take the following form:

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

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 and no large support assumptions are required.

Theorem 4. Identification of the DSC Model with Logit Consideration. Given Assumption 1, 2, and 3a, then $[\gamma_0, \gamma_p, \{s_j^\star(\mathbf{p})\}_{j\in\mathcal{J}}]$ are identified where

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

Proof in Appendix A.

3.5 The Alternative Specific Model

The 'Alternative Specific Consideration' (ASC) model is a consideration set framework employed in the marketing and empirical industrial organization literatures, which allows for richer patterns of consideration across goods than the DSC 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:

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

Our central insight is that, even in this much 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 much richer setting, changes in consideration probabilities can be expressed as a linear function of observables. Let $\bar{\mathbf{p}}_j$ give the price vector $\mathbf{p}$ with p_j taking on a high value. As shown fully in Appendix A, one can express cross derivative differences between default and non-default products as:

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

While cross derivative differences for $j, j' \neq 0$ can be expressed as:

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

Note the power of Equations 3.21 and 3.22: they relate unobservable changes in consideration probabilities to observed cross derivative differences and market shares.

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 with an outside default. For simplicity, in the main text, we provide the just-identified conditions for an inside-default. 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). 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 at very high prices; a weak assumption that is easily tested.

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

Assumption ASC.ii: $s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j) \neq 0$.

Theorem 5. Identification of Consideration Probabilities in the ASC Model. Given Assumption 1, 2, 3c, ASC.i, and ASC.ii, then $\phi_j(p)$ for $j = 1, ..., J$ are constructively identified:

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

Identifying Latent Market Shares "Nominal illusion" facilitates the identification of the 2^J independent latent choice probabilities in the ASC model, $s_j^\star(\mathbf{p}|C)$.[17] Imagine that $N = 2^J$ price shifts are observed (meaning that prices for all goods are perturbed by the same constant). Given

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

quasi-linearity, these price shifts alter consideration probabilities but do *not* alter latent choice probabilities conditional on consideration.[18] In Appendix A, we show how this variation is sufficient to identify latent choice probabilities.

> Theorem 6. Identification of the ASC Model. Given Assumption 1, 2, 3d, ASC.i, ASC.ii and ASC.iii, then $\phi_j(p_j)$ for $j = 1, ..., J$ and latent market shares conditional on consideration set C, $s_j^\star(\mathbf{p}|C)$ are identified for all $C \in \cup_{j=0}^J \mathbb{P}(j)$. Proof in Appendix A.

3.6 The Hybrid Consideration Set Model.

The assumptions made to identify the ASC model are also sufficient to identify a hybrid model that combines the ASC and DSC models. Let the market share of the inside default, good-0, and non-default goods take the form:

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

where $\phi_0(p_0) = 1$ for all $p_0 \in \chi$. Restricting $\phi_j(p_j) = 1$ for all $j > 0$ gives the DSC model. Restricting $\mu(p_0) = 1$ gives the ASC model.

> Theorem 7. (Identification of Hybrid Consideration Set Model) Given Assumption 1, 2, 3d, DSC, ASC.i, ASC.ii and ASC.iii, then $\mu(p_0)$ and $\phi_j(p_j)$, and latent market shares conditional on consideration set C, $s_j^\star(\mathbf{p}|C)$ are identified for all $C \in \cup_{j=0}^J \mathbb{P}(j)$. Proof in Appendix A.

3.7 Overidentification

Given our assumptions, imperfect consideration is the only mechanism giving rise to an asymmetric cross-derivative matrix. Relaxing our background assumptions might, however, give rise to alternative sources of asymmetry that our framework could incorrectly attribute to inattention. We note that our model is over-identified, providing the potential to test the validity of the consideration set

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

model outlined in this paper, and that the asymmetries predicted by our framework have a particular structure.[19] With $J > 2$, the derivative of the log of consideration probabilities are over-identified.[20]

Let the system of equations defined by Equation 3.22 be expressed as:

$$c(\mathbf{p}) = D(\mathbf{p})\phi(\mathbf{p}) \tag{3.25}$$

where $c(\mathbf{p})$ is $\frac{1}{2}J(J+1)$-vector of cross derivative differences at prices $\mathbf{p}$, $\phi(\mathbf{p})$ is the J-vector of log consideration probability derivatives, and $D(\mathbf{p})$ is the coefficient matrix of market share differences.[21]

There are

$$\underbrace{\frac{1}{2}J(J+1)}_{\text{\# Independent Cross Deriv. Diffs}} - \overbrace{(J+1)}^{\text{\# } \phi_{jm} \text{ Derivatives}} \tag{3.26}$$

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

Different models of consideration set formation imply different patterns of cross-derivative asymmetry. These patterns can be tested and used to inform model development. For example, in the DSC model, cross-price derivatives are only asymmetric between the default and non-default products. In the ASC model, however, much richer patterns of cross-derivative asymmetry arise. We suggest practical suggestions for testing this in Section 4.

Furthermore, there are particular patterns in cross-derivative asymmetries that are characteristic of lack of consideration as opposed to other well documented behavioral phenomena. For example, one robustly documented pattern is that consumers respond more to larger proportional changes in prices. This is sometimes referred to as the "Weber-Fechner law of psychophysics" (Thaler 1980). This effect could be captured by allowing indirect utility to be a nonlinear function of price; cross-derivatives will be symmetric conditional on two goods having the same price level. Additionally, this model predicts a different pattern of asymmetries than our consideration set model. In our consideration set model, asymmetries scale with latent utilities. In the Weber-Fechner model, cross-price asymmetries scale with the difference in prices between two goods. In Section 6, we show in one application that cross-derivatives scale in the manner predicted by our consideration set model.

More generally, asymmetries might also arise from different forms of inattention from those modeled here. We have assumed that attention occurs at the level of goods. An alternative possibility,

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

[20] With $N > 2^J$, latent market shares are over-identified.

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

developed in Gabaix (2014), is that inattention occurs at the level of *characteristics*.[22] While a comprehensive treatment of inattention to characteristics is beyond the scope of this paper, we here show that the patterns of asymmetries implied by Gabaix (2014) are distinguishable from those in our model of good-specific attention. Adapting Gabaix (2014) to a discrete choice setting gives an indirect utility function of the form:

$$u_{ij} = \beta(p_d + \theta_j^p(p_j - p_d)) + \epsilon_{ij} \tag{3.27}$$

where θ_j^p represents the attention paid to the price (p) of j relative to the price of a default good, d.

Discussing consumer choice, Gabaix (2014) treats θ_j^p as structural parameters that are fixed independently of the realized characteristics of each good. In this model, the ratio of cross-derivatives is constant but not generally equal to 1 (implying asymmetric cross-derivatives):[23]

$$\frac{\frac{\partial s_{jm}^x}{\partial p_{j'm}}}{\frac{\partial s_{j'm}^x}{\partial p_{jm}}} = \frac{\theta_{j'}^p}{\theta_j^p} \quad \neq 1 \quad \text{when } \theta_{j'} \neq \theta_j \tag{3.28}$$

where the x superscript denotes that we are considering cross-derivatives with inattention to characteristics rather than goods. In our consideration set model, however, the ratio of cross-derivatives is not constant but instead scales with market shares. Thus, while alternative behavioral stories can generate asymmetries, they are typically distinguishable from the models of good-specific consideration we consider here.

4 Estimation

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

However, in practice, the curse of dimensionality renders this approach infeasible in most applied settings of interest.[24] Furthermore, consideration probabilities and $s_j^\star(\cdot)$ are complicated nonlin-

[22]See also Gennaioli and Shleifer (2010) and Bordalo, Gennaioli, and Shleifer (2017) for further work on salience effects distorting the relative weight placed on the characteristics of goods.

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

[24]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

ear functions of observables, so measurement error will make estimation challenging even for low-dimensional problems. Thus, as with most practically-sized discrete choice models, parametric assumptions will be necessary for estimation. Throughout this section and our empirical applications, we will therefore assume that our consideration set model is characterized by the finite-dimensional parameter vector $\beta \in \mathbb{R}^k$ as follows:

$$u_{ij} = v(x_j; \beta) + \epsilon_{ij}(\beta) \tag{4.1}$$

$$\mu = Pr\left(h(x_0; \beta) + \eta_{i0}(\beta) > 0\right) \tag{4.2}$$

$$\phi_j = Pr\left(g_j(x_j; \beta) + \eta_{ij}(\beta) > 0\right) \quad \text{for } j > 0 \tag{4.3}$$

and $\phi_0 = 1$.

Existing applications of consideration set models are typically estimated by maximum likelihood (Goeree 2008).[25] In our setting, the principle downside of this approach is the lack of transparency regarding what variation is driving our ultimate results. Are the estimated consideration probabilities driven by the asymmetries in the choice probabilities or by parametric assumptions made in specifying the model?

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

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

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

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.

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

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

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

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

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

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

where W is a positive definite weighting matrix. Note that the set of structural errors used to generate the simulated data sets are held fixed for different values of β. As the sample size grows large, $\hat{\theta}$ and $\widetilde{\theta}(\beta)$ both converge to the same "pseudo true" value, θ_0, underlying the consistency of the approach (Gourieroux, Monfort, and Renault 1993).[26].

4.1 Auxiliary Model Specification

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

Our identification proof points to the importance of specifying an auxiliary model that permits

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

asymmetric cross elasticities. A functional form we find works well in practice (in simulations and our application below) is to specify a flexible logit model, where we begin with a conventional logit model (which imposes symmetry) and then add additional interaction terms between the characteristics of alternative goods to allow for asymmetries. That is:

$$\widetilde{u}_{ij} = \theta_j + x_{ij}\theta_j^0 + \sum_{j'}\sum_{k'}\sum_{k} \theta_{j,j'}^{k,k'} x_{ij}^k x_{ij'm}^{k'} + e_{ij} \tag{4.7}$$

where k is the index denoting the attribute of a product.[27] We then specify $\mathcal{L}(y; x; \theta)$ as the likelihood function for the conditional logit with the utility function $\widetilde{u}_{ij}$.

This specification has several desirable properties. We show in Appendix B that the consideration set model with logit utility can be rewritten as a full-consideration model where the utility of each alternative j depends directly on the attributes of rival goods.[28] 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).[29] 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. In the related literature on choice-based sampling, the econometrician sees only a subset of goods from which consumers choose. In these models, one can sometimes consistently estimate preferences by controlling for alternative-specific constant (Manski and Lerman 1977; Bierlaire, Bolduc, and McFadden 2008). In our framework, this approach does not work since such constants cannot capture the direct dependence of utility on (variable) attributes of rival goods.

[27] When estimating a DSC model, we find that it suffices to use:

$$\widetilde{u}_{ij} = \theta_j + x_{ij}\theta_j^0 + \sum_{k}\sum_{k'} \theta_j^k x_{ij}^k x_{i0}^{k'} + e_{ij}$$

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

[29] 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.

5 Application 1 : Experimental Validation

In this section, we investigate whether we can recover consideration probabilities from real choices with varying prices. We do so in a lab experiment in which consumers make choices from known subsets of 10 goods. We ask whether we can recover the known consideration probabilities as well as preferences conditional on consideration using information only on observed choices. Using real choices rather than simulations tests whether we can use our identification result to recover consideration probabilities in real-world settings where our functional form assumptions on preferences are not guaranteed to hold.

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

Set-up We selected 10 goods sold at the Yale Bookstore with list prices ranging from \$19.98-\$24.98. These goods and their list prices are shown in Table 9 in Appendix C. Each participant was endowed with \$25 and made 50 choices from randomly chosen subsets of the 10 goods with randomized prices (one third of the list price plus a uniformly distributed amount between \$0 and \$16). After making all 50 choices, one of these choices was randomly selected and they were given that item as well as \$25 minus the price of the item in cash.[30] In total, we ran the experiment with 149 participants, resulting in 7,450 choices.[31] Prior to the experiment, participants were given several examples to illustrate the incentive scheme and were quizzed on their understanding. 70% correctly answered our test of understanding (and all participants were told why their answer was correct or incorrect). Appendix Table 10 reports results using only this subset of users who passed this test and shows that results are qualitatively unchanged.

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

[30] 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.

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

probability varied across goods and with prices such that goods were more likely to be considered if they had a higher price (perhaps mimicing the behavior of a retailer who places their highest margin products where they are most likely to be noticed). The probability that good j was in a participant i's round r consideration set was specified as:[32]

$$\phi_{ijr} = Pr\left(\rho_j + p_{ijr}\gamma - \eta_{ijr} > 0\right) \tag{5.1}$$
$$= \frac{\exp(\rho_j + p_{ijr}\gamma)}{1 + \exp(\rho_j + p_{ijr}\gamma)} \tag{5.2}$$

where η_{ijr} is distributed logistic, p_{ijr} give's the product's price, and ρ_j is a product-specific fixed effect. The coefficients were chosen so that most choice sets would include between 2 and 7 products. See Table 1 for the precise coefficients. Note that in the experiment we do not specify an inside default good as in the DSC model.[33]

To increase the likelihood that participants considered all of the products that they were presented with, we required them to spend at least 10 seconds looking at the screen before finalizing their choices. This allows us to take choices from the generated choice set as representative of consumers' true preferences. A sample product selection screen is shown in Figure 1. Consumers were shown images of all the products in their consideration set along with the (randomly chosen) prices. They click the radio button for the product they want, and can click "Next" after 10 seconds.

Figure 1 Here

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

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

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

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

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

[33] However, closing the model requires us to specify a good that is chosen if the consideration set is empty. We specify this as good 10. At the estimated parameter values, an empty consideration set has a 0.2% chance of occurring so the choice of default does not impact estimation.

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

Results Table 1 compares the estimated parameters from a conditional logit model (estimated by maximum likelihood as if all 10 goods are considered) and from our consideration set model estimated by a) maximum likelihood and b) indirect inference. We also report preference parameters estimated by maximum likelihood using information on the actual choice sets that consumers faced. In contrast, our consideration set model parameters are estimated using only information about the product consumers actually chose and not information about the specific subset of 10 goods they could choose from in each instance. Like in any real-world setting, the underlying utility model could in reality be a nested logit, a random coefficients model, a multinomial probit or something more exotic – but our experiment shows that we can nonetheless recover the (known) underlying attention probabilities.

Table 1 here

First, consider the maximum likelihood preference parameters shown in the top panel of Table 1 (we consider the implied elasticities below). The conditional logit model estimated based on a considered choice set of all 10 goods gives a price effect of -0.05, less than a third of the value recovered from a logit model given actual choice sets. This is because the conditional logit model wrongly infers from the fact that high priced products are more likely to be considered (and thus chosen) that consumers do not really dislike high prices. The consideration set model gives a value of -0.20 (0.03) – the confidence interval includes the 'true' value of -0.17. The conditional logit fixed effects are systematically biased because they conflate attention and utility. Products that are rarely in the choice set are assumed to be low utility. In contrast, the consideration set model recovers fixed effects consistent with those estimated using all information on what products were considered. The intervals are relatively wide, but that is a feature, not a bug relative to the conditional logit model: the consideration set model correctly recognizes that rare products are rare and that only limited information is available about how much consumers value them. The consideration set model confidence intervals on the less rare products (products 6-9 in Table 1) are reasonably precise.

The second panel of Table 1 reports the coefficients which give rise to the consideration probabilities and Figure 2 shows this information graphically. Across products, the confidence intervals on the fixed effects in the consideration equation include the true values with the exception of product 1, which lies close to the boundary of the confidence interval. We also correctly recover the impact of price on consideration. Price has a (known) coefficient of 0.15 – by construction, consumers are more likely to see a product if the price is higher, as might arise in the real world if sellers advertise their premium products. In the consideration set model, we estimate 0.137 (.017).

Figure 2 here

Table 1 also reports structural parameters estimated by indirect inference using a flexible auxiliary model that permits a broad class of asymmetric demand responses.[34] These estimates are slightly less precise than the maximum likelihood estimates, but otherwise share the qualitative features noted above. This approach allows us to test directly whether the ASC model can account for the patterns of asymmetries we see in the data.

Elasticities An additional question of interest is whether the implied price elasticities of consideration set models differ from full consideration approaches. We can compute the price elasticities given preferences estimated assuming consideration sets are known and compare them to the price elasticities implied by a variety of models. We consider a few alternatives: our consideration set model (the ASC model), a random coefficients model which allows each individual to have a separate price coefficient,[35] and standard conditional logit models with quadratic and alternative-specific price parameters. We compare these to the "Full Information" elasticities, where preferences are estimated using a logit model with known consideration sets and elasticities are computed given the known function relating consideration to prices (the elasticity reported is still the reduced form elasticity – how demand changes as prices change, combining the impact of prices on consideration and the impact of prices on preferences).

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

[34] See Section 4 and Appendix 4 for details.

[35] "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.

Figure 3 shows the average own-price elasticities by good in each model. For goods 1-4, true own-price elasticities are positive because a higher price makes a good more likely to be considered. As noted above, this is an intentional feature of the model designed to mimic the fact that in some real world settings, consumers might be more likely to see higher priced items. Conditional on consideration, Table 1 shows that price responses are negative as expected; the reduced form elasticities in Figure 3 are positive in some cases if the increase in consideration probability at higher prices outweighs the decrease in utility.

The logit, random coefficients and quadratic models all badly fail to characterize how elasticities vary across goods. With a separate price coefficient for each good, the product-specific model is able to capture these patterns as is the ASC model. But the product-specific model still performs badly in capturing cross-elasticities. The average magnitude of the 90 full information cross-elasticities in the data is 0.090. The logit model has an average absolute deviation of 0.083, the random coefficients logit model has an average absolute deviation of 0.168, the quadratic model has an average deviation of 0.068, the product-specific model has an average deviation of 0.080, and the ASC model has an average deviation of 0.027, less than half of any of the alternative models. As a function of the full-information elasticities, the bias is on average 45.5 percentage points smaller in the ASC model than in any other model.

Figure 3 here

6 Application 2: Smart Defaults Medicare Part D

In this section, we use our model to evaluate whether the observed inertia in Medicare Part D plans is due to inattention, utility-relevant switching costs or both. This allows us to evaluate a recent "smart default" policy proposal in which consumers are automatically defaulted into the lowest cost plan available. Such a policy has not been implemented and thus we must rely on existing variation in the data and structural methods both to predict how consumers would respond and to normatively evaluate the results.

This application also serves as a roadmap for empirical researchers seeking to apply our results. Specifically, we highlight four steps:

1. *Choose a consideration set model*: the DSC model, the ASC model, or a hybrid specification combining features of both. Following the literature on insurance choice, we will here focus on the DSC model.

2. *Estimate reduced form cross-derivatives*: this tests whether the identifying variation is present in the reduced form data. In this application, we will consider the differential sensitivity of switching decisions to changes in the characteristic of a default good versus rival goods.

3. *Estimate the full model*: one can either estimate the model by maximum likelihood (checking consistency with the reduced form results) or indirect inference (identified directly from the reduced form results). As discussed in Section 3, maximum likelihood is likely to be appropriate in most circumstances unless one is particularly concerned about showing a direct connection between cross-derivative asymmetries and the structural parameters of interest.

4. *Conduct overidentification tests*: In this application, we will consider whether our results adequately capture average cross-derivative asymmetries and how these asymmetries scale with market shares.

6.1 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. Beneficiaries in the median choice set have 48 plans that they can choose among, including both plans which provide only prescription drug coverage and plans which provide broader medical insurance ("Medicare Advantage"). Our analysis focuses on stand-alone prescription drug insurance plans (PDP plans).

The main plan attributes that we observe are premiums, deductibles, indicators for whether plans provide coverage in the Part D "donut hole",[36] the number of the top 100 drugs included in the formulary, and a quality rating based on customer service, member complaints, and "member experience" with the drug plan. Additionally, we construct a "calculator" which determines for a given set of beneficiary drug claims what the beneficiary would pay out of pocket for those claims in each alternative plan in their choice set (Abaluck and Gruber 2016).[37] The average beneficiary in our sample pays \$869 in annual out of pocket costs and \$451 in plan premiums each year.

Table 2 here

[36]The donut hole is a gap in coverage included in Part D in order 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.

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

Beneficiaries We consider two main classes of Medicare Part D beneficiaries. The first is a random sample of 100,000 beneficiaries that are not eligible for low income subsidies (LIS). We also use variation generated by the random reassignment of LIS beneficiaries into new plans. LIS beneficiaries, who constitute 35 percent of all Medicare Part D enrollees, are consumers with income below some multiple of the federal poverty level, depending on family size. They are more likely to be non-white and male than beneficiaries who do not qualify for the program (Table 2). 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.[38] For our analysis, we will consider the plan choices of both the 100,000 random sample of non-LIS Medicare beneficiaries, and the 2,852 individuals whose eligibility for LIS varies during the sample period.[39]

6.2 Smart Defaults & Inertia in PDP Plan Choice

Switching between plans is rare; 90% of beneficiaries in PDP plans choose to remain enrolled in the same plan as last year (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). Motivated partly by these patterns, Handel and Kolstad (2018) propose a 'smart default' policy in which an individual is defaulted into the lowest cost plan available in each year, provided that their monetary gain from such a switch exceeds some threshold. 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.

Ex-ante, the welfare benefits of smart defaults are ambiguous. Would this policy make individuals better off by assigning them to lower cost plans, or do individuals have utility-relevant switching

[38] In Appendix E.2, we also consider a sample of 6,971 beneficiaries which includes those under age 65 who qualify for Medicare due to kidney disease or disability payments.

[39] As these samples are demographically very different, in Appendix D we repeat our analysis using a sample of 100,000 non-LIS eligible beneficiaries who demographically resemble the reassigned sample and find no qualitative differences.

costs that might outweigh the benefits of reassignment? Evaluating the welfare impact of smart defaults requires an understanding of why consumers are inertial. If switching rates are low because of utility-relevant switching costs, then those costs must be weighed against any potential benefits of the policy. Yet if inertia reflects inattention alone, then individuals may be made better off enrolling in alternative plans. The DSC model allows us to separate inattention from utility-relevant explanations using asymmetries in demand responses.

To evaluate the policy of interest it is important to distinguish between two sources of utility-relevant switching costs. First, there may be non-trival "paperwork costs" to switch plans and enroll in a new scheme (Luco 2019). Second, consumers may face "acclimation costs" when navigating new products, such as scheduling deliveries for mail-order drugs or learning which of several chemically equivalent drugs are covered by any given plan. 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.[40] Heiss, McFadden, Winter, Wupperman, and Zhou (2016) attempt to distinguish between inattention and utility-relevant switching costs in Part D (finding that most inertia is driven by utility-relevant switching costs) but do not distinguish between paperwork and acclimation costs and thus cannot evaluate "smart defaults". Additionally, they impose several exclusion restrictions to identify inattention (such as that certain demographics impact cognitive ability but not preferences and that consideration is independent of the price level of the default); we test these restrictions below.

All of the mechanisms that we have discussed so far are forms of what Heckman (1981) calls structural state dependence; past purchases *directly* impact choice probabilities. This is distinguished from spurious state dependence, wherein inertia arises due to unobserved serially correlated factors impacting purchasing decisions (e.g. a consumer is willing to pay a premium for certain brands). 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.[41]

[40]Dubé, Hitsch, and Rossi (2010) propose an alternative decomposition of structural state dependence into loyalty, learning and search effects. In Appendix D.1, we discuss this decomposition and how it relates to our work here. To summarize: their proposed method of identifying "search costs" explicitly rules out the inattention channel which we find is responsible for the majority of inertia. Additionally, rather than considering "learning" as a separate possible determinant of inertia, we see it as a factor that can impact inertia via all of the usual channels. Likewise, (Luco 2019) decomposes inertia into what we call paperwork costs and "the cost of evaluating financial information". The latter is analogous to what we call "rational inattention"; we allow for the latter without assuming that inattention is rational in the sense of being responsive to the potential cost savings from being awake.

[41]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 have large acclimation costs (a form of structural state dependence), they may return to their original plan because choosing

To separate these, we 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).[42] 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.

A decomposition of the potential reasons for inertia is summarized in Figure 4. To summarize, we will focus on disentangling different components of structural state dependence to evaluate a smart default policy. We differ from the prior literature in that we will separately identify both paperwork and acclimation costs, and will allow for both inattention to be both 'naive', i.e. independent of the potential benefits of searching, and 'rational', i.e. proportional to the expected benefits via the value of the default plan.

Figure 4 Here

6.3 Choice Model & Plan Reassignment

Before presenting reduced form evidence on inattention and switching costs, we formally define each of the above determinants of inertia in the context of the DSC model in which consumers are either inattentive and choose the default good or, if the default good becomes sufficiently unsuitable, they "wake up" and make an active choice from the full feasible choice set. Following the existing literature, we assume that consumers compare plans only in the current year when making a choice.[43]

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.

[42]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".

[43]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,

As discussed in detail in Section 3, under the DSC model the probability of selecting option j is expressed as:

$$s_{ijt} \equiv s_{jt}(\mathbf{x}_{it}) = (1 - \mu_t(\mathbf{x}_{idt}))\, Default_{ijt} + \mu_t(\mathbf{x}_{idt}) s^{\star}_{jt}(\mathbf{x}_{it}) \tag{6.1}$$

where d is used to denote the current default, $Default_{ijt} = 1$ if good-j is individual i's default at time t and is zero otherwise, and $s^{\star}_{jt}(\cdot)$ denotes the probability of choosing j conditional on being awake at t. The identification argument of Section 3 demonstrates how consideration probabilities can be identified from cross-derivative asymmetries given independent price variation.

Conditional on being awake, the utility of individual i from choosing plan j at time t is given by:

$$u_{ijt} \quad = \quad \mathbf{x}_{ijt}\beta + (\alpha + \rho) Default_{ijt} + \epsilon_{ijt} \tag{6.2}$$

where $\mathbf{x}_{ijt}\beta$ gives the utility of plan characteristics $\mathbf{x}_{ijt}$,[44] and the remaining terms parametrize inertia: 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. Following the earlier literature, we allow consumers to err 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),[45] and we allow financial plan characteristics to matter for 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). Thus, our model allows for inattention both at the level of goods and at the level of attributes.

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. When these beneficiaries no longer qualify for full premium subsidies, utility is given by:

$$u_{ijt} \quad = \quad x_{ijt}\beta + (\alpha + \rho) Default_{ijt} + \alpha\, (Default_{ij,t-1} \times Reassigned_{ijt}) + \epsilon_{ijt} \tag{6.3}$$

and Town (2015) and Abaluck, Gruber, and Swanson (2018) both estimate that Medicare Part D consumers are highly myopic in their drug choices).

[44] Precisely, $\mathbf{x}_{ijt}\beta = \pi_{jt}\beta_0 + \mu^*_{ijt}\beta_1 + \sigma^2_{ij}\beta_2 + \delta q_{jt} + \mathbf{x}_{jt}\gamma$, where π_{jt} denotes premiums, μ^*_{ijt} denotes expected out of pocket costs, σ^2_{ij} denotes the variance of out of pocket costs, q_{jt} denotes plan quality ratings and x_{jt} includes other plan covarates such deductibles, # of top 100 drugs on the formulary, average cost-sharing (a percentage, averaged across all beneficiaries), as well as plan quality ratings.

[45] 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.

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.[46]

Welfare Changes In order to evaluate optimal defaults, we must take a stand on normative utility. Following Abaluck and Gruber (2011) and Heiss, McFadden, and Winter (2007) and supported by our finding of no spurious state dependence, we assume that–apart from switching costs–normative utility depends only on total cost and other observable factors. We denote this utility by v_{ij} (we suppress the subscript t, although plan attributes vary over time as well).[47] In other words, normative utility is given by 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).

Consider the smart default policy under which individuals are reassigned to an alternative plan and given the option of immediately switching back if they desire without enrolling in the new plan for a year. Denote the new default by d, the old default by o. In the "old" status quo regime without the smart default policy, regime 0, the expected welfare of consumer i is given by:

$$W_i^0 = \mu_i^0 \left[s_{id}^{\star 0} \left(v_{id} - \alpha - \rho\right) + s_{io}^{\star 0} v_{io} + \sum_{j \neq d,o} s_{ij}^{\star 0} \left(v_{ij} - \alpha - \rho\right) \right] + (1 - \mu_i^0) v_{io} \tag{6.4}$$

$$= s_{id}^0 \left(v_{id} - \alpha - \rho\right) + s_{io}^0 v_{io} + \sum_{j \neq d,o} s_{ij}^0 \left(v_{ij} - \alpha - \rho\right) \tag{6.5}$$

Latent choice probabilities and inattention probabilities – and thus market shares – will be all be affected by the change in switching costs and default option bought about by the smart default policy. There is also the potential for a direct impact of the policy on consideration, which we discuss below. In the regime with the active default policy, expected welfare can be expressed as:

$$W_i^1 = \mu_i^1 \left[s_{id}^{\star 1} \left(v_{id} - \alpha\right) + s_{io}^{\star 1} (v_{io} - \rho) + \sum_{j \neq d,o} s_{ij}^{\star 1} \left(v_{ij} - \alpha - \rho\right) \right] + (1 - \mu_i^1)(v_{id} - \alpha) \tag{6.6}$$

$$= s_{id}^1 \left(v_{id} - \alpha\right) + s_{io}^1 (v_{io} - \rho) + \sum_{j \neq d,o} s_{ij}^1 \left(v_{ij} - \alpha - \rho\right) \tag{6.7}$$

[46] 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 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.

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

The change in welfare can therefore be expressed as:

$$\begin{aligned}\triangle W_i &= W_i^1 - W_i^0 \\ &= \rho\left(s_{id}^1 - s_{io}^0\right) + \alpha \triangle s_{io} + \sum_j \triangle s_{ij} v_{ij} \end{aligned} \tag{6.8}$$

Defaults 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).

Estimating the degree of inattention is required to bring Equation 6.8 to the data for two reasons. First, estimates of ρ and α will depend on the degree of inattention. Second, in order to recover Δs_{ijt}, we need to simulate the impact of "smart defaults"; this in turn will depend on the degree of inattention as well as the structural parameters.

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 a switch to the smart default policy to have a direct effect on attention. It is plausible that defaulting consumers to a different plan might "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. In the DSC model, attentive consumers always put in effort to make an active choice and inattentive consumers do not. Thus, if smart defaults shift the degree of inattention, we might worry (normatively) 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.[48]

[48] 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.

6.4 Results

Reduced form results Before giving our structural estimates, we first provide reduced form evidence of inattention and switching costs consistent with our framework. Inattention and utility-driven switching costs are separately identified following Section 3 using asymmetries in how the decision to remain inertial depends on default vs. rival plan characteristics.[49] 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).

$$y_{it} = x_{idt}\alpha_d + \bar{x}_{ijt}\alpha_x + \delta_i + \delta_t + e_{it} \tag{6.9}$$

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 both unweighted averages of rival plan characteristics and market-share weighted averages (using pre-period weights).[50]

The results are shown in Table 3. In nearly all cases where we detect a significant effect, we see that switching decisions are more sensitive to default characteristics than rival characteristics. A \$100 increase in premiums or deductibles for the default plan increases switching probabilities by 13-14%, while we find no significant impact of rival premiums and deductibles on switching behavior. This evidence of asymmetry in responses to the changes in characteristics of the default versus non-default products is consistent with the findings of Ho, Hogan, and Scott Morton (2015) 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 many consumers do not actively search each period but can be induced to make an active choice if the default plan becomes bad enough – in this case, we will see greater responsiveness to attributes of the default plan which impact choices both via utility and via prompting consumers actively to consider other available plans. These tests imply a high degree of inattention.

Table 3 here

Next, we consider variation which separately identifies paperwork costs and acclimation costs.

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

[50] This (panel) regression is only possible in our non-reassigned beneficiaries who are exposed to the financial consequences of plan choices for multiple years. We report this regression here using non-reassigned beneficiaries, and later estimate the DSC model using a pooled sample which includes both reassigned and non-reassigned beneficiaries but allows their preferences to differ.

Both of these are utility-relevant sources of inertia and thus identified by the degree to which the observed amount of inertia is larger than we would predict given the probability of attention. To distinguish paperwork from acclimation costs, we consider the choices of beneficiaries who are randomly reassigned due to LIS eligibility but subsequently lose their eligibility (and thus face non-zero premiums). If consumers who are randomized away disproportionately prefer to return to their original plan rather than alternative plans (given observables), this identifies acclimation costs.[51] If consumers disproportionately prefer to stay enrolled in the plan which they were randomized to, this identifies paperwork costs. In practice, we observe that only 0.7% of beneficiaries return to their original plan after being reassigned. However, given the extremely high degree of inattention and the large number of possible options, this is still more than we would expect by chance and implies non-trivial acclimation costs. Our structural results below imply that 20% of beneficiaries are attentive. Given 53 plans in the typical choice set, this implies that an average plan would have enrollment of 0.4%. Enrollment in the original plan is thus about twice as high as we would otherwise predict, which suggests non-trivial acclimation costs. We use a choice model to quantify these costs and assess their policy implications.

Structural Results 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.[52]

We start by estimating conditional logit models which do not allow for inattention (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. In terms of dollars of premiums, ignoring inattention, the fact that consumers are overwhelmingly likely to choose the default plan implies paperwork costs of $1,078, and the fact that they are somewhat more likely to return to their original plan after being reassigned than we would predict given observables implies acclimation costs of $196.

Next, we estimate the DSC model in the same sample. Consider first the attention probability

[51] Assuming as we show above that spurious state dependence plays a limited role.

[52] In Appendix D, Table 4 we show that our results are robust to preference differences that might be expected given the differences in demographic characteristics between our LIS and non-LIS beneficiaries. Specifically, we re-estimate our models with non-LIS beneficiaries who demographically resemble the LIS beneficiaries according to a Mahalanobis distance metric. No qualitative differences in results or our estimated welfare benefits of the smart default policy arise.

estimates. The impact of plan attributes on attention is intuitive: consumers are more likely to pay attention if, for their default plan, premiums or out of pocket costs are higher, if deductibles are larger, if the plan lacks donut hole coverage, or if the plan has less generous cost-sharing or a lower quality rating. In a simple rational inattention model, we might expect the attention coefficients to be proportional to the utility coefficients (as consumers would choose to pay attention based on their assessed value of their current plan). This is not quite the case: attention is more responsive to donut hole coverage, average cost sharing and plan quality relative to premiums than rational inattention would predict. Thus, we statistically reject both a rational attention model and a fully naive model where attention is random, but rational inattention is closer to correct than the naive alternative. The preference coefficients are similar in sign to the conditional logit estimates but slightly larger in magnitude, reflecting the fact that the same behavioral response implies a greater sensitivity to attributes conditional on paying attention (partially offset by the fact that some of the behavioral response captured in the conditional logit coefficients reflects the impact of those attributes on attention rather than utility).

The implied attention probability is about 20%. Reassigned consumers are more likely to choose their original plans everything else held equal than rival plans (identifying acclimation costs) and they are more likely still to choose the plan they were reassigned to, even conditional on paying attention (identifying paperwork costs). Estimated acclimation costs are \$127 and paperwork costs are \$283.

Table 4 here

Overidentification Next, we consider overidentification tests both to evaluate the fit of the DSC model in this setting and to test assumptions used in the previous literature to identify inattention. To do so we apply a strategy inspired by the indirect inference estimation approach that was outlined in Section 4. We estimate a model where the utility of good j is allowed to depend directly on attributes of the default good in a more flexible manner than that permitted by the DSC model, while nesting this framework (see Appendix E for the formal details). We then test whether the cross derivative differences implied by the more flexible model are statistically different from those estimated conditional on assuming the Default Specific Model.

Figure 5 gives the predicted cross derivative difference between default and non-default goods for four plan characteristics. We graph the estimated cross derivative difference yielded by the flexible discrete choice model and by the DSC model against the predicted market share of plan j; the DSC model implies a linear relationship between cross-derivative differences and predicted market shares (Equation 3.14). To capture the uncertainty in the estimated cross-derivatives, we bootstrap estimation of the flexible discrete choice model and graph the resulting 95% confidence interval.

In all graphs, the green dots indicate the empirical cross-derivatives with respect to premiums estimated conditional on the DSC framework – this is exactly the same data in all graphs, and is included for scale (the green dots are absent in the premium graph itself since they would overlap perfectly with the red dots). For each variable, the red dots indicate the predicted cross-derivative difference from the DSC model and the grey region indicates the 95% confidence interval on the cross-derivative difference from the more flexible specification. We can see that the DSC model cross-derivatives match up well with empirical cross-derivatives.

Figure 5 Here

In Appendix E.1, 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 all of 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).

6.5 Simulation Results

Given our parameter estimates, we simulate welfare under the smart default policy conditional on the structural parameters reported in Table 4.[53] 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. Table 5 gives the distribution of cost savings achieved under the policy. By choosing the lowest cost plan, consumers save $383 on average. About 5% of the sample save more than $1,000. Those who switch back to the old default save no money (by definition), and those who instead actively choose another plan save $98 on average relative to their plan before reassignment.

Table 5 here

However, these monetary savings do not take into account switching costs that beneficiaries

[53]To conduct this simulation, we restrict attention to the random sample. We report the results where choices are simulated using the model estimated with the pooled 100K and LIS samples.

are exposed to by the policy and which lower the utility benefits from moving plans. We thus consider the overall welfare impact of the policy, taking into account risk protection, plan quality, the paperwork and acclimation costs, as well as any additional costs of paying attention. We compare our results to those implied by the conditional logit structural parameters which are estimated under the assumption that consumers fully consider all products.

We find that the welfare benefits of the smart default policy are likely to be significantly understated if consumer inattention is ignored. Table 6 gives the average welfare change from the smart policy. The first row gives the average welfare change implied by the conditional logit structural preference parameters, which are estimated assuming perfect consideration. In this model, the cost savings from defaulting consumers to lower cost plans are offset by the fact that some consumers end up paying extremely large paperwork costs to opt out of their new default. In a model in which choices were rational, this would imply that the potential benefits were even larger than these paperwork costs, but recall that our model also allows consumers to err by overweighting the value of plan attributes. In the perfect consideration model, consumers sometimes pay very high paperwork costs because they (wrongly) think their new plan is unsuitable, and this offsets the cost savings from defaulting them into a new plan.

Using the structural preference parameters implied by the DSC model changes one's conclusions. The second row of Table 6 assumes that the act of defaulting a consumer into new plan as per the smart default policy does not have any direct impact on consideration; attention probabilities are as prescribed by our baseline structural estimates. We find that, on average, consumers are made \$177 better off by the policy; approximately half the average monetary saving from the policy are offset by paperwork and acclimation costs. Relative to the model with full attention, the implied paperwork costs are substantially lower (since stickiness is now partly rationalized by inattention), and estimated acclimation costs are lower as well. Each subsequent row of Table 6 considers an alternative assumption about how the smart default policy directly perturbs the probability of paying attention. Each column shows an alternative assumption about the cost that paying attention imposes on beneficiaries. In the first row, where the smart default policy does not change the attention probability, the cost of paying attention is irrelevant. In subsequent rows, a higher cost of paying attention counts against the smart default policy.[54]

Our results suggest that cost savings generally outweigh acclimation costs for reassigned beneficiaries. These results suggest that reassignment would likely make the average consumer better

[54] It is also possible that consumers would become less likely to pay attention once the smart default were in place. This might occur if they take the new default as a normative recommendation or were rationally inattentive given the quality of the default. In this case, the \$177 in Table 6 would understate the benefits of the policy when the costs of paying attention were large.

off unless the policy induces many consumers to pay large attention costs.[55] Consumers report spending an average of 3 hours choosing plans (Kling, Mullainathan, Shafir, Vermeulen, and Wrobel 2012), so attention costs in the \$50-\$100 seem plausible.

Table 6 here

An alternative policy would only reassign those consumers who stand to benefit 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 7. In this case, the average benefits are about \$50 larger since we avoid forcing some consumers with little to gain to nonetheless pay acclimation costs. These results suggest that a smart default policy can improve welfare, but would be better implemented such that consumers are only defaulted if potential cost savings exceed the estimates of acclimation costs.

Table 7 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 well-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). Such a policy is simulated in Table 8, again only reassigning 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 benefits of the policy (in partial equilibrium) are slightly smaller than in Table 6 but still

[55] The results in the matched sample are similar, but somewhat more favorable to the policy as the estimated acclimation costs are smaller.

substantial: consumers save less on premiums and out of pocket costs (partially offset by enrolling in plans with higher quality), but the implied savings are still sufficiently large that consumers are left better off unless the policy forces consumers to pay large attention costs.

Table 8 here

7 Conclusion

Discrete choice models with consideration sets relax the strong assumption that beneficiaries consider all of the options available to them before making a choice. In the 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 unnecessary. We show that a broad class of such models are identified from variation already available in the data. Consideration set probabilities are constructively identified by deviations from Slustky symmetry, i.e. asymmetries in the matrix of cross-derivatives of choice probabilities with respect to characteristics of rival goods.

We illustrate a number of practical applications of the model. In a lab experiment, using only data on observed choices, we recover consideration probabilities and obtain accurate estimates of the elasticities that we would estimate if we observed consideration sets. In data from Medicare Part D, we validate the model further by demonstrating that the cross-derivative asymmetries follow the specific pattern predicted by our model of inattention. We also show that the model can be used to test identifying restrictions used in other work, many of which are rejected. Our model implies that, while most inertia is driven by inattention in Part D, there remain non-trivial adjustment costs. We simulate the welfare effect of a "smart default" policy, finding that defaulting consumers into lower cost plans produces large benefits, which are 4.5 times larger those that implied by structural models which ignore consumer inattention.

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

It is important to note that consideration set models are not necessarily behavioral – search costs or unobserved constraints can lead to a lack of full consideration, and inattention of any kind can be rationalized by sufficiently large search or computational costs (Simon 1971). That being said, consideration set models relax 'full' rationality in the sense that consumers are not necessarily choosing the best option given their utility functions and the choices observable to the econometrician. This allows for a general measure of the quality of consumers' choices, measured as the welfare loss relative to what consumers would choose given full consideration with the estimated utility function.[56] While it is sometimes argued that relaxing full rationality leads to a lack of discipline, we show that the consideration set models we develop are over-identified and their validity can thus be empirically determined.

While we show that deviations from Slutsky symmetry are indicative of imperfect attention in a large class of models, our constructive identification results use the additional structure imposed by the widely applied Default Specific Consideration and Alternative Specific Consideration frameworks. One direction for future work is to characterize more generally when consideration probabilities can be recovered from choice data. One important case are the K-rank models considered in Honka (2014) and Honka, Hortaçsu, and Vitorino (2015) in which consumers consider the K-goods which are highest ranked according to some index, thus violating the independence assumption of the ASC model. 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)). We hope that the sufficient conditions given here will make it possible to adapt consideration set models to a wider range of settings than they have previously been applied.

With additional structure, 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.

[56] We note that this implicitly assumes that the quality of consumer choices is independent of the size of the choice set.

References

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

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

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.

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.

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.

Bordalo, P., N. Gennaioli, and A. Shleifer (2017). Memory, attention, and choice. Technical report, National Bureau of Economic Research.

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.

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 (2017). Inferring cognitive heterogeneity from aggregate choices.

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

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.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics 41*(3), 417–445.

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

Gabaix, X. (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*.

Gennaioli, N. and A. Shleifer (2010). What comes to mind. *The Quarterly journal of economics 125*(4), 1399–1433.

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).

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., 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.

Kling, J. R., S. Mullainathan, E. Shafir, L. C. Vermeulen, and M. V. Wrobel (2012). Comparison friction: Experimental evidence from medicare drug plans. *The Quarterly Journal of Economics 127*(1), 199–235.

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.

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

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The computer journal 7*(4), 308–313.

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.

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

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*(S1).

Swait, J. and M. Ben-Akiva (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological 21*(2), 91–102.

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

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization 1*(1), 39–60.

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

Van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research 47*(1), 63–74.

8 Tables in Main Text

Table 1: Experimental Data Estimation Results

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

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

Table 2: Sample Demographics and Plan Characteristics

Beneficiary Characteristics:	Non-LIS Sample		LIS Sample	
	Mean	Std Dev	Mean	Std Dev
Age	75.6	9.1	75.6	7.4
Female	0.626	0.484	0.546	0.498
White	0.934	0.248	0.640	0.480
Plan Characteristics:	Chosen Plans		All Plans	
	Mean	Std Dev	Mean	Std Dev
Total Part D Costs	$1,320	$808	$1,389	$811
Out of Pocket Costs	$869	$765	$882	$803
Deductible	$64	$117	$111	$133
Donut Hole Coverage	13%	34%	27%	44%

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.

Table 3: Sensitivity of Switching to Default and Rival Plan Characteristics

	Unweighted, Sensitivity to:		Weighted, Sensitivity to:	
	Default	Avg. of Rival Plan	Default	Avg. of Rival Plan
Annual Premium (hundreds)	0.130***	-0.091	0.132***	-0.028
	(0.028)	(0.050)	(0.028)	(0.033)
Annual Out of Pocket Costs (hundreds)	0.016*	-0.015*	0.015*	-0.017*
	(0.007)	(0.007)	(0.007)	(0.007)
Variance of Costs (millions)	-0.145	0.128	-0.135	0.177
	(0.109)	(0.102)	(0.108)	(0.110)
Deductible (hundreds)	0.141**	0.076	0.138**	-0.082
	(0.047)	(0.234)	(0.046)	(0.056)
Donut Hole Coverage	-0.037	0.792	-0.055	0.004
	(0.132)	(0.685)	(0.125)	(0.096)
Average Consumer Cost Sharing %	-0.215	1.760	-0.254	-0.061
	(0.129)	(2.249)	(0.144)	(0.655)
Normalized Quality Rating	0.004	0.062	0.009	0.056
	(0.021)	(0.180)	(0.023)	(0.054)
# of Top 100 Drugs in Formulary	0.011	-0.033	0.014	-0.027
	(0.007)	(0.028)	(0.007)	(0.076)

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, weighted and unweighted average attributes of rival plans, and individual and time fixed effects. The first and third columns report the coefficients on the characteristics of the plan the beneficiaries were enrolled in the year prior (the default plan); the second and fourth columns report the coefficients on the (either unweighted or weighted) average value of the characteristics of rival plans. Weights are determined by estimating the model for individuals who are entering the market for the first time. The regression also includes an indicator for plans that are missing the # of top 100 drugs in formulary variable. Standard errors in parentheses. *** denotes significance at the 1% level, ** at 5% level, and * at 10%.

Table 4: Conditional Logit & DSC Model

	Conditional Logit		DSC Model	
Utility:				
Original Plan	0.988***	(0.238)	1.314***	(0.257)
Assigned Plan	6.428***	(0.012)	4.240***	(0.078)
Annual Premium (hundreds)	-0.505***	(0.005)	-1.034***	(0.010)
Annual Out of Pocket Costs (hundreds)	-0.214***	(0.007)	-0.297***	(0.012)
Variance of Costs (millions)	2.246***	(0.089)	2.579***	(0.165)
Deductible (hundreds)	-0.516***	(0.009)	-0.724***	(0.013)
Donut Hole Coverage	0.691***	(0.027)	0.335***	(0.051)
Average Consumer Cost Sharing %	-1.181***	0.107	-4.128***	0.163
# of Top 100 Drugs in Formulary	0.038***	(0.004)	0.172***	(0.006)
Normalized Quality Rating	0.438***	(0.010)	0.515***	(0.015)
Attention:				
Annual Premium (hundreds)	-	-	0.062***	(0.014)
Annual Out of Pocket Costs (hundreds)	-	-	0.030*	(0.012)
Variance of Costs (millions)	-	-	-0.627***	(0.159)
Deductible (hundreds)	-	-	0.069***	(0.020)
Donut Hole Coverage	-	-	-0.761***	(0.052)
Average Consumer Cost Sharing %	-	-	-1.447***	(0.219)
# of Top 100 Drugs in Formulary	-	-	-0.002	(0.010)
Normalized Quality Rating	-	-	-0.511***	(0.019)
Acclimation Costs	$196		$127	
Paperwork Costs	$1078		$283	
Attention Probability			19.7%	

Notes: The table shows conditional and attentive logit estimates. Estimates in the attentive logit models are the coefficients in the utility and attention equations (not marginal effects). The coefficients in the attention equation are the coefficients on the listed characteristics of the default good. 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: Monetary Savings Relative to Current Plan

	New Default	Old Default	Other Plan
Savings from Plan:			
<$0	0.000	0.000	0.011
$0 - $100	0.081	0.048	0.008
$100 - $200	0.156	0.000	0.008
$200 - $300	0.255	0.000	0.005
$300 - $400	0.193	0.000	0.002
$400 - $500	0.080	0.000	0.001
$500 - $600	0.030	0.000	0.001
$600 - $700	0.028	0.000	0.000
$700 - $800	0.023	0.000	0.000
$800 - $900	0.019	0.000	0.000
$900 - $1000	0.011	0.000	0.000
$1000 +	0.038	0.000	0.001
Mean Savings:	$383	$0	$98

Notes: The table shows the proportion of beneficiaries who choose the lowest cost plan (new default), stay in their current plan (old default) or switch to a third plan (other plan) when choices for a random sample of 100,000 beneficiaries are simulated using the DSC parameter estimates reported in Column 1 of Table 4. For beneficiaries who choose each type of plan, the table shows the distribution of total cost savings relative to remaining enrolled in last year's plan.

Table 6: Welfare Impact of Smart Default Policy

	Attention Cost				
	\$0	\$50	\$100	\$200	\$300
Conditional Logit Parameters	\$31	\$31	\$31	\$31	\$31
DSC Parameters	\$177	\$177	\$177	\$177	\$177
Direct Effect on Attention Probability					
25%	\$172	\$170	\$168	\$164	\$160
50%	\$144	\$129	\$115	\$86	\$57
75%	\$112	\$85	\$58	\$4	-\$50
100%	\$77	\$37	-\$2	-\$81	-\$161

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.

Table 7: Welfare Impact of Smart Default Policy: Only Reassigned if Cost Savings Exceed $125

	Attention Cost				
	$0	$50	$100	$200	$300
DSC Parameters	$222	$222	$222	$222	$222
Direct Effect on Attention Probability					
25%	$215	$213	$210	$204	$199
50%	$184	$168	$153	$122	$91
75%	$150	$122	$95	$39	-$17
100%	$114	$74	$35	-$45	-$124

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. In this case, only beneficiaries for whom the cost savings from the new default plan exceeds the estimated acclimation costs are reassigned.

Table 8: Welfare Simulations - Overall Welfare Impact Restricted Reassignment into High Quality Plans

	Attention Cost				
	\$0	\$50	\$100	\$200	\$300
DSC Parameters	\$184	\$184	\$184	\$184	\$184
Direct Effect on Attention Probability					
25%	\$169	\$164	\$159	\$148	\$138
50%	\$131	\$113	\$96	\$60	\$25
75%	\$93	\$63	\$32	-\$28	-\$88
100%	\$55	\$13	-\$29	-\$112	-\$196

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. In this case, only beneficiaries for whom the cost savings from the new default plan exceeds the estimated acclimation costs are reassigned.

9 Figures in Main Text

Figure 1: Lab Experiment: Sample Product Selection Screen

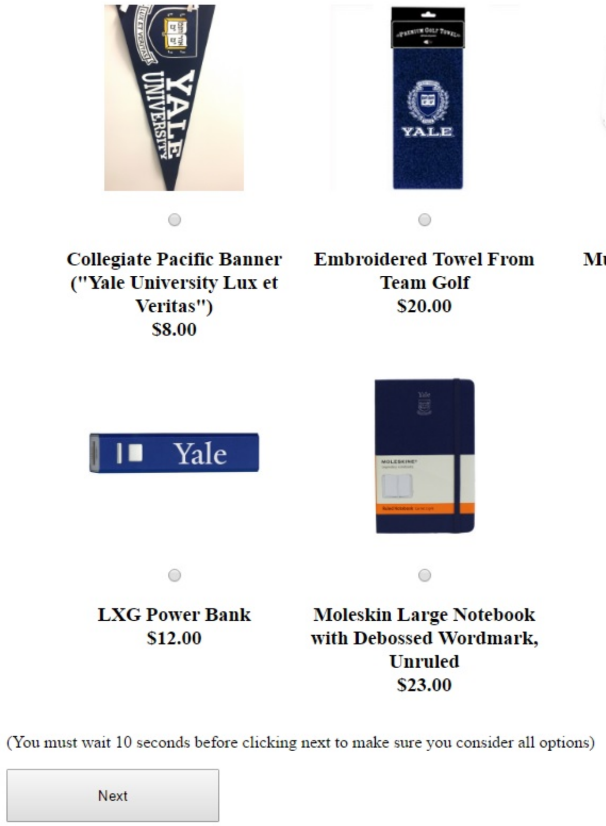


Figure 2: Product Fixed Effects in Attention: Truth vs. ASC Model

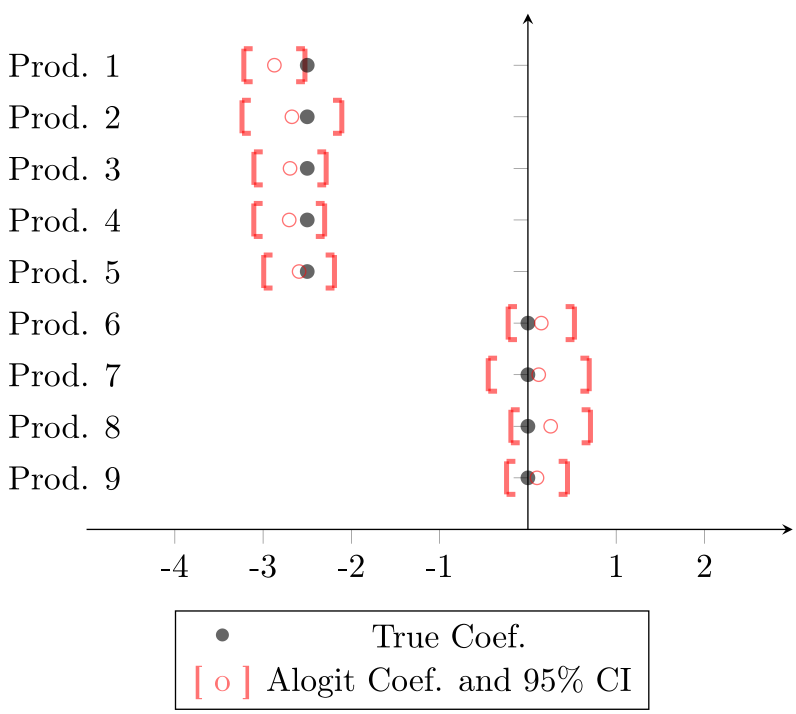
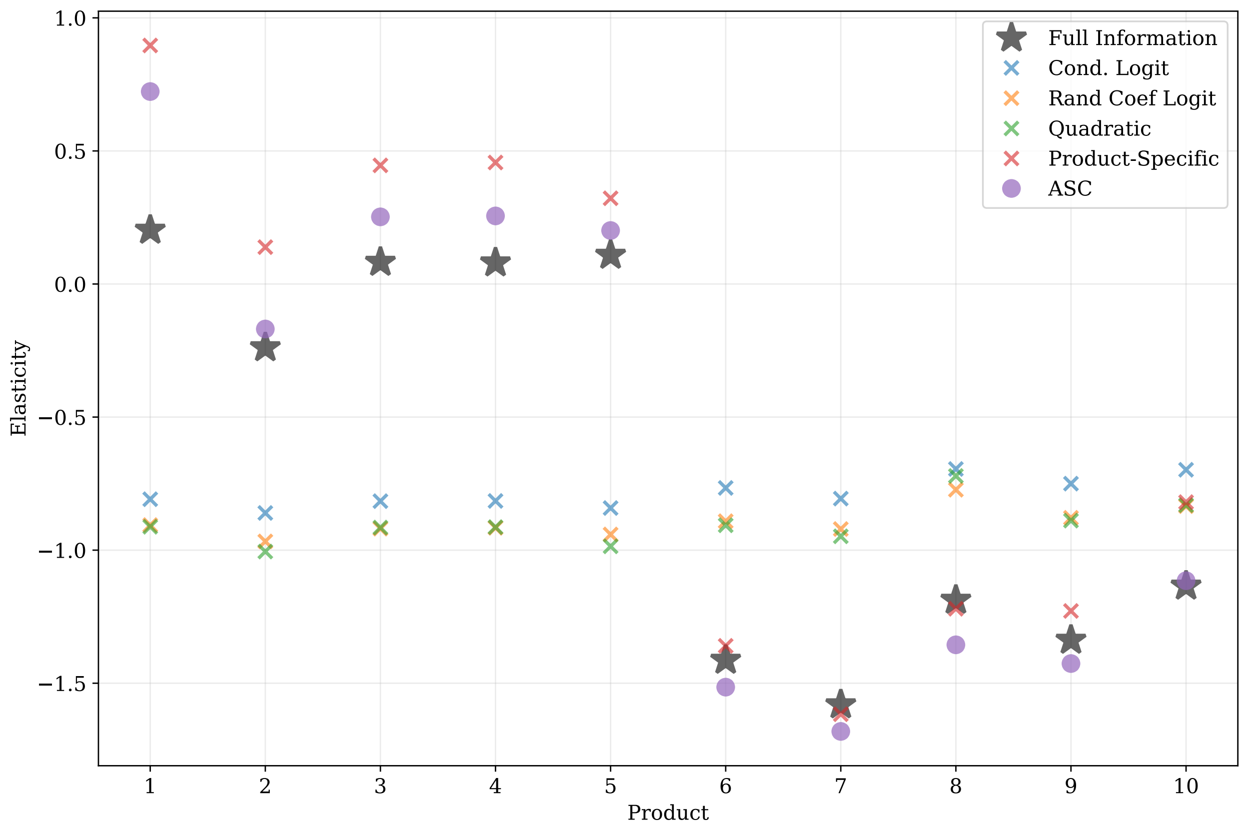


Figure 3: Experimental Data: Own-Price Elasticities by Good



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

Figure 4: Overview of Reasons for Inertia

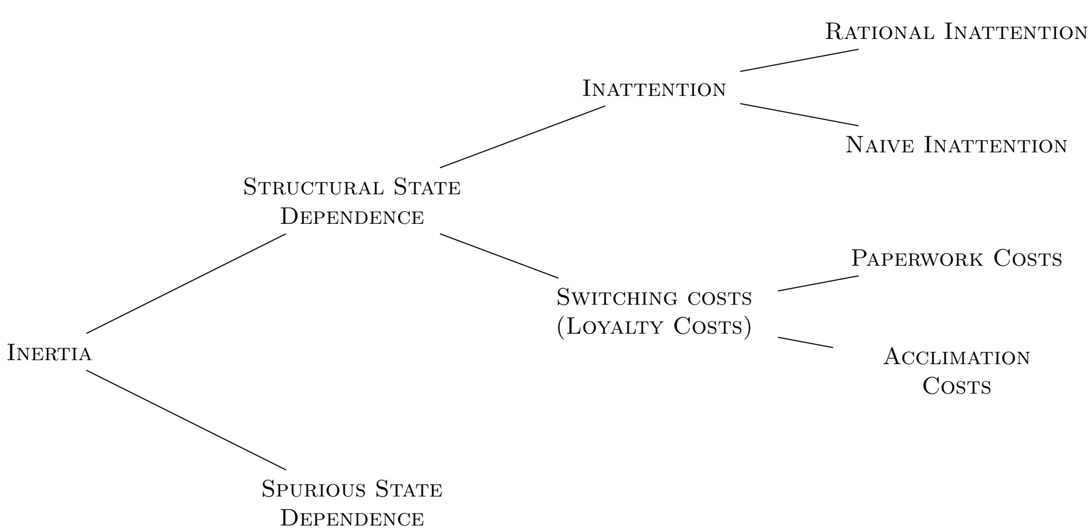
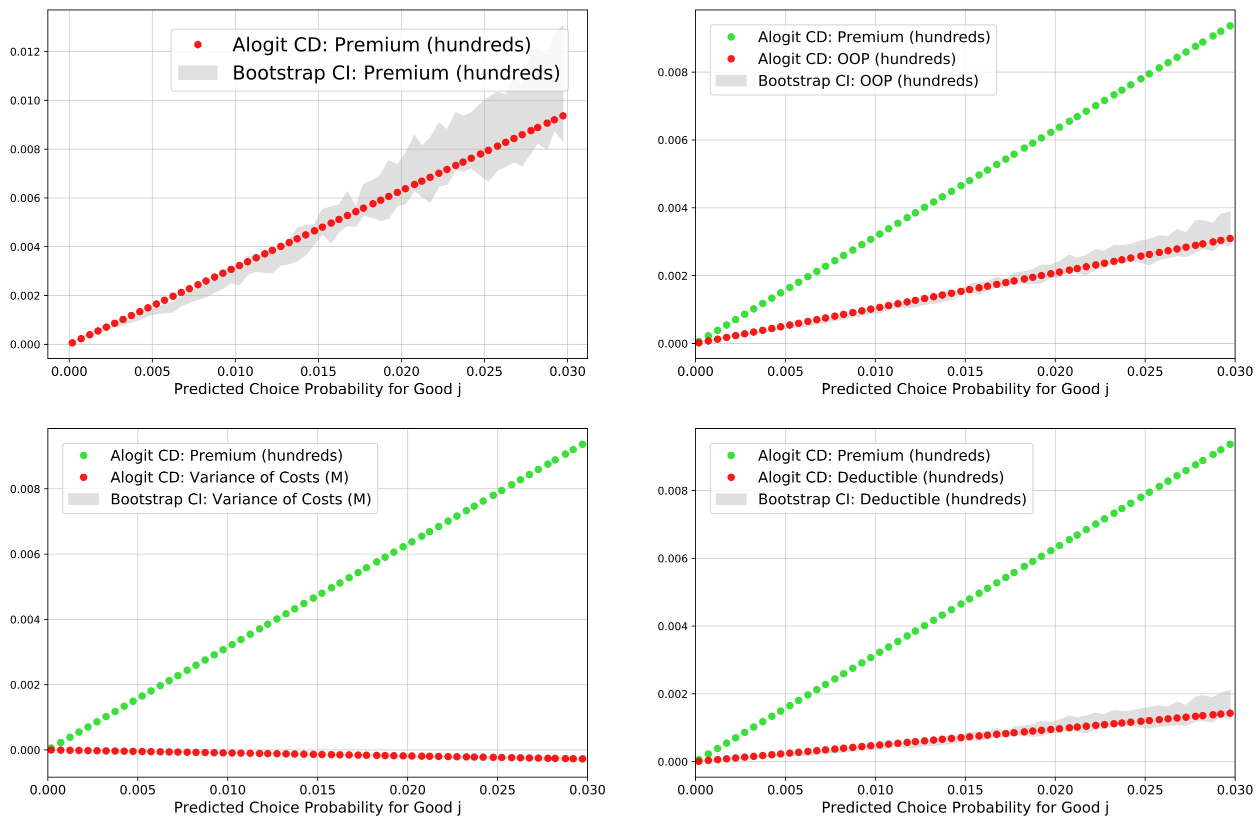


Figure 5: Empirical vs. Model Predicted Cross-derivatives



Notes: Figure shows the estimated cross derivative difference yielded by the flexible discrete choice model (outlined in Appendix E) and by the DSC model against the predicted market share of plan j. In all graphs, the green dots indicate the empirical cross-derivatives with respect to premiums estimated conditional on the DSC framework and is included for scale. For each variable, the red dots indicate the predicted cross-derivative difference from the DSC model and the grey region indicates the 95% confidence interval on the cross-derivative difference from the more flexible specification.

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.

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

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

Thus, non-zero cross-derivative asymmetries imply:

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

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

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

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

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

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 $\pi_{\mathcal{J}} < 1$.

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 $f(\gamma) = 0$.

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

$$\begin{aligned}
s_j(\mathbf{p}) &= \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C) && \text{(A.18)}\\
&= \phi_{j'}(p_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C) + (1 - \phi_{j'}(p_{j'})) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C/j') && \text{(A.19)}\\
&= \phi_{j'}(p_{j'}) \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C)(s_j^\star(\mathbf{p}|C) - s_j^\star(\mathbf{p}|C/j')) + \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C/j') && \text{(A.20)}
\end{aligned}$$

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:

$$\begin{aligned}
s_j(\mathbf{p}|\mathcal{J}/j') &= \sum_{C \in \{\mathbb{P}(j)/\mathbb{P}(j')\}} \prod_{l \in C} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C) && \text{(A.21)}\\
&= \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \prod_{l \in C/j'} \phi_l(p_l) \prod_{l' \notin C} (1 - \phi_{l'}(p_{l'}))\, s_j^\star(\mathbf{p}|C/j') && \text{(A.22)}\\
&= \sum_{C \in \mathbb{P}(j) \cap \mathbb{P}(j')} \widetilde{\pi}(C) s_j^\star(\mathbf{p}|C/j') && \text{(A.23)}
\end{aligned}$$

Given that changes in latent market shares cancel out within a consideration set, cross derivative differences take the form:

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

$$\begin{aligned}
&= \frac{\partial \phi_{j'}}{\partial p_{j'}} \frac{1}{\phi_{j'}} (s_j(\mathbf{p}) - s_j(\mathbf{p}|\mathcal{J}/j')) - \frac{\partial \phi_j}{\partial p_j} \frac{1}{\phi_j} (s_{j'}(\mathbf{p}) - s_{j'}(\mathbf{p}|\mathcal{J}/j)) && \text{(A.25)}\\
&= \frac{\partial \log \phi_{j'}}{\partial p_{j'}} (s_j(\mathbf{p}) - s_j(\mathbf{p}|\mathcal{J}/j')) - \frac{\partial \log \phi_j}{\partial p_j} (s_{j'}(\mathbf{p}) - s_{j'}(\mathbf{p}|\mathcal{J}/j)) && \text{(A.26)}
\end{aligned}$$

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 forms for preferences). 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 $\bar{\mathbf{p}}_{j'}$ be the price vector $\mathbf{p}$ with only $p_{j'} \approx \infty$.

ASC Identification Without Relying Just on Default Cross-Derivative Asymmetries. Cross derivative differences for $j, j' \neq 0$ can be expressed as:

$$\frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} - \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} = \frac{\partial \log(\phi_{j'})}{\partial p_{j'}} \big(s_j(\mathbf{p}) - s_j(\bar{\mathbf{p}}_{j'})\big) - \frac{\partial \log(\phi_j)}{\partial p_j} \big(s_{j'}(\mathbf{p}) - s_{j'}(\bar{\mathbf{p}}_j)\big) \tag{A.30}$$

Let the system of equations defined by Equation A.30 be expressed as:

$$c(\mathbf{p}) = D(\mathbf{p}) \partial \phi(\mathbf{p}) \tag{A.31}$$

where $c(\mathbf{p})$ is a $\frac{1}{2}J(J+1)$-vector of cross derivative differences at prices $\mathbf{p}$, $\partial\phi(\mathbf{p})$ is the J-vector of log consideration probability derivatives, and $D(\mathbf{p})$ is the coefficient matrix of market share differences.[57] As there are typically more than J cross-derivative differences, it is convenient to work with the system:[58]

$$D'(\mathbf{p})c(\mathbf{p}) = D(\mathbf{p})' D(\mathbf{p}) \partial \phi(\mathbf{p}) \tag{A.32}$$

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. We discuss the requirements for this in the

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

[58] Alternative weighting matricies, W_m, can be used: $D'_m W_m D_m$.

context of the Hybrid model to prevent repetition. ASSUMPTION HYBRID.i *Rank Condition*: The matrix $D'(\mathbf{p})D(\mathbf{p})$ is full rank.

A strength of our approach is that the rank condition is testable given market share data. If the rank condition holds, then the derivatives of log consideration probabilities are given as:

$$\phi(\mathbf{p}) = \left(D'(\mathbf{p})D(\mathbf{p})\right)^{-1} D'(\mathbf{p})c(\mathbf{p}) \tag{A.33}$$

For the rank condition to hold, we must have that the number of independent cross-derivative differences is at least as large as the number of derivatives of the log of consideration probabilities:

$$\frac{1}{2}J(J+1) \geq J+1 \tag{A.34}$$

$$J \geq 2 \tag{A.35}$$

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

$$s_j(\mathbf{p}|\mathcal{J}) \neq s_{j'}(\mathbf{p}|\mathcal{J}) \tag{A.36}$$

$$\frac{s_l(\mathbf{p}|\mathcal{J}) - s_l(\mathbf{p}|\mathcal{J}/j)}{s_{j'}(\mathbf{p}|\mathcal{J}) - s_{j'}(\mathbf{p}|\mathcal{J}/j)} \neq \frac{s_l(\mathbf{p}|\mathcal{J}) - s_l(\mathbf{p}|\mathcal{J}/j')}{s_j(\mathbf{p}|\mathcal{J}) - s_j(\mathbf{p}|\mathcal{J}/j')} \tag{A.37}$$

$$s_j(\mathbf{p}|\mathcal{J}) - s_j(\mathbf{p}|\mathcal{J}/j') \neq 0 \tag{A.38}$$

for all $j, j', l \in \mathcal{J}$ with $j, j' > 0$. Equation A.38 will be met when good j' is considered with strictly positive probability and good j' is purchased with strictly positive probability from some choice set that includes j. Equation A.37 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:

$$\begin{bmatrix} -(s_0(\mathcal{J}) - s_0(\mathcal{J}/1)) & 0 & s_1(\mathcal{J}) \\ 0 & -(s_0(\mathcal{J}) - s_0(\mathcal{J}/2)) & s_2(\mathcal{J}) \\ -(s_2(\mathcal{J}) - s_2(\mathcal{J}/1)) & (s_1(\mathcal{J}) - s_1(\mathcal{J}/2)) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \log(\phi_1)}{\partial p_1} \\ \frac{\partial \log(\phi_{2m})}{\partial p_2} \\ \frac{\partial \log(\mu_m)}{\partial p_0} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_1}{\partial p_1} - \frac{\partial s_0}{\partial 0_1} \\ \frac{\partial s_2}{\partial p_0} - \frac{\partial s_0}{\partial 0_2} \\ \frac{\partial s_1}{\partial p_2} - \frac{\partial s_2}{\partial 0_1} \end{bmatrix} \tag{A.39}$$

The determinant of D is:

$$\det(D) = s_2(\mathcal{J})\left(s_0(\mathcal{J}) - s_0(\mathcal{J}/1)\right)\left(s_1(\mathcal{J}) - s_1(\mathcal{J}/2)\right) \\ - s_{1m}(\mathcal{J})\left(s_0(\mathcal{J}) - s_0(\mathcal{J}/2)\right)\left(s_2(\mathcal{J}) - s_2(\mathcal{J}/1)\right) \tag{A.40}$$

When D is singular:

$$\frac{1}{s_1(\mathcal{J})}\left(\frac{s_0(\mathcal{J}) - s_0(\mathcal{J}/1)}{s_2(\mathcal{J}) - s_2(\mathcal{J}/1)}\right) = \frac{1}{s_2(\mathcal{J})}\left(\frac{s_0(\mathcal{J}) - s_0(\mathcal{J}/2)}{s_1(\mathcal{J}) - s_1(\mathcal{J}/2)}\right) \tag{A.41}$$

A.2 Identification of Latent Market Shares

"Nominal illusion" facilitates the identification of the 2^J independent latent choice probabilities in the ASC model, $s_j^\star(\mathbf{p}|C)$.[59] 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 quasi-linearity, these price shifts alter consideration probabilities but do *not* alter latent choice probabilities conditional on consideration.[60] 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 $j > 0$,[61] define the matrices:

$$\Pi_j(\mathbf{p}, \delta) = \begin{bmatrix} \pi_{j1}(\delta_1) & \cdots & \pi_{j\kappa}(\delta_1) \\ \vdots & \ddots & \vdots \\ \pi_{j1}(\delta_N) & \cdots & \pi_{j\kappa}(\delta_N) \end{bmatrix} \tag{A.42}$$

$$s_j^\star(\mathbf{p}) = \left[s_j^\star(\mathbf{p}|C_{j1}), ..., s_j^\star(\mathbf{p}|C_{j\kappa})\right] \tag{A.43}$$

$$s_j^\delta(\mathbf{p}) = \left[s_j(\mathbf{p} + \delta_1), ..., s_j(\mathbf{p} + \delta_N)\right] \tag{A.44}$$

where

$$\pi_{jC}(\delta) = \prod_{l \in C} \phi_l(p_l + \delta) \prod_{l' \notin C} \left(1 - \phi_{l'}(p_{l'} + \delta)\right) \tag{A.45}$$

[59] 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.

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

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

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

$$\Pi_j(\mathbf{p}, \delta) s_j^\star(\mathbf{p}) = s_j^\delta(\mathbf{p}) \tag{A.46}$$

$$s_j^\star(\mathbf{p}) = \Pi_j^{-1}(\mathbf{p}, \delta) s_j^\delta(\mathbf{p}) \tag{A.47}$$

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

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

To avoid too much repetition, we here give the conditions assuming the Hybrid Model, which subsumes the ASC case.

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

$$\frac{\phi_j(p_j + \delta_i)}{1 - \phi_j(p_j + \delta_i)} \neq \frac{\phi_j(p_j + \delta_{i'})}{1 - \phi_j(p_j + \delta_{i'})} \tag{A.48}$$

$$\phi_j(p_j + \delta_i) \neq \phi_{j'}(p_{j'} + \delta_i) \quad \text{at, at least one } i = 1, ..., N \tag{A.49}$$

for $j, j' > 0$. To see the logic of these conditions, consider the just identified case where $J = 2$ and $N = 2$. Suppressing dependence on product characteristics, the coefficient matrix then takes the form:

$$\Pi_1(\delta) = \begin{bmatrix} \mu(\delta_1)\phi_1(\delta_1)(1 - \phi_2(\delta_1)) & \mu(\delta_1)\phi_1(\delta_1)\phi_2(\delta_1) \\ \mu(\delta_2)\phi_1(\delta_2)(1 - \phi_2(\delta_2)) & \mu(\delta_2)\phi_1(\delta_2)\phi_2(\delta_2) \end{bmatrix} \tag{A.50}$$

$$\Pi_2(\delta) = \begin{bmatrix} \mu(\delta_1)\phi_2(\delta_1)(1 - \phi_1(\delta_1)) & \mu(\delta_1)\phi_1(\delta_1)\phi_2(\delta_1) \\ \mu(\delta_2)\phi_2(\delta_2)(1 - \phi_1(\delta_2)) & \mu(\delta_2)\phi_1(\delta_2)\phi_2(\delta_2) \end{bmatrix} \tag{A.51}$$

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

$$\frac{1 - \phi_2(\delta_1)}{\phi_2(\delta_1)} = \frac{1 - \phi_2(\delta_2)}{\phi_2(\delta_2)} \tag{A.52}$$

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

$$\frac{1-\phi_1(\delta_1)}{\phi_1(\delta_1)} = \frac{1-\phi_1(\delta_2)}{\phi_1(\delta_2)} \tag{A.53}$$

When $J > 2$, we 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:

$$s_j = \sum_{c \in \mathbb{P}(j)} \prod_{l \in c} \phi_l(p_0, p_l) \prod_{l' \notin c} (1 - \phi_{l'}(p_0, p_{l'})) \tag{A.54}$$

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:

$$\frac{\partial s_j}{\partial p_0} - \frac{\partial s_0}{\partial p_j} = \frac{\partial \log(\phi_j)}{\partial p_0} s_j(\mathcal{J}) + \sum_{j' \neq \{j,0\}} \frac{\partial \log(\phi_{j'})}{\partial p_0} \left(s_j(\mathcal{J}) - s_j(\mathcal{J}/j')\right) - \frac{\partial \log(\phi_j)}{\partial p_j} \left(s_0(\mathcal{J}) - s_0(\mathcal{J}/j)\right) \tag{A.55}$$

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 6):

$$\frac{1}{2} J(J+1) \geq 2J \tag{A.56}$$

$$J \geq 3 \tag{A.57}$$

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. Other than this restriction, the rest of the proof in Section 3 goes through without modification.

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 which rely on variation in characteristics of the default good. In this case, the order condition for the identification of the derivative of log consideration probabilities changes (Assumption 6). We now require:

$$\frac{1}{2}J(J-1) \geq J \tag{A.58}$$

$$J \geq 3 \tag{A.59}$$

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

A.5 ASC Identification with an Inside Default Good with $\phi_0 < 1$

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

$$s_0 = \prod_{j \in \mathcal{J}} (1-\phi_j) + \sum_{C \in \mathbb{P}(0)} \prod_{l \in C} \phi_l \prod_{l' \notin C} (1-\phi_{l'}) s_0^\star(C) \tag{A.60}$$

$$s_j = \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l \prod_{l' \notin C} (1-\phi_{l'}) s_j^\star(C) \tag{A.61}$$

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

$$\frac{\partial s_0}{\partial p_j} - \frac{\partial s_j}{\partial p_0} = \frac{\partial \log(\phi_j)}{\partial p_j} \left(s_0(\mathcal{J}) - s_0(\mathcal{J}/j)\right) - \frac{\partial \log(\phi_0)}{\partial p_0} \left(s_j(\mathcal{J}) - s_j(\mathcal{J}/0)\right) \tag{A.62}$$

This expression might seem somewhat odd given that 'leave-zero-out' variation is required. How natural this assumption is might vary across contexts. If default goods are randomly assigned in the population, this variation (or permitting the market share of good-0 to go to zero) might be plausible.

A.6 Identification of Features in General Consideration Set Model

Our proof of constructive point identification relies on the structure imposed by the ASC and DSC frameworks. However, features of a more general model of consideration sets can still be identified from cross-derivative asymmetries. This remains the case with correlation between the unobservables driving consideration probabilities. To illustrate, let $g_j = x_j\gamma$ and assume that the impact of characteristics on attention 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:

$$\frac{\partial s_j}{\partial x_{j'}^k} - \frac{\partial s_{j'}}{\partial x_j^k} = \sum_{C\in\mathbb{P}(j)} \frac{\partial \pi_C(g_0,...,g_J)}{\partial x_{j'}^k} s_j^\star(C) - \sum_{C'\in\mathbb{P}(j')} \frac{\partial \pi_{C'}(g_0,...,g_J)}{\partial x_j^k} s_{j'}^\star(C') \tag{A.63}$$

$$= \gamma_k \sum_{C\in\mathbb{P}(j)} \frac{\partial \pi_C(g_0,...,g_J)}{\partial g_{j'}} s_j^\star(C) - \sum_{C'\in\mathbb{P}(j')} \frac{\partial \pi_{C'}(g_0,...,g_J)}{\partial g_j} s_{j'}^\star(C') \tag{A.64}$$

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

$$\frac{\frac{\partial s_j}{\partial x_{j'}^k} - \frac{\partial s_{j'}}{\partial x_j^k}}{\frac{\partial s_j}{\partial x_{j'}^{k'}} - \frac{\partial s_{j'}}{\partial x_j^{k'}}} = \frac{\gamma_k}{\gamma_{k'}} \tag{A.65}$$

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.7 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:

$$\frac{\partial s_k}{\partial p_j} - \frac{\partial s_j}{\partial p_k} = \frac{\partial \mu(\mathbf{p})}{\partial p_j} s^\star_k(\mathbf{p}|\mathcal{J}) \tag{A.66}$$

$$= \frac{\partial \log \mu(\mathbf{p})}{\partial p_j} s_k(\mathbf{p}|\mathcal{J}) \tag{A.67}$$

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

$$\frac{\partial \log \mu(\mathbf{p})}{\partial p_k} = \frac{\partial \log \mu(\mathbf{p})}{\partial p_{k'}} \tag{A.68}$$

then changes in consideration probabilities can be recovered by inverting a system of linear equations. For example, consider the $J = 2$ case.

$$\begin{bmatrix} -s_1 & (s_0 - 1) \\ -s_2 & (s_0 - 1) \end{bmatrix} \begin{bmatrix} \frac{\partial \log(\mu)}{\partial p_0} \\ \frac{\partial \log(\mu)}{\partial p_{\neg 0}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_0}{\partial p_1} - \frac{\partial s_1}{\partial p_1} \\ \frac{\partial s_0}{\partial p_2} - \frac{\partial s_2}{\partial p_1} \end{bmatrix} \tag{A.69}$$

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

A.8 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, but 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:

$$s_{jm} = \sum_{n=1}^{N} \omega^n(w) s^n_j(x_m) \tag{A.70}$$

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

$$s^n_{jm} = \mu^n_m \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi^n_l \prod_{l' \notin C} (1 - \phi^n_{l'})\, s^{n\star}_{jm}(C) \tag{A.71}$$

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.70 (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. The probability of choosing any specific alternative as a function of the parameters $\beta = (\theta, \gamma)$ is given by:

$$s_j(x;\beta) = \sum_{c\in\mathbb{P}(j)} \prod_{l\in c} \phi_l(x_l;\beta) \prod_{l'\notin c} (1-\phi_{l'}(x_{l'};\beta)) s_j^{\star}(x|C,\beta) \tag{B.1}$$

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

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{N} \sum_{j=0}^{J} y_{ij} \log\left(s_j(\beta)\right) \tag{B.2}$$

We note that a major computational issue arises with larger choice sets - there are 2^J possible consideration sets to sum over to construct choice probabilities. To deal with this problem, we advocate that researchers follow the simulated likelihood approach outlined in Goeree (2008) and we refer to readers to https://sites.google.com/view/alogit/home for further practical details.

B.2 Indirect Inference

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

$$\widetilde{u}_{ij} = \theta_j^0 + x_{ij}\theta_j + \sum_{j'} \sum_{k} \sum_{k'} \theta_{j,j'}^{k} x_{ij}^{k} x_{ij'}^{k} + e_{ij} \tag{B.3}$$

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

$$\widetilde{f}_{ij} = \frac{\exp\left(\theta_j^0 + x_{ij}\theta_j + \sum_{j'}\sum_k \theta_{j,j'}^k x_{ij}^k x_{ij'}^k\right)}{\sum_{l=0}^J \exp\left(\theta_j^0 + x_{il}\theta_l + \sum_{l'}\sum_k \theta_{l',l}^k x_{il'} x_{il}\right)} \tag{B.4}$$

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

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

$$\widetilde{\theta}^m(\beta) = \arg\min_\theta \sum_{i=1}^N \sum_{j=0}^J y_{ij} \log\left(f_{ij}^m(\beta)\right) \tag{B.6}$$

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

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

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

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

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

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

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

$$u_{ij} = v_j + \epsilon_{ij} \tag{B.9}$$

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

$$s_1 = \phi_1(x_1) s_1^\star(\mathbf{x}) \tag{B.10}$$

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

$$\widetilde{u}_{ij} = v_j + \psi_{j=1} + \epsilon_{ij} \tag{B.11}$$

where $\psi_{j=1} = \psi_1$ for plan 1 and is 0 otherwise, where ψ_1 is given by:

$$\psi_1 = \ln \left(\frac{\phi_1(x_1)\exp(v_0)}{(1-\phi_1(x_1))\exp(v_1) + \exp(v_0)} \right) \tag{B.12}$$

This follows since:

$$s_1 = \frac{\exp(v_1 + \psi_{j=1})}{\exp(v_1 + \psi_{j=1}) + \exp(v_0)} = \phi_1(x_1)\frac{\exp(v_1)}{\exp(v_1) + \exp(v_0)} \tag{B.13}$$

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

$$\psi_j = \ln \left(\frac{\phi_j \sum_{k \neq j} \exp(v_k + \psi_k)}{(1-\phi_j)\exp(v_j) + \sum_{k \neq j} \exp(v_k + \psi_k)} \right) \tag{B.14}$$

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:

$$s_j^a = \sum_{C \in \mathbb{P}(j)} \prod_{l \in C, l \neq j} \phi_l \prod_{l' \notin C} (1 - \phi_{l'})\, s_j^\star(C) \tag{B.15}$$

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

$$s_J = \phi_J s_J^a \tag{B.16}$$

By the inductive hypothesis, we have:

$$s_J^a = \frac{\exp(v_J)}{\exp(v_J) + \sum_{k \neq J} \exp(v_k + \psi_k)} \tag{B.17}$$

Therefore,

$$s_J \quad = \quad \phi_J \frac{\exp(v_J)}{\exp(v_J) + \sum_{k \neq J} \exp(v_k + \psi_k)} \tag{B.18}$$

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:

$$u_{iJ} = v_J + \psi_J + \epsilon_{iJ} \tag{B.19}$$

where:

$$\psi_j = \ln\left(\frac{\phi_j \sum_{k \neq j} \exp(v_k + \psi_k)}{(1 - \phi_j)\exp(v_j) + \sum_{k \neq j} \exp(v_k + \psi_k)}\right) \tag{B.20}$$

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 ψ_j with respect to the attributes of rival goods around the point where all of these attributes are 0 so that $\psi_j = 0$.

Next, consider the DSC model.

$$\begin{aligned} s_0 &= (1 - \mu) + \mu s_0^\star(\mathcal{J}) \\ s_j &= \mu s_j^\star(\mathcal{J}) \quad \text{ for } j > 0 \end{aligned} \tag{B.21}$$

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

$$u_{ij} = v_j + \psi_{j=d} + \epsilon_{ij} \tag{B.22}$$

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

$$\psi = \ln\left(\frac{1 + (1 - \mu)\sum_{k \neq d} \exp(v_k - v_d)}{\mu}\right) \tag{B.23}$$

C ASC Additional Data and Robustness Checks

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

Table 9 shows the products used in the experiment and their list prices. A sample product selection screen is shown in Figure 1.

Table 9: Product Names and Prices

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

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

Table 10 reports estimates of the ASC model for the subset of experimental participants who correctly answered the question testing their understanding of the instructions. The results are very comparable to Table 1 in the text.

Table 10: Experimental Data Estimation Results

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

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

D Application 2: Optimal Defaults Medicare Part D

Relationship to General Theorem To nest the DSC model within our general framework, let $\phi_{ijt} = 1$ for all $j > 0$. Different theoretical models of inattention suggest different functional forms for the probability of making an active choice, μ_{it}. In the model of Ho, Hogan, and Scott Morton (2015), consideration is driven by health shocks while in Heiss, McFadden, Winter, Wupperman, and Zhou (2016) changes in plan characteristics can trigger a consumer to search. In the rational inattention models, a consumer will actively search only if her expected benefits exceed the costs. For example, in a framework inspired by Moshkin and Shachar (2002), a consumer will actively consider products if:

$$x_{idt}\beta + \nu_{idt} \leq \max_{j \neq d} \{x_{ij,t-1}\beta\} - \rho - \alpha - c \tag{D.1}$$

where c is the cost of paying attention.

We employ a specification that encompasses these different models as

$$\mu_{it} = \Pr\left(f\left(z_{it}, x_{idt}, x_{i1,t-1}, ..., x_{iJ,t-1}\right) \geq \eta_{it}\right) \tag{D.2}$$

$$= \frac{\exp\left(f\left(z_{it}, x_{idt}, x_{i1,t-1}, ..., x_{iJ,t-1}\right)\right)}{1 + \exp\left(f\left(z_{it}, x_{idt}, x_{i1,t-1}, ..., x_{iJ,t-1}\right)\right)} \tag{D.3}$$

The probability of selecting option j is expressed as:

$$s_{ijt} = (1 - \mu_{it})\, Default_{ijt} + \mu_{it} s^{\star}_{jt}(x_{ijt}) \tag{D.4}$$

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

D.1 Discussion of Dube et. al. Decomposition

In the text, we decompose structural state dependence into inattention and switching costs. Dubé, Hitsch, and Rossi (2010) instead decompose structural state dependence into loyalty, learning, and search effects. They define loyalty effects as psychological switching costs which consumers must incur to switch products, learning as the degree to which experience alters the degree of state dependence, and search as the degree to which direct information provision about rival products impacts choices (in their empirical work, aisle displays).

We have three objections to this decomposition: first, insensitivity to information about rival products does not rule out search costs; in a rational inattention / search model, consumers may be inattentive to information about rival goods because that information is costly to process or because that information is incomplete (and obtaining more information would be costly). Second, consumers may be naively inattentive–they may be inattentive, and their decision of whether or not to attend to information about other goods may not be rational. Third, rather than calling "learning" a separate reason for structural state dependence, we prefer to think of learning as a factor that can impact both switching costs and search (as both can change over time with experience).

These distinctions are important for welfare – whereas Dubé, Hitsch, and Rossi (2010) would conclude that consumers who were insensitive to a mailing providing information must have large switching costs and would be worse off were they forced to switch, the DSC model allows for the possibility that such consumers are not paying attention and might be better off with a different choice.

E Practical Overidentification Test in the DSC Model

The DSC model models choice probabilities as:

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

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

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

As noted in Section 6, the DSC model with linear utility and logit errors can be written as a random utility model where utility depends on the characteristics of rival goods:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

E.1 Testing the Validity of Additional Exclusion Restrictions

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

We test these assumptions by estimating the DSC model and allowing each of the attributes listed above to potentially impact both attention and utility. The utility coefficients relevant for these tests are reported in Table 11. 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.[62] The remainder of Table 11 reports interactions between each of the preference parameters and age dummies, non-white dummies and experience dummies (experience is defined as the number of years since 2006 for which you enrolled in Part D). Heiss, McFadden, Winter, Wupperman, and Zhou (2016) assume that all of these terms are 0 – in other words, preferences are invariant to these attributes. We find several cases where this assumption is rejected – more 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 increases from 13.1% in the model with these additional terms allowed to impact attention but not utility to 15.4% when they are excluded

[62] 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.

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 11: 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%.

E.2 Robustness to Alternative Specifications and Samples

As shown in Table 2, non-LIS and reassigned LIS beneficiaries have different demographic characteristics. To check the robustness of our results to these differences, we generate a sample of 100,000 non-LIS beneficiaries that is chosen to resemble LIS beneficiaries in terms of available demographics using a Mahalanobis distance metric. Column 2 of Table 12 reports structural estimates of the conditional logit and DSC model using the matched sample in lieu of the 100K sample. The estimates are qualitatively similar to those reported in the text. The estimates of adjustment costs and paperwork costs are slightly smaller, meaning that the implied benefits of smart defaults are accordingly larger. Column 3 reports estimates using the sample of all reassigned beneficiaries regardless of age rather than only those of age 65+. These estimates are extremely close to those in the main text.

Table 12: Part D Data: Conditional and Attentive Logit Estimates

	100k Sample and Final LIS Sample		Matched Sample and Final LIS Sample		100k Sample and Final LIS Sample All Ages	
	Clogit	Alogit	Clogit	Alogit	Clogit	Alogit
Utility:						
Original Plan	0.988***	1.314***	1.283**	1.215**	0.708***	1.248***
	(0.238)	(0.257)	(0.453)	(0.472)	(0.176)	(0.191)
Assigned Plan	6.428***	4.240***	6.553***	2.983***	6.461***	4.220***
	(0.012)	(0.078)	(0.017)	(0.141)	(0.012)	(0.080)
Annual Premium (hundreds)	-0.505***	-1.034***	-0.590***	-1.241***	-0.509***	-1.030***
	(0.005)	(0.010)	(0.008)	(0.015)	(0.005)	(0.010)
Annual Out of Pocket Costs (hundreds)	-0.214***	-0.297***	-0.236***	-0.407***	-0.214***	-0.298***
	(0.007)	(0.012)	(0.011)	(0.022)	(0.007)	(0.012)
Variance of Costs (millions)	2.246***	2.579***	2.321***	3.886***	2.241***	2.607***
	(0.089)	(0.165)	(0.151)	(0.297)	(0.088)	(0.163)
Deductible (hundreds)	-0.516***	-0.724***	-0.401***	-0.555***	-0.498***	-0.714***
	(0.009)	(0.013)	(0.012)	(0.018)	(0.009)	(0.013)
Donut Hole Coverage	0.691***	0.335***	0.884***	0.498***	0.704***	0.329***
	(0.027)	(0.051)	(0.040)	(0.090)	(0.027)	(0.051)
Average Consumer Cost Sharing %	-1.181***	-4.128***	-0.256*	-1.939***	-1.045***	-4.002***
	(0.107)	(0.163)	(0.142)	(0.235)	(0.105)	(0.162)
# of Top 100 Drugs in Formulary	0.038***	0.172***	0.011***	0.125***	0.031***	0.173***
	(0.004)	(0.006)	(0.005)	(0.008)	(0.004)	(0.006)
Normalized Quality Rating	0.438***	0.515***	0.481***	0.430***	0.437***	0.510***
	(0.010)	(0.015)	(0.014)	(0.023)	(0.010)	(0.015)
Attention:						
Annual Premium (hundreds)	-	0.062***	-	0.134***	-	0.070***
	-	(0.014)	-	(0.019)	-	(0.014)
Annual Out of Pocket Costs (hundreds)	-	0.030*	-	0.037	-	0.033**
	-	(0.012)	-	(0.020)	-	(0.012)
Variance of Costs (millions)	-	-0.627***	-	-0.621*	-	-0.654***
	-	(0.159)	-	(0.275)	-	(0.158)
Deductible (hundreds)	-	0.069***	-	0.211***	-	0.029
	-	(0.020)	-	(0.024)	-	(0.020)
Donut Hole Coverage	-	-0.761***	-	-0.502***	-	-0.802***
	-	(0.052)	-	(0.081)	-	(0.052)
Average Consumer Cost Sharing %	-	-1.447***	-	0.150	-	-1.770***
	-	(0.219)	-	(0.248)	-	(0.217)
# of Top 100 Drugs in Formulary	-	-0.002	-	-0.036***	-	0.043***
	-	(0.010)	-	(0.010)	-	(0.009)
Normalized Quality Rating	-	-0.511***	-	-0.720***	-	-0.498***
	-	(0.019)	-	(0.024)	-	(0.019)
Adjustment Costs	\$196	\$127	\$218	\$97	\$139	\$121
Paperwork Costs	\$1,078	\$283	\$894	\$142	\$1,131	\$288
Attention Probability	-	19.7%	-	11.6%	-	19.2%

The remaining tables repeat our simulations using estimates from the DSC model when the reassigned population includes individuals under the age of 65 and is thus substantially larger. The results below exactly parallel those in the text using individuals only over the 65 and they are qualitatively identical. Thus, we report the tables with no further comment. Table 13 shows whether individuals in this simulation save money relative to remaining in their current plan as a function of whether they choose the new default (the lowest cost plan), the old default (the original plan), or some other plan.

Table 13: Welfare Simulations - Savings Relative to Current Plan for All Ages

	New Default	Old Default	Other Plan
Savings from Plan:			
<$0	0	0	1,487
$0 - $100	11,689	6,742	1,035
$100 - $200	22,611	0	1,054
$200 - $300	36,895	0	633
$300 - $400	27,986	0	245
$400 - $500	11,658	0	107
$500 - $600	4,420	0	68
$600 - $700	4,095	0	42
$700 - $800	3,283	0	23
$800 - $900	2,740	0	-
$900 - $1000	1,672	0	-
$1000 +	5,564	0	82
Mean Savings:	$423	$0	$124

Notes: The table shows the number of beneficiaries with no age restrictions who choose the lowest cost plan (new default), stay in their current plan (old default) or switch to a third plan (other plan). The table also reports how much money individuals save relative to remaining in their original plan.

Next, we consider the overall welfare impact, taking into account risk protection, plan quality, the adjustment and paperwork costs, as well as any additional costs of paying attention. These results are shown in Table 14.

Table 14: Welfare Simulations - Overall Welfare Impact for All Ages

	Attention Cost				
	\$0	\$50	\$100	\$200	\$300
DSC Parameters	\$184	\$184	\$184	\$184	\$184
Direct Effect on Attention Probability					
25%	\$178	\$174	\$171	\$165	\$158
50%	\$149	\$133	\$117	\$86	\$54
75%	\$116	\$88	\$60	\$3	-\$53
100%	\$79	\$39	-\$2	-\$83	-\$165

Notes: The table shows overall welfare impacts of alternative assumption about the probability of paying attention and the cost of paying attention with no age restrictions. Each row shows a different assumption about the probability of paying attention, while each column shows alternative assumptions about the cost of paying attention.

What if we reassign only those beneficiaries for whom the potential cost savings exceed our estimated adjustment costs? Table 15 shows results.

Table 15: Welfare Simulations - Overall Welfare Impact Restricted Reassignment for All Ages

	Attention Cost				
	$0	$50	$100	$200	$300
DSC Parameters	$228	$228	$228	$228	$228
Direct Effect on Attention Probability					
25%	$218	$214	$210	$201	$192
50%	$186	$169	$152	$118	$84
75%	$152	$122	$93	$34	-$24
100%	$115	$74	$33	-$48	-$130

Notes: The table shows overall welfare impacts of alternative assumption about the probability of paying attention and the cost of paying attention with no age restrictions. Each row shows a different assumption about the probability of paying attention, while each column shows alternative assumptions about the cost of paying attention. In this case, only beneficiaries for whom the cost savings from the new default plan exceeds the estimated adjustment costs are reassigned.