Endogenous Information and Simplifying Insurance Choice*

Zach Y. Brown
University of Michigan & NBER

Jihye Jeon
Boston University

August 2023

Abstract

In markets with complicated products, individuals may choose how much time and effort to spend understanding and comparing alternatives. Focusing on insurance choice, we find evidence consistent with individuals acquiring more information when there are larger consequences from making an uninformed choice. Building on the rational inattention literature, we develop and estimate a parsimonious demand model in which individuals choose how much to research difficult-to-observe characteristics. We use our estimates to evaluate policies that simplify choice. Reducing the number of plans can raise welfare through improved choice as well as savings in information costs. Capping out-of-pocket costs generates larger welfare gains than standard models. The empirical model can be applied to other settings to examine the regulation of complex products.

Keywords: insurance, information frictions, rational inattention
JEL Classification: L15, I13, D83

*We thank Ying Fan, Giuseppe Forte, Michael Grubb, Kate Ho, Marc Rysman, Fiona Scott-Morton, and Mo Xiao for helpful comments and suggestions. We are also grateful to seminar and conference participants at ASSA, Boston IO, Boston U, CeMENT Workshop, Harvard, IIOC, MIT, NBER IO Winter, NBER Summer Institute, Princeton, SITE, SMU, U Michigan, U Pennsylvania, WUSTL, and Yale. Zach Brown received support from the National Bureau of Economic Research, Michigan Institute for Teaching and Research in Economics, and National Institute on Aging, Grant Number T32-AG000186. We thank Chuqing Jin and Juan Sebastián Fernández for excellent research assistance.

1 Introduction

In markets where products have complicated features that are difficult to observe or understand, individuals may incur significant cost conducting research before making a choice. Moreover, the amount of research that an individual does may depend on the perceived benefits of information and the information cost they face, with the possibility of choosing dominated options. Therefore, evaluating a policy in this setting requires understanding how the policy will affect incentives to acquire information, and ultimately, choice quality. This issue is particularly relevant for insurance choice. While premiums are easy to observe, out-of-pocket costs can be difficult to compare given that insurance contracts often have complicated non-linear designs with different reimbursement and cost sharing policies for different types of claims.

We develop a tractable micro-founded framework for empirically examining demand when individuals choose how much information to acquire, which can be applied to a broad range of settings featuring complicated product attributes. A key prediction of the model is that individuals acquire more information when facing higher *stakes*, or consequences from making a poor choice. This can be contrasted with standard discrete-choice demand models in which there is no scope for the stakes to affect demand. Using data from Medicare prescription drug insurance (Part D), we provide evidence consistent with our model's predictions. Estimating the empirical model incorporating costly information acquisition, we show that costly information has important welfare consequences.

The model builds on theoretical work incorporating rational inattention in discrete choice models (Matějka and McKay 2015). In the model, individuals know plan premiums and other easy-to-observe product characteristics and then decide how much to research difficult-to-observe characteristics such as out-of-pocket costs given their prior beliefs. The more research individuals do, the more accurate their beliefs will tend to be. While there is a growing theoretical literature on rational inattention, it is difficult to separately identify heterogeneous preferences from information frictions. Moreover, the complexity of the rational inattention model makes estimation challenging. We derive a novel analytical solution for choice probabilities that incorporates preference heterogeneity, allowing for a feasible estimation strategy. We show

how the model can be identified by leveraging the fact that some characteristics are always observed.

The model has distinct implications for how choices change when individuals face higher stakes. We document evidence consistent with the model using administrative data from Medicare prescription drug insurance. Focusing on individuals forced to make an active choice, we find that the quality of decision making is affected by the stakes. In order to help address concerns that this finding is driven by preferences that are correlated with the stakes, we show that the results hold when exploiting within-individual variation in the stakes. In other words, in years in which an individual faces higher stakes, such as when she is expecting to be in the coverage gap, she makes choices that are consistent with having acquired more information.

We estimate an empirical model with endogenous information and recover individuals' marginal cost of information, which is a key structural parameter capturing the cost of reducing uncertainty by one unit. Importantly, the model allows for heterogeneous marginal cost of information across individuals to account for the fact that researching plans may be easier for certain individuals, such as younger Medicare enrollees and those with previous experience choosing a plan.

Empirical results imply that endogenous information frictions play an important role in our setting. If individuals had full information, they would choose plans that had somewhat higher average annual premiums ($642 vs. $570) in exchange for significantly lower out-of-pocket costs ($601 vs. $713). In addition, costly information also causes individuals to choose plans with suboptimal quality and risk. Overall, the average annual increase in welfare of full information is $412 per individual, of which $127 is the information cost.

We use the estimates to examine the implications of simplifying choice. In standard demand models, restricting the choice set strictly decreases welfare, which seems at odds with individuals' strong desire for a reduced and simplified choice set as documented in existing surveys (Altman et al. 2006). By contrast, limiting options can potentially improve consumer welfare in our model by reducing the amount of research that individuals need to do and reducing the probability of choosing plans that would not be optimal under full information. Our counterfactual experiment reveals that removing a quarter of the plans with the lowest mean utility increases annual welfare by $55 per enrollee, approximately 80% of which is due to a reduction in individuals' chosen research effort. However, if the choice set is restricted too much, individuals with heterogeneous preferences cannot find a plan that is a good match, reducing welfare. We also consider a cap on out-of-pocket costs, which is particularly important given the recent growth in cost sharing. Imposing the cap can also impact consumer welfare through both lowering information acquisition costs and improving choices in our framework, leading to a larger welfare gain than implied by commonly used demand models.

Overall, we argue that a micro-founded framework that endogenizes information is able to rationalize key facts in the market we examine. More generally, the framework has important implications for simplifying choice and consumer protection in markets featuring complex choices.

Our model of endogenous information acquisition builds on the rational inattention model originally developed by Sims (2003). We leverage theoretical results from Matějka and McKay (2015) that link rational inattention models to discrete choice demand.[1] There is limited work applying this model to structural estimation (e.g., Joo 2023).[2] We develop a tractable model with both observed and initially unobserved product characteristics and a novel identification strategy that can be applied to a variety of settings.

Our work is related to the large literature on choice frictions in health insurance markets. There is an influential literature documenting that individuals choose expensive or dominated health insurance plans (e.g. Abaluck and Gruber 2011; Heiss et al. 2013; Bhargava et al. 2017). Individuals do not fully understand health insurance plans (Handel and Kolstad 2015) and respond to easy-to-use information (e.g. Kling et al. 2012). Some papers argue inattention may be a driving force of inertia in insurance plan choice (e.g. Handel 2013; Heiss et al. 2016; Ho et al. 2017). Our work also complements papers that assess the rationality of individual choices in Medicare Part D markets (e.g. Ketcham et al. (2015)) and papers that evaluate policies re-

[1] In related work, Fosgerau et al. (2020) generalize the cost function of information.

[2] There is also limited work testing the rational inattention framework in real-world setting (e.g. Bhattacharya and Howard (2020)).

ducing choice in Part D (e.g. Lucarelli et al. (2012)). Relative to this literature, our paper proposes a micro-founded approach for modeling information and focuses on the role of endogenous information. We show that endogenous information matters for explaining choice behavior and evaluating policy in our setting.

Finally, our approach is related to the literature on consumer search (see Honka et al. (2019) for a recent survey) and consideration set models (e.g. Coughlin 2019; Abaluck and Adams-Prassl 2021). Unlike these models, the rational inattention framework implies that individuals may choose to acquire partial information about any of the options in their choice set. In general, search models are well suited to situations with a large number of simple options while the rational inattention approach is a natural framework for analyzing markets with complicated products.

Section 2 presents the general framework. Section 3 discusses background and data as well as motivating evidence. Section 4 presents an empirical framework and Section 5 presents counterfactual results. Section 6 concludes.

2 Theoretical Framework

In this section, we present a discrete choice model in which individuals maximize expected utility when part of the utility, such as out-of-pocket payments, is initially unobserved unless individuals acquire costly information. We leverage theoretical results from Matějka and McKay (2015) linking the rational inattention framework with discrete choice models. Matějka and McKay (2015) focus on the conditions necessary for equivalence between rational inattention and random utility models. In contrast, our model is useful for clarifying how demand with endogenous information acquisition differs from standard demand models when attributes are initially partially observed and individuals may have an idiosyncratic taste shock. We show that, under relatively innocuous assumptions, one can derive a straightforward expression for choice probabilities that nests logit choice probabilities as a special case.

Consider individual i choosing option $j \in \mathcal{J}$ where the choice set is defined by $\mathcal{J}$. Each alternative has two components of cost, p_j and v_{ij}. Individual i observes component p_j, but does not initially observe v_{ij}, unless the individual acquires costly information. We also allow for other characteristics X_j^u, which are initially unknown to the individual, and characteristics X_j^k, which are initially known. Utility is given by

$$u_{ij} = \underbrace{\alpha v_{ij} + \beta X_j^u}_{\text{Initially Unknown}} + \underbrace{\alpha p_j + \theta X_j^k + \epsilon_{ij}}_{\text{Known}}. \tag{1}$$

We include an idiosyncratic taste shock, ϵ_{ij}, which is assumed to be iid with variance normalized to $\pi^2/6$. We assume that the taste shock is initially known to the individual, but not to the econometrician.[3]

In the case of insurance choice, p_j is the premium and v_{ij} is expected out-of-pocket costs with rational expectations. Information on plan premiums is readily available, often listed on websites or in published material. Conversely, individual-specific expected out-of-pocket costs are difficult to observe as it requires forming expectations about claims and mapping those claims to out-of-pocket costs via complicated insurance contracts that potentially involve deductibles, copays, coinsurance, and catastrophic coverage. Equation (1) can be considered the "full-information" utility if individuals perfectly researched insurance plans.[4] In the case of insurance choice, individuals may also be risk adverse. We incorporate risk aversion into our empirical model in Section 4.

Let $\xi_{ij} \equiv \alpha_i v_{ij} + \beta_1 X_j^u$ be the component of utility that is initially unknown to the individual but can be observed with costly information acquisition. Individuals have a prior about ξ_{ij} for each option in the choice set. Let this multivariate distribution have CDF given by $G_i(\boldsymbol{\xi_i})$ where $\boldsymbol{\xi_i} = (\xi_{i1}, \cdots, \xi_{iJ})$.

We assume that individuals have prior mean ξ_{ij}^0, which may differ across options. Prior variance is σ_i^2, which is common to all options in an individual's choice set. As we describe in greater detail below, σ_i^2 plays a key role in the model. The prior distributions for each option are assumed to be independent. Expected utility before

[3]We assume that the econometrician observes all terms other than ϵ_{ij} in equation (1).

[4]In the case of health insurance, there may be health shocks that are realized after choosing a plan. This means that there may be a difference between the realized out-of-pocket costs and the expected out-of-pocket costs even under rational expectations. We define v_{ij} as expected out-of-pocket costs before the realization of these shocks.

information acquisition is

$$\mathbb{E}_G[u_{ij}] = \xi_{ij}^0 + \alpha p_j + \theta X_j^k + \epsilon_{ij}. \tag{2}$$

Following Matějka and McKay (2015), we can consider the decision problem having two stages. In the first stage, an individual optimally chooses how much information to acquire based on their prior and the cost of information. In particular, the individuals may choose to receive more precise signals about certain options in the choice set. Given the prior and signals, an individual then forms posterior beliefs about each option. In the second stage, an individual chooses an option that maximizes expected utility given these beliefs.

As is standard in the rational inattention literature, we adopt the entropy-based cost function for information. Given constant marginal cost of information λ, total cost of information takes the form

$$\lambda \left(H(G_i) - \mathbb{E}_{\mathbf{s}_i}[H(F_i(\boldsymbol{\xi}_i|\mathbf{s}_i))] \right) \tag{3}$$

where $F_i(\boldsymbol{\xi}_i|\mathbf{s}_i)$ is the posterior belief about $\boldsymbol{\xi}_i$ after receiving signal $\mathbf{s}_i$, and $H(F)$ is the entropy of belief F, which is a measure of uncertainty and is given by $H(F) = -\int_{\mathbf{x}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}$ when F has a pdf f. The total cost of information acquisition is proportional to the change in entropy between the prior and posterior. Thus, it can be thought of as a measure of the reduction in uncertainty after signals are received, often referred to as mutual information. This cost function is meant to reflect the time and cognitive load necessary to acquire and process information.[5]

Matějka and McKay (2015) characterize choice probabilities when individuals optimally choose the distribution of signals to maximize their expected payoff given the entropy-based cost function. Results from Matějka and McKay (2015) show that individual's choices after information acquisition are *as if* they maximize expected

[5]It is also possible to make the cost function more general by replacing λ with an alternative marginal cost function. See discussion in Cabrales et al. (2013) and Mackowiak et al. (2018) for further motivation for the cost function.

utility

$$\underbrace{u_{ij}}_{\text{Actual Utility}} + \underbrace{\lambda \log P^0_{ij}}_{\text{Contribution of Prior}} + \underbrace{\lambda \varepsilon_{ij}}_{\text{Belief Error}} \tag{4}$$

where P^0_{ij} is the expected choice probability based on the prior before the realization of signals, and can be obtained by solving

$$\max_{P^0_{i1},..,P^0_{iJ}} \int_{\boldsymbol{\xi_i}} \lambda \log \Sigma_j P^0_{ij} e^{(\xi_{ij}+\alpha p_j+\theta X^k_j+\epsilon_{ij})/\lambda} G_i(d\boldsymbol{\xi_i}) \text{ s.t. } \sum_{j\in\mathcal{J}} P^0_{ij} = 1,\, P^0_{ij} \geq 0\ \forall j. \tag{5}$$

Solving for P^0_{ij} using Equation (5) is computationally demanding and poses a significant challenge for estimating an empirical model based on the framework. We develop a tractable model of demand with endogenous information that can be easily applied to data. We do this by assuming that the distributions of the prior, $G_i(\xi_{ij})$, and the taste shock, $M(\epsilon_{ij})$, follow the Cardell distribution.[6] This leads to a closed-form expression for choice probabilities given by

$$P_{ij} = \frac{\exp\left[a(\sigma_i,\lambda)\left(\alpha v_{ij} + \beta X^u_j + \frac{\xi^0_{ij}}{\ell(\sigma_i,\lambda)}\right) + b(\sigma_i,\lambda)\left(\alpha p_j + \theta X^k_j\right)\right]}{\sum_{k\in\mathcal{J}} \exp\left[a(\sigma_i,\lambda)\left(\alpha v_{ik} + \beta X^u_k + \frac{\xi^0_{ik}}{\ell(\sigma_i,\lambda)}\right) + b(\sigma_i,\lambda)\left(\alpha p_k + \theta X^k_k\right)\right]} \tag{6}$$

where

$$a(\sigma_i,\lambda) \equiv \frac{\ell(\sigma_i,\lambda)-1}{(\ell(\sigma_i,\lambda)^2 + \lambda^2(\ell(\sigma_i,\lambda)-1)^2)^{\frac{1}{2}}},$$

$$b(\sigma_i,\lambda) \equiv \frac{\ell(\sigma_i,\lambda)}{(\ell(\sigma_i,\lambda)^2 + \lambda^2(\ell(\sigma_i,\lambda)-1)^2)^{\frac{1}{2}}},$$

$$\ell(\sigma_i,\lambda) \equiv \left(\frac{6\sigma_i^2}{\pi^2\lambda^2} + 1\right)^{\frac{1}{2}}.$$

We present the derivation of equation (6) and the discussion of our distributional assumption in Online Appendix A-1.[7]

[6] This implies that when the random variable is added to a random variable with a type 1 extreme value distribution, the resulting distribution is scaled type 1 extreme value. See Cardell (1997) and Galichon (2022) for details about this distribution. This distribution is also an integral part of the nested logit demand system.

[7] In Online Appendix G, we conduct a Monte Carlo exercise to assess the importance of the distributional assumption regarding the prior and argue the the model is an accurate approximation even if the distribution of the prior is misspecified and is actually normally distributed.

Equation (6) resembles choice probabilities from a standard logit model, but there are additional coefficients $a(\sigma_i, \lambda)$ and $b(\sigma_i, \lambda)$ that are determined by the variance of the prior, σ_i, and the marginal cost of information, λ. Equation (6) nests choice probabilities from the standard logit model when the marginal cost of information goes to zero. This is given by

$$P_{ij} = \frac{\exp\left[\alpha(v_{ij} + p_j) + \beta X_j^u + \theta X_j^k\right]}{\sum_{k \in \mathcal{J}_{it}} \exp\left[\alpha(v_{ik} + p_k) + \beta X_k^u + \theta X_k^k\right]}. \tag{7}$$

Given equation (6), expected utility after information acquisition can be expressed as

$$\mathbb{E}_F[u_{ij}] = a(\sigma_i, \lambda)\left(\alpha v_{ij} + \beta X_j^u + \frac{\xi_{ij}^0}{\ell(\sigma_i, \lambda)}\right) + b(\sigma_i, \lambda)\left(\alpha p_j + \theta X_j^k\right) + e_{ij} \tag{8}$$

where e_{ij} is distributed iid EV1 and represents the combined error due to idiosyncratic beliefs and the idiosyncratic taste shock.

We can consider the simple case in which there is no taste shock and individuals simply wish to minimize cost given by $p_j + v_{ij}$. When $v_{ij}^0 = v_i^0$ for all j so individuals have a homogenous prior, expected utility after information acquisition becomes

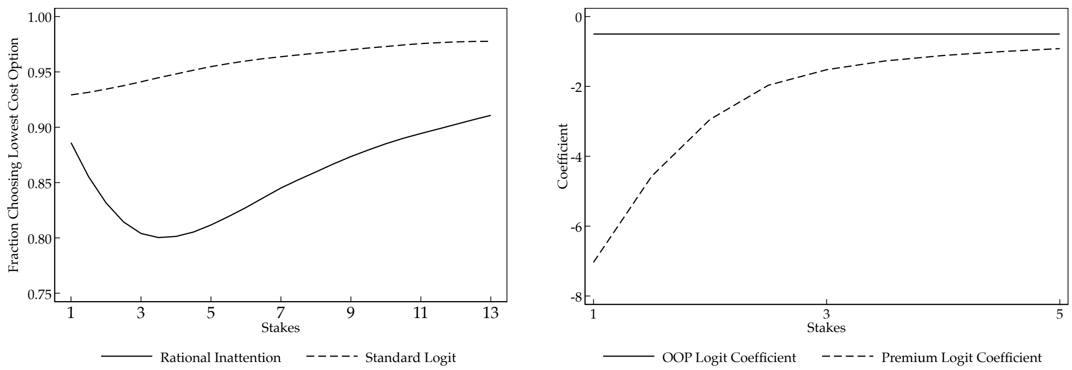
$$\mathbb{E}_F[u_{ij}] = -\underbrace{\frac{1}{\lambda}}_{\substack{\text{OOP}\\ \text{Weight}}} v_{ij} - \underbrace{\frac{\ell(\sigma_i, \lambda)}{\lambda(\ell(\sigma_i, \lambda) - 1)}}_{\substack{\text{Premium}\\ \text{Weight}}} p_j + \underbrace{e_{ij}}_{\substack{\text{Normalized}\\ \text{Belief Error}}}. \tag{9}$$

Under full information, an individual is indifferent between a marginal charge in v_{ij} and p_j. In contrast, when it is costly to observe v_{ij}, choices are as if individuals put more weight on p_j than v_{ij}. This can be seen by noting that the coefficient on p_j is larger than the coefficient on v_{ij} in equation (8) and equation (9). Specifically, the ratio of the coefficients on v_{ij} and p_j is given by

$$\frac{a(\sigma_i, \lambda)}{b(\sigma_i, \lambda)} = \frac{\ell(\sigma_i, \lambda) - 1}{\ell(\sigma_i, \lambda)}. \tag{10}$$

This ratio is determined by two key parameters in the model: the standard deviation of the prior, σ_i, and the marginal cost of information, λ. The ratio increases with the standard deviation of the prior. We interpret the standard deviation of the prior

Figure 1
Fraction Choosing Lowest Cost Plan and Logit Coefficients by Stakes



a. Fraction Choosing Lowest Cost Plan by Stakes

b. Logit Coefficient on Premium and Out-of-Pocket

Notes: Charts show simulations for the simplified discrete choice model with costly information acquisition (see equation (9). Panel a shows the mean fraction of individuals choosing the lowest cost option as a function of the stakes. In addition to showing the simulated results from the model with endogenous information, the dashed line shows the simulated choices from a logit model where utility is given by $u_{ij} = -v_{ij} - p_j + \varepsilon_{ij}$. Panel b shows implied coefficients as a function of the stakes. The solid line shows the coefficient on v_{ij} and the dashed line shows the coefficient on p_j. Simulations assume 3 options, $\lambda = 2$, p_j standard deviation of 4, and standard deviation of v_{ij} ranges from 1 to 13. The prior standard deviation σ_i is determined by the standard deviation of v_{ij}.

as a measure of the stakes. When individuals have a less precise prior, i.e. when σ_i is large, individuals are more worried about making a suboptimal choice when uninformed so there is more incentive to acquire information. In other words, individuals acquire more information when the stakes are high. Individuals also acquire more information when the marginal cost of information is low.

In Figure 1, we simulate choices from the simple version of the model in which choice probabilities are determined by equation (9). We assume that individuals know the distribution of v_{ij} across the choice set and this determines the variance of their prior. Therefore, $\sigma_i^2 = Var_j\left[v_{ij}\right]$. In the figure, choices are simulated for different values of $Var_j\left[v_{ij}\right]$ within the choice set, which determines the stakes.

Figure 1 Panel a shows the fraction of individuals choosing the lowest cost plan as a function of the stakes. A key implication of the model is that there is a non-monotonic relationship between the stakes and overspending. When the stakes are low, plans have similar out-of-pocket costs. Despite the fact that individuals exert low research effort, they often choose correctly just by choosing a plan with low premiums. As the stakes grow and comparisons become more complex, it becomes more difficult for individuals to choose the lowest cost plan despite the fact that they are acquiring more information. This implies a positive relationship between stakes and overspending. However, once the stakes are large enough, individuals become highly informed given the strong incentive to acquire information. In this range, there is a negative relationship between stakes and overspending.

Our model of endogenous information acquisition can be contrasted with standard demand models assuming full information. If utility is only a function of the cost and there is no idiosyncratic taste shock ($u_{ij} = -v_{ij} - p_j$), the stakes will have no effect on choices. In a logit demand model with a taste shock ($u_{ij} = -v_{ij} - p_j + \varepsilon_{ij}$ where ε_{ij} is EV1), there is a monotonic relationship between stakes and probability of choosing the least expensive plan. As the variance of v_{ij} grows, the taste shock becomes less important, generating a positive relationship. This can be seen in Figure 1 Panel a.

Moreover, the model has stark predictions for the effective weight that decision makers place on p_j and v_{ij}. When it is costly to observe v_{ij}, the weight that individuals appear to place on characteristics is endogenous and differs for p_j and v_{ij}. As shown in Figure 1 Panel b, the magnitude of the coefficient on p_j decreases when the stakes increase. As individuals acquire more information about v_{ij}, the weights on p_j and v_{ij} converge.[8]

In Section 3.2 we examine whether Medicare prescription drug insurance choice are affected by the stakes in a manner consistent with the model above. This motivates the structural model in Section 4.

3 Data and Motivating Evidence

We focus on Medicare prescription drug insurance, known as Medicare part D for our application. When individuals choose a Medicare prescription drug plan, it is

[8] Consequently, the elasticity of demand with respect to p_j and v_{ij} converges as well. The elasticities are derived in Online Appendix A-1.

easy to compare premiums either on the Medicare website or in printed material. As with other types of insurance, expected out-of-pocket costs are difficult to calculate. Individuals must know their likely drug usage over the coming year, including dosage and frequency. In addition, individuals must understand how this maps into out-of-pocket costs. Given the complexity of deductibles, copayments, coinsurance, the donut hole, and catastrophic coverage, this may require significant time and effort, especially for the older population that is eligible for Medicare Part D. Resources for patients often note that it is especially important for those with complex health care needs to research their Medicare plans.[9]

The Medicare website provides an online tool, PlanFinder, that helps individuals compare out-of-pocket costs across plans after entering information about drug usage. However, the tool is still difficult to use, especially for older patients that may not be familiar with the Internet. In surveys, individuals often report that the plans are still too complicated and difficult to compare.[10] The difficulty in comparing out-of-pocket costs is also highlighted by Kling et al. (2012), who find that individuals would choose less expensive plans with easier-to-use information.

3.1 Data

In order to construct out-of-pocket costs, we use a 20 percent sample of Medicare Part D beneficiaries from 2010 to 2015, 13.9 million unique individuals. We focus on the period starting in 2010 since this is the period in which we have detailed drug formulary data allowing us to construct out-of-pocket cost.[11]

In the context of our model, we wish to construct a measure of each individual's expected out-of-pocket cost for each plan in their choice set that reflects the beliefs

[9]For example, cancercare.org notes that "Choosing a Medicare plan, however, can be very challenging. Because costs are so high, it's especially important for people with cancer to understand how plans cover care and treatment." See https://www.cancercare.org/blog/choosing-the-right-medicare-program-when-you-have-cancer.

[10]For example, Altman et al. (2006) find in their survey that 73% of seniors, 91% of pharmacists, and 92% of doctors agree that the Medicare prescription drug benefit is too complicated. Additionally, 68% of seniors favor reducing the number of available plans. Also see Cummings et al. (2009).

[11]Due to a change in plan identifiers, we are not able to construct a comparable sample of individuals for 2013. For this reason, 2013 is removed from the sample.

of individuals if they used all available information. We construct two measures of out-of-pocket cost that closely follow Abaluck and Gruber (2016). For our primary measure, based on the rational expectations assumption, we compute out-of-pocket costs for each individual for each plan by applying the plan's formulary and cost sharing rules to observed drug utilization.[12] Then, we obtain expected out-of-pocket costs for each plan by averaging out-of-pocket costs across individuals with similar characteristics. Similarly, a plan's risk is calculated by considering the variance in out-of-pocket costs among similar individuals. We describe the procedure for constructing out-of-pocket costs in greater detail in Online Appendix B. Our alternative measure is based on a perfect foresight assumption. In this case, an individual's realized claims is used to construct her own out-of-pocket costs. This approach abstracts from moral hazard.

Similarly to Abaluck and Gruber (2016), we focus on individuals that are forced to make a choice due to the fact that they are new enrollees or their previous plan is no longer available, mitigating potential concerns about inertia.[13] The plan can become unavailable, for example, when the enrollee moves to a different market in which the plan is not offered or the insurer stops offering the plan. Individuals forced to make an active choice constitute 22.0 percent of the sample. For this sample, we argue that individuals are unlikely to start with information specific to certain plans. Finally, we eliminate choice situations in which individuals face stakes higher than $1,500, where stakes are defined below. This removes 2.2 percent of observations.[14] We use a 5 percent sample for the motivating analysis, which includes 206,851 choice situations. For the structural analysis, we use a 1 percent sample due to computational constraints.

Table 1 describes the final sample of active choice makers that we use for the descriptive analysis. The claims data contain information on age, sex, and chronic conditions of each individual. In addition, we use individuals' zip code to merge on

[12]As in Abaluck and Gruber (2016), we allow for substitution to equivalent drugs in less expensive tiers.

[13]The previous literature has documented the importance of consumer inertia in plan choice (e.g. Handel 2013; Polyakova 2016; Ho et al. 2017).

[14]The out-of-pocket cost calculator appears to be less accurate for those with extremely idiosyncratic drug needs, including those using very uncommon, expensive drugs.

Table 1
Summary of Insurance Choice for Active Choice Makers

	Mean	SD
Demographics:		
Age	76.2	7.4
Female	0.602	0.489
Zip income (1,000s)	77.3	35.1
Zip education (pct BA)	29.9	17.1
Rural	0.074	0.262
Years enrolled in Part D	5.53	2.31
Alzheimers	0.086	0.281
Lung disease	0.101	0.302
Kidney disease	0.157	0.364
Heart failure	0.132	0.339
Depression	0.118	0.322
Diabetes	0.268	0.443
Other chronic condition	0.303	0.460
Chosen option:		
Annual premium	674.1	378.1
Out-of-pocket cost (RE)	672.1	923.9
Out-of-pocket cost (PF)	678.6	1115.0
Relative to least expensive option:		
Difference (RE)	601.2	553.3
Percent difference (RE)	0.42	0.19
Difference (PF)	635.5	852.8
Percent difference (PF)	0.44	0.20
Plans in Choice Set	25.6	4.9
Number of individuals	90,187	
Choice situations	206,851	

Notes: Sample constructed from Medicare Part D beneficiaries that made an active plan choice from 2011 to 2015. Out-of-pocket cost (RE) measures expected annual out-of-pocket cost if individuals had rational expectations. Out-of-pocket cost (PF) measures annual out-of-pocket cost if individuals had perfect foresight.

education and income from the American Community Survey. The demographics of individuals that are forced to make an active choice are similar to the demographics of the overall Medicare Part D population. Consistent with the previous evidence, we find that the difference between the cost of an individual's chosen plan and the cost of the least expensive plan in their choice set is quite large on average, based on either measure of out-of-pocket costs.

We now turn to the definition of the stakes. The stakes are defined as the prior variance for the part of utility that is initially unobserved (ξ_{ij}). For our empirical setting, ξ_{ij} depends not only on v_{ij} and X_j^u, but also on the preference parameters

α and β_1, which require a full structural estimation. Therefore, for the purpose of the descriptive analysis in this section, we construct our measure of the stakes based simply on the expected out-of-pocket cost (v_{ij}), given that it is the major source of uncertainty. We account for other plan characteristics that may be initially unknown in our structural analysis in Section 4.

Individuals may understand the variance of v_{ij} across alternatives, forming the basis of their prior. For example, those currently taking new branded drugs that are not covered by all plans may understand that their out-of-pocket costs could vary widely depending on their plan choice. Motivated by this, we define the stakes as the standard deviation in expected out-of-pocket costs across plans in an individual's choice set.[15]

Individuals face mean stakes of \$227 and standard deviation of \$268, suggesting that there is significant variation in the stakes across choice situations. There is significant variation even within individual across years. The average within-individual standard deviation of de-meaned stakes is \$48.7. This variation may arise due to changes in health status or changes in the plans available. Our measure of the stakes is significantly correlated with health, including whether a patient has a chronic condition.[16] However, it is important to note that the stakes are not always higher when individuals face higher out-of-pocket costs. For instance, if individuals face very high out-of-pocket costs, they may hit the catastrophic coverage portion of Medicare Part D plans, leading to low variance in cost across plans. In this case, the individual could face relatively low stakes.

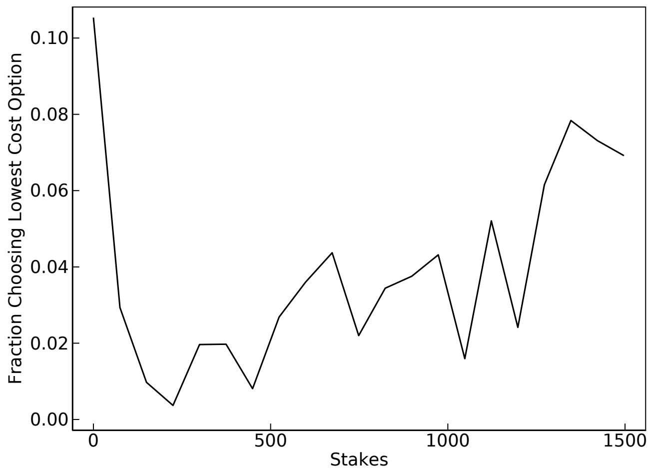
3.2 Motivating Evidence

Motivated by the predictions of the model in Section 2, we now examine how insurance plan choice is affected by the stakes using individual-level data on Medicare prescription drug plan choice.

[15] This is analogous to the standard assumption in the search literature that individuals know the distribution of prices in their choice set.

[16] The correlation is 0.85 with respect to total Part D spending and 0.22 with respect to having a chronic condition.

Figure 2
Fraction Choosing Lowest Cost Plan by Stakes



Notes: Chart shows mean fraction of individuals choosing lowest cost option as a function of the stakes, defined as the standard deviation in out-of-pocket costs across plans in an individual's choice set. Standard error bars show 95% confidence interval for the mean.

Stakes and Overspending

We start by examining the relationship between the fraction of individuals choosing the lowest cost plan and the stakes. Figure 2 shows that there is a U-shaped relationship. The relationship is consistent with the predictions of the model in Figure 1 Panel b showing that individuals with the easiest choice and those with the most incentive to acquire information are the most likely to choose a low-cost plan. We interpret this as initial evidence consistent with the model. However, there are concerns that individuals facing high stakes have different preferences or different information costs from individuals facing low stakes.

In order to help address these concerns, we exploit within-individual variation and focus on the sample of individuals making active choices multiple times in the sample period. We define an indicator variable, denoted by y_{it}, for whether individual i chose the option with the lowest total cost, where the total cost is defined as the sum of the annual premium plus and the annual expected out-of-pocket cost calculated using rational expectations assumption. Then, for individual i in year t, we estimate the following linear probability model

$$y_{it} = \beta_0 + \alpha_1 Stakes_{it} + \alpha_2 Stakes_{it}^2 + \beta X_{it} + \eta \widetilde{\sigma}_{it}^2 + \gamma_i + \theta_t + \varepsilon_{it} \tag{11}$$

where γ_i are individual fixed effects, θ_t are year fixed effects, and X_{it} are characteristics of the choice including average star quality, average deductible, average generic coverage, average coverage in the donut hole, average cost sharing, and the number of plans in the choice set.[17] To account for risk aversion, we also include $\widetilde{\sigma}_{it}^2$, the average within-plan out-of-pocket cost variance. This measure is meant to capture variation in cost due to unpredictable health shocks after enrolling in a plan. By including individual fixed effects, identification of α_1 and α_2 exploits within-individual variation in the stakes across years. Year fixed effects control for changes in plans offered over the period. The primary hypothesis is that there is a U-shaped relationship between stakes and the dependent variable, i.e. $\alpha_1 < 0$ and $\alpha_2 > 0$.

Estimates are presented in Table 2. Across specifications including different controls and fixed effects, we consistently find that $\alpha_1 < 0$ and $\alpha_2 > 0$. The coefficients are all highly statistically significant. The preferred specification, presented in column 3, includes both individual and year fixed effects. The coefficients imply that individuals are initially less likely to choose the lowest cost plan as the stakes increase. However, once the stakes are higher than \$376, individuals are more likely to choose the lowest cost plan as the stakes increase. Controlling for plan characteristics and the number of plans in the choice set has little effect on the estimates, implying that the U-shaped relationship is not driven by differences in the choice set that may be correlated with the stakes. In columns 5 and 6, we allow more flexibility by letting the effect of stakes to vary by quintiles. These specifications also imply a non-monotonic effect. Specifically, the probability of choosing the lowest cost plan is lowest when the stakes are in the middle quintiles. Individuals may also face different stakes because of differences in the offered plans in their market.[18] In column 4, we include market fixed effects in order to examine the effect of within-market variation in the stakes

[17] The donut hole (coverage gap) refers to the range of drug costs for which out-of-pocket costs are high in Medicare Part D. It begins after total drug costs paid by the enrollee and her plan reach a certain limit and ends when the catastrophic coverage starts.

[18] There are 34 geographic regions that define markets for Medicare Part D.

Table 2
Non-Monotonic Effect of Stakes on Choice of Lowest Cost Insurance Plan

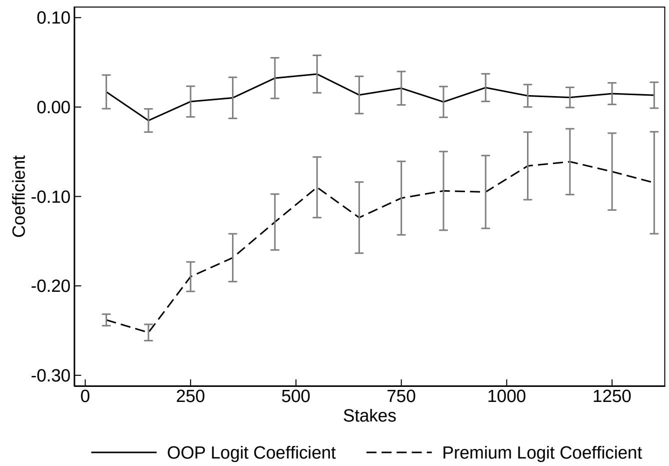
	(1)	(2)	(3)	(4)	(5)	(6)
Stakes (100s)	-0.0234^{***}	-0.0225^{***}	-0.0040	-0.0214^{***}		
	(0.0028)	(0.0030)	(0.0027)	(0.0031)		
Stakes Squared	0.0021^{***}	0.0020^{***}	0.0006^{***}	0.0019^{***}		
	(0.0003)	(0.0003)	(0.0002)	(0.0003)		
Stakes quintile 2					-0.0507^{***}	-0.0094^{**}
					(0.0042)	(0.0036)
Stakes quintile 3					-0.0585^{***}	-0.0171^{***}
					(0.0046)	(0.0039)
Stakes quintile 4					-0.0622^{***}	-0.0225^{***}
					(0.0051)	(0.0053)
Stakes quintile 5					-0.0491^{***}	-0.0093^{**}
					(0.0040)	(0.0043)
Individual FEs	No	No	Yes	No	No	Yes
Year FEs	No	No	Yes	Yes	No	Yes
Market FEs	No	No	No	Yes	No	No
Controls for Plan Characteristics & Number of Plans	No	Yes	Yes	Yes	Yes	Yes
Implied minimum	554.9	553.5	355.4	550.6		
Adjusted R2	0.009	0.012	0.299	0.016	0.024	0.299
Observations	199,783	193,745	183,402	193,745	193,745	183,402

Notes: Estimates from linear probability model where dependent variable is the indicator variable for whether the individual chooses the lowest cost plan. Stakes is defined as the standard deviation in expected out-of-pocket costs across plans in an individual's choice set. Standard errors clustered at the market level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and also find a U-shaped relationship.

Our findings on the relationship between the stakes and choice quality are robust to alternative measures of the stakes and choice quality. First, we use our perfect-foresight measure of out-of-pocket costs based on each individual's realized utilization to address a concern that there can be measurement error in our baseline measure. Second, we restrict our sample to new enrollees who might display different behavior and find that the U-shaped relationship is even more pronounced. Third, we explore a variety of alternative measures of choice quality such as the fraction of individuals choosing a plan in the lowest decile and quintile of out-of-pocket costs among the plans in their choice set. We also consider choice quality measures based on plan riskiness and quality. To the extent that these are also initially unobserved unless individuals conduct costly research, we would expect a similar relationship. Across all of these alternative measures, we find evidence of a U-shaped relationship between the stakes and choice quality. We provide details of these robustness results in Online

Figure 3
Logit Coefficient on Premium and Expected Out-of-Pocket
Cost by Stakes



Notes: Chart shows logit coefficient on annual out-of-pocket cost and annual premium interacted with indicators for the stakes. Logit specification includes controls for risk aversion (within-plan OOP variance), plan quality rating, deductible, generic coverage, coverage in the donut hole, and cost sharing. Standard error bars show 95% confidence interval.

Appendix D.

Stakes and Logit Coefficients

Another prediction we draw from the model in Section 2 is that the relative weight that individuals place on out-of-pocket cost and premiums varies with the stakes. To investigate this relationship in the data, we estimate a model based on the standard logit framework in this section.

We start by considering a specification for observable utility of plan j given by

$$\nu_{ijt} = \alpha_1 p_{jt} + \alpha_2 p_{jt} Stakes_{it} + \gamma_1 v_{ijt} + \gamma_2 v_{jt} Stakes_{it} + \theta \widetilde{\sigma}^2_{ijt} + \beta X_{ijt}. \qquad (12)$$

The specification controls for risk aversion by including $\widetilde{\sigma}^2_{ijt}$ as well as other plan characteristics, X_{ijt}. Given an additive iid EV1 error, choice probabilities are $P_{ijt} = \exp[\nu_{ijt}] / \left(\sum_k \exp[\nu_{ikt}]\right)$.

If the assumptions of the standard logit model hold, we would expect $\alpha_1 = \gamma_1$

since both coefficients should be equal to the negative marginal utility of income. The stakes do not affect decisions in the standard model; therefore $\alpha_2 = \gamma_2 = 0$. In contrast to the standard logit model, the model presented in Section 2 predicts $\alpha_1 < \gamma_1$ and $\alpha_2 > 0$, since individuals acquire more information about out-of-pocket costs when the stakes are high.

Figure 3 presents the results in graphical form by interacting stake bins with coefficients on premium and out-of-pocket cost.[19] When the stakes are low, individuals appear to place a high value on reducing premiums relative to the value that they place on reducing out-of-pocket cost, i.e. the coefficient on premium is low relative to the coefficient on out-of-pocket cost. This is consistent with the idea that individuals do not have incentive to become informed about out-of-pocket costs. As the stakes rise, the relative weight that individuals appear to place on premiums declines, consistent with the model predictions depicted in Figure 1.

The results using the specification described in equation (12) are presented in Table 3 column 2. Consistent with the model, the interaction of premium and stakes is positive and statistically significant. The interaction of out-of-pocket cost and stakes is very small and statistically insignificant, also consistent with the model.

The primary concern is that the results reflect heterogeneity in preferences that are correlated with the stakes rather than endogenous information acquisition. We address this in a few ways. First, we allow for heterogeneity in the price coefficients by including separate coefficients on observable individual characteristics interacted with the stakes. Observable individual characteristics include age, gender, race indicators, average chronic conditions, zip code income and education, and an indicator for rural locality. The results, presented in Table 3 column 3, are qualitatively similar.

To address the concern that there still may be unobserved preference heterogeneity, we include a separate coefficient on the interaction between premium and an individual's average stakes during the period. We also include out-of-pocket cost interacted with an individual's average stakes during the period. Therefore, within-individual variation in the stakes identifies the coefficients on $p_{jt} \times Stakes_{it}$ and $v_{jt} \times Stakes_{it}$.

[19] Formally, the logit specification assumes observable utility $v_{ijt} = \sum_g \alpha_g p_{jt} D_{ijtg} + \sum_g \gamma_g v_{jt} D_g + \theta \tilde{\sigma}^2_{ijt} + \beta Z_{ijt}$ where $Stakes_{it}$ is divided into groups indexed by g and $D_{ijtg} = 1$ if $Stakes_{it}$ is in group g and $D_{ijtg} = 0$ otherwise.

Table 3
Interaction of Stakes and Price Coefficient in Standard Logit Model

	(1)	(2)	(3)	(4)	(5)	(6)
Premium (100s)	-0.233^{***}	-0.276^{***}	-0.477^{***}	-0.291^{***}	-0.477^{***}	-0.477^{***}
	(0.003)	(0.003)	(0.020)	(0.003)	(0.021)	(0.021)
Premium × Indiv. avg stakes				0.019^{***}	0.017^{***}	0.017^{***}
				(0.001)	(0.001)	(0.001)
Premium × Stakes		0.020^{***}	0.017^{***}	0.008^{***}	0.007^{***}	
		(0.001)	(0.001)	(0.001)	(0.001)	
Premium × Stakes × $\mathbb{1}(\Delta > 0)$						0.005^{***}
						(0.001)
Premium × Stakes × $\mathbb{1}(\Delta < 0)$						0.011^{***}
						(0.001)
Out-of-Pocket Cost (100s)	-0.017^{***}	0.018^{***}	0.011	0.020^{***}	0.011	0.005
	(0.002)	(0.005)	(0.014)	(0.004)	(0.014)	(0.014)
OOP × Indiv. avg stakes				0.000	0.000	0.000
				(0.000)	(0.000)	(0.000)
OOP × Stakes		-0.000	-0.000	-0.000	-0.001^{**}	
		(0.000)	(0.000)	(0.000)	(0.000)	
OOP × Stakes × $\mathbb{1}(\Delta > 0)$						-0.001^{**}
						(0.000)
OOP × Stakes × $\mathbb{1}(\Delta < 0)$						-0.000
						(0.000)
Premium × Z_i	No	No	Yes	No	Yes	Yes
OOP × Z_i	No	No	Yes	No	Yes	Yes
Log Likelihood	-114,187	-113,814	-113,391	-113,654	-113,251	-113,230
Observations	1,025,674	1,025,674	1,025,674	1,025,674	1,025,674	1,025,674

Notes: Shows estimates from a logit demand model of plan choice (see equation (12)). Stakes are defined as the standard deviation in out-of-pocket costs across an individual's choice set and are measured in hundreds of dollars. All specifications include controls for risk aversion (OOP variance), plan quality rating, deductible, generic coverage, coverage in the donut hole, and cost sharing. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results, with and without the interaction of observable characteristics, are presented in Table 3 columns 4 and 5. The coefficient on premium interacted with the within-individual stakes remains positive and statistically significant in both specifications, although smaller in magnitude. We also examine the results using the perfect foresight measure of out-of-pocket costs and obtain qualitatively similar results.[20]

There is concern that individuals with high stakes may be older and have more experience choosing plans, leading to better choices for reasons other than the incentive to research plans. In column 6 of Table 3, we estimate separate coefficients when the stakes increase versus decrease and find qualitatively similar results. Addi-

[20]The detailed results using the perfect foresight measure of out-of-pocket costs are available in Online Appendix Table A-5.

tionally, we show that first-time enrollees have a similar relationship between stakes and choice quality as the rest of our sample (see Table A-4). These findings suggest that our results are not solely driven by individuals gaining more experience or stakes increasing over time as individuals age.

The descriptive evidence above implies that there is a relationship between the stakes and choice quality. This relationship holds when controlling for individual fixed effects. While time-varying unobservables may play a role, they are unlikely to fully explain the results. For these reasons, the evidence suggests that the relationship between stakes and choices is at least due in part to the fact that individuals respond to incentives to acquire information. This motivates us to estimate a model incorporating endogenous information. The empirical model allows us to quantify the welfare effects of costly information and provide insight into the implications for simplifying choice.

4 Empirical Model and Estimation

In this section, we develop and estimate an empirical model of insurance demand with endogenous information acquisition based on the framework in Section 2. The model seeks to identify the marginal cost of information in addition to preferences.

Individual i chooses plan $j \in \mathcal{J}_{it}$ in year t where the choice set is defined by $\mathcal{J}_{it}$. Consider expected utility if individuals used all available information to form expectations about the cost of each plan. Following the previous literature, we assume that utility follows from an approximation to a CARA utility function, implying that the certainty equivalent expected cost of a plan can be expressed as $p_{jt} + v_{ijt} + \frac{1}{2}\gamma^2\widetilde{\sigma}^2_{ijt}$ where γ is the coefficient of risk aversion.[21] As in the previous section, p_{jt} is the annual premium, v_{ijt} is the expected annual out-of-pocket cost, and $\widetilde{\sigma}^2_{ijt}$ is a measure of plan risk. Adding preferences over non-cost characteristics, indirect utility can be

[21] Consider CARA utility, $-\exp(-\gamma(W - C_{ijt}))$, where individuals have wealth W and plan cost is distributed $C_{ijt} \sim N(p_{jt} + v_{ijt}, \widetilde{\sigma}^2_{ijt})$. Expected utility can be expressed as $u(\mu, \widetilde{\sigma}^2) = -\tau \exp\left(\gamma\mu + \frac{1}{2}\gamma^2\widetilde{\sigma}^2\right)$ where $\tau = -exp(\gamma W)$. Using a first-order Taylor expansion around $(\mu', \widetilde{\sigma}^{2\prime})$ and dropping constant terms, $u(\mu, \widetilde{\sigma}^2) \approx u(\mu', \widetilde{\sigma}^{2\prime}) - \tau\gamma u(\mu', \widetilde{\sigma}^{2\prime})\mu - \frac{1}{2}\tau\gamma^2 u(\mu', \widetilde{\sigma}^{2\prime})\widetilde{\sigma}^2$. See Abaluck and Gruber (2011) and Heiss et al. (2013).

written as

$$u_{ijt} = \underbrace{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma}^2_{ijt}}_{\text{Initially Unknown}} + \underbrace{\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt}}_{\text{Known}}. \tag{13}$$

In a similar way as Section 2, individuals initially observe premium, p_{jt}, and other known characteristics. The magnitude of parameter α_i can be interpreted as the marginal utility per dollar when individuals are fully informed. Individuals whose previous plan is no longer available may have a preference for the same insurer, and therefore we include an indicator for previous insurer as a known characteristics, X^k_{jt}. We also include insurer by chronic condition fixed effects, $\zeta_{b(j)d(it)}$, where $b(j)$ represents the function mapping each plan j to the insurer and $d(it)$ represents the function mapping each individual i at time t to a major diagnosis. The insurer fixed effects capture quality differences between insurers observed by enrollees but unobserved by the researcher.[22] We interact the insurer fixed effects with an indicator for each of the most common diagnoses. In particular, we include separate insurer fixed effects for individuals diagnosed with diabetes, chronic kidney disease, congestive heart failure, and other chronic diagnosis. In this way, unobserved preferences are allowed to differ for individuals with different chronic conditions. Finally, individuals have a known idiosyncratic taste shock, ϵ_{ijt}. The taste shock is important for capturing unobserved preferences which could provide an alternative explanation for why some individuals choose plans that appear dominated.

Given known plan characteristics and individuals' prior, individuals choose how much information to acquire about the initially unknown portion of utility for each option. The unobserved portion of utility, ξ_{ijt}, includes expected out-of-pocket costs, v_{ijt} and the riskiness of the plan, $\widetilde{\sigma}^2_{ijt}$, both of which require researching contract terms and potential health shocks. We also include plan quality, as measured by star ratings, as an additional plan characteristics that is initially unknown, X^u_{jt}. As in Section 2, individuals maximize expected utility after information acquisition. Given the distributional assumptions on the prior and taste shock in Section 2, choice prob-

[22]We group insurers with less than 1% market share into a single category given that it is difficult to estimate a separate fixed effect.

abilities after information acquisition take a closed-form following equation (6). After choosing a plan, health shocks and consumption are realized.

The model requires assumptions about the mean and variance of individuals' priors. We construct a measure of an individual's prior variance, σ^2_{it}, using the variance in the unknown portion of the utility across options in the individual's choice set, $Var_j\left[\xi_{ijt}\right]$. In this way, the prior variance depends on the variance in unknown plan characteristics, including expected out-of-pocket costs, and preferences for those characteristics. The prior variance is common to all options in the choice set. In the baseline specification, we also assume that an individual has a common prior mean across options in the choice set. This is motivated by the fact that the sample is limited to individuals that were not previously enrolled in any of the plans and therefore are less likely to start with any information about specific plans. Given a prior mean that is common across options, ξ^0_{ijt} can be normalized to zero for every option.[23] We also consider specifications in which individuals start with additional information about plans, i.e. allow for a heterogeneous prior across choices. We discuss these specifications in detail in Section 5.4.

We now describe the specific assumptions we make regarding heterogeneity across individuals in the price coefficient, α_i, and the marginal cost of information, λ_{it}. We allow for observable heterogeneity in price sensitivity by assuming

$$\alpha_i = -\exp(\beta^{\alpha} Z_i) \tag{14}$$

where Z_i are time-invariant individual characteristics (including a constant). Similarly, we also allow for heterogeneity in the marginal cost of information by assuming

$$\lambda_{it} = \exp(\beta^{\lambda 1} Z_i + \beta^{\lambda 2} W_{it}) \tag{15}$$

where W_{it} are time-varying characteristics including the individual's health status and experience with Medicare Part D.[24]

[23]Since choice probabilities only depend on differences in expected utility and there is no outside option, the normalization of the prior is inconsequential.

[24]Although λ_{it} varies across individuals, we assume that it is common to all options in an individual's choice set. This is consistent with the fact that Medicare Part D plans all have similar benefits

To summarize our key assumptions, as in Matějka and McKay (2015) we adopt an entropy-based cost function. In addition, we assume that individual preferences are approximated by a CARA utility function. The prior over the initially unobserved component of utility, ξ_{ij}, and the taste shock, ϵ_{ij}, follow the Cardell distribution. The prior variance is common to all options in the choice set.

The estimation strategy is straight-forward. Given that we derive closed-form choice probabilities, we employ maximum likelihood. The likelihood function is similar to the standard likelihood function for a multinomial logit; however the parameter vector β^{λ} enters representative utility non-linearly. The log-likelihood function is reported in Online Appendix A-1.

4.1 Alternative Models without Endogenous Information

We compare the results of the endogenous information model to three alternative demand models. As a benchmark, we estimate a *standard logit model* assuming that individuals have full information about both premiums and expected out-of-pocket cost, thus putting equal weights on the two objects. Next, we estimate a model in which demand is a function of premium and coverage characteristics, such as the deductible, rather than expected out-of-pocket cost. This approach is widely used in the literature estimating demand for insurance.[25] We call this model the *coverage characteristics model*. Finally, we estimate a *differential weight model* that allows different coefficients on premium and expected out-of-pocket cost. This approach has been previously applied in the context of Medicare Part D.[26] The details of these alternative models and parameter estimates are presented in Online Appendix C.

4.2 Welfare

With costly information acquisition, individuals choose plans that maximize expected utility given beliefs after information acquisition, but do not necessarily maximize ex-

designs, making them equally complicated.

[25]This general approach has been used by Bundorf et al. (2012), Handel (2013), Polyakova (2016), and others.

[26]See, for example, Abaluck and Gruber (2011) and Abaluck and Gruber (2016). Ho et al. (2017) also uses a related approach.

pected utility given rational expectations. Our model decomposes the idiosyncratic error term into a taste shock and an informational error which can be seen in equation (A-11). Since part of the idiosyncratic error is due to limited information rather than a taste shock in our model, eliminating products with the same representative utility will result in a smaller welfare loss, and can even result in a welfare increase. Welfare must take into account the fact that there is a difference between the expected utility anticipated at the time of decision-making and utility with rational expectations, leading to choices that are incorrect ex-post. In addition, total welfare should account for individual's information acquisition cost.

Consumer surplus with endogenous information for individual i in year t is given by

$$CS_{it}^{RI} = \frac{1}{-\alpha_i} \log \sum_j e^{\tilde{\nu}_{ijt}} + \frac{1}{-\alpha_i} \sum_j P_{ijt}[\nu_{ijt} - \tilde{\nu}_{ijt}] \tag{16}$$

where rational expectations utility (with full information) excluding the iid shock is

$$\nu_{ijt} = \alpha_i v_{ijt} + \beta_1 X_{jt}^u + \beta_2 \widetilde{\sigma}_{ijt}^2 + \alpha_i p_{jt} + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} \tag{17}$$

and the expected utility after information acquisition excluding the iid shock is

$$\tilde{\nu}_{ijt} = a(\sigma_{it}, \lambda_{it}) \left(\alpha_i v_{ijt} + \beta_1 X_{jt}^u + \beta_2 \widetilde{\sigma}_{ijt}^2\right) + b(\sigma_{it}, \lambda_{it}) \left(\alpha_i p_{jt} + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)}\right). \tag{18}$$

The first term in equation (16) is the expected welfare calculated as if beliefs after information acquisition were correct. Note that $-\alpha_i$ is the marginal utility of income. The second term adjusts for the fact that there may be a difference between expected utility after information acquisition and full-information utility. This term is the weighted average of the difference between expected consumer surplus after chosen information acquisition and consumer surplus evaluated with full information where the weights are the probability of choosing each option.[27] The welfare loss due to

[27] P_{ijt} is given in equation (A-17). Note equation (16) follows from Train (2015) who consider welfare in discrete choice models when beliefs, $\tilde{\nu}_{ijt}$, and actual utility differ and beliefs are observed. Further detail is provided in Online Appendix A-4.

information frictions is then given by

$$\Delta CS_{it} = CS_{it}^{FullInfo} - CS_{it}^{RI} + \hat{C}_{it} \tag{19}$$

where $CS_{it}^{FullInfo}$ is consumer surplus under full information given by $\frac{1}{-\alpha_i} \log \sum_j e^{\nu_{ijt}}$. The total cost of information, $\hat{C}_{it}$, is determined by the mutual information following the assumptions of the rational inattention model and is given by equation (A-20).

4.3 Identification

In many settings, an individual that does not understand a product characteristic may make similar choices as an individual that does not care about a product characteristic. Therefore, our primary identification concern is separately identifying preference parameters, including the coefficients on the price and other product characteristics, separately from the marginal cost of acquiring information.

For identification, we leverage the fact that individuals can observe premiums but do not initially observe out-of-pocket costs unless they acquire costly information. Given our assumptions about risk preferences and taste shocks, choice probabilities after information acquisition imply that individuals put weight $b(\sigma_{it}, \lambda_{it})\alpha_i$ on premiums and weight $a(\sigma_{it}, \lambda_{it})\alpha_i$ on expected out-of-pocket cost. When information is free, these weights are equal and an individual will be indifferent between a marginal charge in premiums and a marginal charge in expected out-of-pocket costs. If observed choices are equally sensitive to premiums and out-of-pocket costs, then we conclude that there are no information frictions.

In contrast, when information is costly, $\frac{a(\sigma_{it},\lambda_{it})}{b(\sigma_{it},\lambda_{it})} = \frac{\ell_{it}(\sigma_{it},\lambda_{it})-1}{\ell_{it}(\sigma_{it},\lambda_{it})} < 1$ and individuals appear to be less sensitive to variation in expected out-of-pocket cost, with the ratio depending both on the unit cost of information and the stakes. Therefore, identification requires variation in premium and variation in expected out-of-pocket cost in order to identify a separate coefficient on both. Given an assumption about the prior variance, the ratio of these coefficients identifies the marginal cost of information. In particular, by estimating coefficients on both premium and expected out-of-pocket costs and allowing the coefficients to vary with the stakes and other individual characteristics in a flexible way, one can pin down the marginal cost of information and how it varies with demographics.[28]

In this way, the identification of our model is related to previous work estimating preference parameters for premium and out-of-pocket cost in insurance markets.[29] The coefficient on expected out-of-pocket cost, as well as the coefficient on risk, is identified by variation across plans, time, and markets within the same insurer given the inclusion of insurer fixed-effects. Much of this variation in the premium and out-of-pocket costs is due to the fact that insurers offer a menu of plans with different benefits within the same market. Contracts also vary across time due, in part, to policy changes in minimum standards imposed by CMS. Insurers charge different premiums for each plan they offer within a market, and premiums often differ across markets for the same plan. Variation in health risks results in further variation in out-of-pocket costs across individuals even for the same plan. In the model with costly information, we also wish to identify how coefficients interact with the stakes. Variation in health status and choice sets, including variation for the same individual across time, leads to variation in the stakes.

As in a standard discrete-choice model, endogeneity issues are a potential concern if a plan's premium is correlated with plan characteristics that we do not observe but are valued by enrollees. However, as discussed by prior work examining the Medicare Part D market (e.g. Ho et al. 2017), the institutional features of the market considerably reduce concerns about endogeneity given that differentiation among plans is limited to specific dimensions. Insurers submit plans to CMS which ensurers that plan benefits meet minimum actuarial standards. Plans may offer contracts that exceed those minimum standards, generating variation in benefits across plans. Our measure of an individual's expected out-of-pocket cost for each plan is determined by the plan's drug contract terms, including deductible, donut hole coverage, and formulary. Concern about measurement error in this out-of-pocket cost measure is mitigated by the fact that we also include insurer fixed effects. Since insurers often use the same formulary across plans, the insurer fixed effects help capture any insurer

[28]Online Appendix F presents a more formal discussion of identification based on this argument.
[29]See, for instance, discussion in Ho et al. (2017) and Polyakova (2016).

benefits that are not reflected in the out-of-cost measure.

There may also be differences in non-pecuniary characteristics across plans, such as customer service. Insurer fixed effects absorb variation in these non-pecuniary benefits. We also allow insurer fixed effects to vary by health status to address concerns that those with chronic conditions may have different preferences for non-pecuniary characteristics of plans.

While we take a number of steps to address potential endogeneity issues, it is still possible that unobserved plan quality is correlated with premiums, in which case the coefficient on the premium would be biased toward zero. This would also mean that λ_{it}, which determined by the ratio between the coefficients on the premium and the out-of-pocket cost, is underestimated. This would imply that our measure of the welfare losses from information frictions is also an underestimate.

4.4 Empirical Model Estimates and Fit

The parameter estimates from the baseline demand model are presented in Table 4. Average price sensitivity, α_i, is estimated to be -0.12. The coefficient on income is negative indicating that individuals in high income zip codes are less price sensitive, however the estimate is not statistically significant. As expected, individuals also have a strong preference for the previously chosen insurer while preferring less risk and higher star ratings.

The average marginal cost of information, which converts bits of information to utils, is estimated to be 2.5, although there is a large degree of heterogeneity.[30] The marginal cost of information may reflect either an individual's mental difficulty in comparing plans or the opportunity cost of time. In addition, many older Medicare patients may receive help from family, nursing home staff, or others. In this case, the estimated marginal cost of information would apply to the decision maker in question.

Individuals in more educated zip codes have lower marginal cost of information, consistent with the idea that it is easier for more educated individuals to research plans. However, this parameter is not statistically significant. Older individuals

[30]Note that cost coefficient is in hundreds of dollars since premium and out-of-pocket costs are scaled for estimation.

Table 4
Estimates for Demand Model with Endogenous Information Acquisition

	Estimate	SE
Price Sensitivity (β^{α})		
Constant	−2.1368	(0.0207)
Income	−0.0008	(0.0005)
Other Plan Characteristics		
Previous insurer	6.4433	(0.0662)
Risk	−0.0464	(0.0029)
Star rating	1.6125	(0.1181)
Marginal cost of information (β^{λ})		
Constant	2.9742	(0.1852)
Zip Income	−0.0004	(0.0010)
Zip Education	−0.0008	(0.0023)
Age	0.5721	(0.0955)
Age^2	−0.0034	(0.0006)
Female	0.0141	(0.0435)
Part D Experience	−0.4243	(0.0392)
Rural	0.2281	(0.0574)
Has alzheimers	0.0823	(0.0710)
Has lung disease	0.1504	(0.0687)
Has kidney disease	−0.0985	(0.0568)
Has heart failure	0.0722	(0.0618)
Has depression	−0.0177	(0.0636)
Has diabetes	0.0523	(0.0502)
Has other chronic condition	0.0022	(0.0499)
Mean price sensitivity	−0.1181	
Mean marginal cost of information	2.5425	
LL	-50,468.22	
Observations	1,035,319	

Notes: Shows MLE estimates from demand model with endogenous information. Premium and out-of-pocket cost are in hundreds of dollars. Continuous individual characteristics (income, education, age, and age squared) are demeaned. Specification includes insurer by chronic condition fixed effects. Standard errors in parentheses.

may have more difficulty researching plans. The coefficient on age is positive and highly significant; however, the coefficient on age squared is negative. This implies the marginal information cost is increasing in age for individuals age 65 to 84 before slightly declining, perhaps due to the fact that the oldest individuals may receive help researching plans from others. Overall, the standard deviation of the marginal cost of information is 2.1, quite large relative to the mean. Along with the variation in the stakes, this implies large differences in the total realized cost of information acquisition across individuals.

Table 5 shows actual mean premium and out-of-pocket costs for individuals' chosen plans versus the mean cost for plans chosen in the simulated baseline. The fit is quite good. The model predicts that individuals choose plans with average out-of-pocket cost of $713 while the actual mean is $719. For premiums, it is $570 and $566, respectively. In addition, the model is able to accurately rationalize the difference between the cost of the chosen option and the plan with the lowest total cost. In contrast, the standard demand model cannot rationalize why individuals choose plans with low premiums and high out-of-pocket costs. This can be seen in the second column of Table 5. Although the standard logit model accurately predicts the total cost, the out-of-pocket cost and premium both differ by over $50.

We also evaluate model fit by using the baseline specification to simulate the probability of choosing the lowest cost plan and the weight that individuals appear to place on premium and out-of-pocket costs as a function of the stakes. We confirm that the results can match the patterns presented in the descriptive analysis in Figures 2 and 3. We also use the estimates from each of the three alternative models that do not allow for endogenous information. All of the alternative models have difficulty rationalizing how choice quality changes when the stakes change.[31]

5 Counterfactual Results

In this section, we explore counterfactual demand under full information, restricted choice sets, and a cap on out-of-pocket cost. The results highlight the role of endogenous information and implications for proposed policies aimed at simplifying choice.

5.1 Full Information Counterfactual

We start by simulating insurance demand under full information in order to shed light on the welfare effects of information acquisition costs in Medicare Part D. This counterfactual can be viewed as scaling up information intervention to the limit. The results, presented in Table 5, indicate that the welfare effects of costly information are substantial. Under full information, individuals would choose plans with out-of-pocket costs that are $112 lower; however these plans have premiums that are $72

[31] The detailed evaluation of the model fit is presented in Figure A-1.

higher. Given that individuals on average choose a plan that is $565 more expensive than the least expensive option, this suggests that individuals have strong preferences over non-cost characteristics. Our analysis focuses on active switchers that are not low-income and care should be taken generalizing these results to the full population of enrollees. However, a simple back-of-the-envelope calculation assuming that these results apply to all enrollees implies that, holding premiums and out-of-pocket costs fixed, removing information frictions would result in total savings of $376 million per year.[32]

Under full information, individuals would choose plans with higher quality and lower risk. Overall, this implies that welfare, excluding information acquisition costs, increases by $285 per enrollee on average. Information acquisition costs are also substantial, with an average of $127. Kling et al. (2012) find that Part D beneficiaries on average spend three hours on plan consideration in their 2007 survey. Bundorf et al. (2019) find that 75 percent of individuals spend more than one hour choosing their Part D plan. We think our information cost estimates are reasonable given that information acquisition in our model encompasses not only researching and choosing plans, but also researching health risks (e.g. drugs that may be needed in the future) and insurance terminology (e.g. definitions of donut hole coverage and deductibles). Relatedly, Kling et al. (2012) report that simply making information available and free to individuals through the Plan Finder does not lead them to use it, potentially because of costs associated with understanding the information. This is consistent with the high level of information cost that we estimate.

When calculating welfare, we make the standard assumption that the taste shock contributes to welfare, implying a mechanical welfare gain from a large number of plans. In order to examine the role of the taste shock, we also calculate the welfare effects excluding the taste shock.[33] As seen in Table 5, the implied welfare gains of full information are even larger when the taste shock is excluded.

In the baseline case, the estimated elasticity of demand with respect to premiums

[32] Average enrollment of individuals in stand-alone Part D plans that do not receive the low-income subsidies is 9.4 million per year over the sample. The calculation includes individuals who do not make an active choice and assumes savings also apply to individuals who remain enrolled in their previous plan.

[33] For this exercise, we define welfare as $CS_{it} = \sum_j P_{ijt} v_{ijt}$.

Table 5
Counterfactual Spending and Welfare Under Full Information

			Endogenous Information Model			
		Standard	All Individual		Individuals w/ High Information Cost	
	Actual	Model	Baseline	Full Info	Baseline	Full Info
Out-of-pocket cost of chosen plan	719	651	713	601	862	701
Premium of chosen plan	566	634	570	642	558	621
Total cost of chosen plan	1285	1285	1282	1244	1420	1323
Cost difference compared to lowest cost plan	565	576	569	538	686	605
Δ welfare ex. info acquisition cost				285		201
Δ info acquisition cost				127		300
Δ welfare ex. info acquisition cost (no taste shock)				379		331
Out-of-pocket Elasticity			-0.35	-1.59	-0.69	-1.81
Premium Elasticity			-1.51	-1.59	-1.66	-1.81

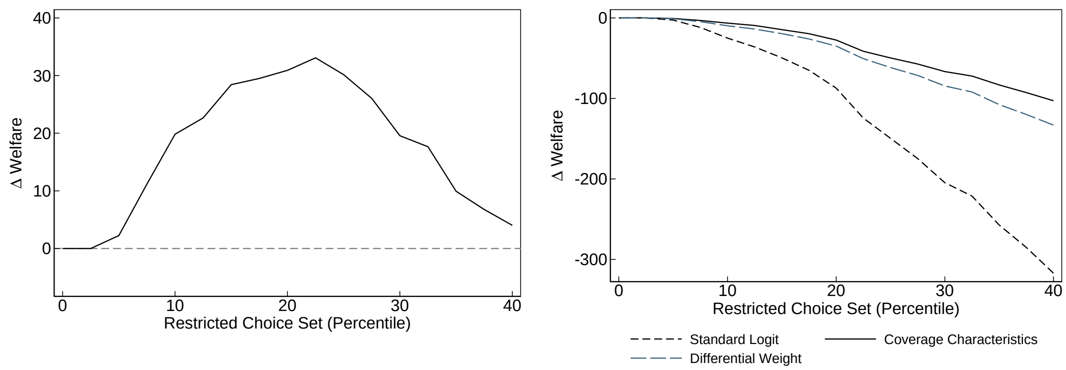
Notes: Shows counterfactual simulations for endogenous information model using parameter estimates from Table 4. Full information counterfactual assumes individuals know expected out-of-pocket cost and other initially unobserved characteristics of each plan. Individuals with high information cost defined as those with total cost of information, $\hat{C}_{it}$, in the top quartile. Standard demand refers to multinomial logit specification described in Appendix Section C.

is -1.5, however the elasticity with respect to expected out-of-pocket costs is only -0.4.[34] Elasticity with respect to premium (out-of-pocket cost) can be interpreted as the percent change in demand from a 1 percent change in cost due to premiums (out-of-pocket costs). The large difference in elasticities reflects the importance of costly information. Under full information, the elasticity of demand is -1.6, the same for both premiums and expected out-of-pocket costs.

Table 5 also shows the results for individuals with the total incurred information cost, $\hat{C}_{it}$, in the top quartile. These individuals may face higher stakes and therefore have more incentive to acquire information, or have higher marginal costs of acquiring information. For these individuals, the total cost saving is $97 in the full information case. Although the welfare effects excluding information acquisition costs are lower than for the population as a whole, the information acquisition costs are more than double. Under full information, their demand is quite elastic, about -1.8.

[34]For comparison, Abaluck and Gruber (2011) report an average elasticity with respect to premium ranging from -0.75 to -1.17.

Figure 4
Counterfactual Welfare Effects of Restricted Choice Set



a. Endogenous Information Model b. Alternative Models

Notes: Chart shows counterfactual average change in welfare per enrollee from removing plans with mean utility below a given percentile where average utlity is computed by plan, year, and age. Counterfactual estimates from model with endogenous information acquisition are contrasted with counterfactual welfare estimates from alternative models of plan demand without endogenous information (see Online Appendix Section C).

5.2 Restricted Plan Choice Counterfactual

We use the model to examine the implications for restricting plan choice. In the Medicare Part D market, many individuals can choose between over 35 plans. The large number of options may make it difficult to research plans and choose correctly. We ask whether welfare can be increased by showing individuals only a subset of the plans based on their age, thus restricting the choice set.[35] For each plan we calculate average utility for each year and enrollee age. We then simulate choices and calculate welfare after removing plans with average utility below a given percentile. We assume that individuals are aware that "poor" plans are removed, thus affecting their incentive to research plans.[36]

The change in welfare from restricting the choice set accounting for endogenous information is depicted in Figure 4 Panel a. As seen in the figure, there is a trade-off.

[35]To a certain extent, insurance regulators already do this through allocation policies that set minimum standards for plans. This is also related to standardization of health exchanges (Ericson and Starc 2016).

[36]Formally, σ_{it}^2 and $\hat{C}_{it}$ are recomputed for each counterfactual simulation.

On the one hand, simplifying the choice set can reduce the chance that individuals accidentally choose poor plans as well as reduce information costs. However, restricting the choice set too much does not allow individuals with heterogeneous preferences to find a plan that is a good fit. For this reason, welfare is increasing until about a quarter of plans are removed from individuals' choice sets. When too many plans are removed, welfare decreases. The result that reducing the size of the choice set can increase welfare at the margin is consistent with a number of surveys of Medicare Part D enrollees indicating that individuals would prefer to see a smaller set of recommended options.

The counterfactual results examining restricted plan choice are summarized in Table 6. Eliminating plans in the lowest quartile results in individuals choosing plans that have better non-cost characteristics as well as slightly lower cost, resulting in welfare gains of $12 per individual. In addition, individuals face lower stakes and therefore choose to acquire less information, resulting in total information acquisition costs that are $43 lower. Therefore, the benefit of providing consumers with a tailored choice set is primary due to the reduction in research in this case. Removing plans in the bottom 10th percentile leads to smaller welfare gains of $22 including the reduction in information cost.

These results can be contrasted with alternative models that do not account for endogenous information (Figure 4 Panel b). In all of these models, restricting the choice set implies a welfare reduction, the opposite of what is implied by the endogenous information model. This is because, in these alternative models, the failure of individuals to choose plans with low out-of-pocket costs is rationalized through heterogeneous preferences rather than information frictions. The effect on spending and welfare is detailed in Table A-2.

The welfare gains of restricting choices are even larger if a social planner can provide a personalized list of plans to each individuals.[37] Under this scenario, welfare is maximized when more than 75% of plans are removed. Results from this counterfactual experiment are available in Section E of the Online Appendix.

[37] I.e., if the social planner restricts the choice set based on representative utility given by $\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma}^2_{ijt} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}$.

Table 6
Counterfactual Spending and Welfare for Restricted Choice Set and Out-of-Pocket Cap

	Restricted Choice Set		Out-of-Pocket Cap	
	10th Percentile Cutoff	25th Percentile Cutoff	\$5,000 Cap	\$15,000 Cap
Δ Premium	-1.5	-4.0	-9.6	-6.8
Δ Out-of-pocket cost	1.2	3.2	-374.3	-180.8
Δ Spending	-0.3	-0.9	-384.0	-187.6
Δ Welfare ex. info	2.7	11.5	374.5	182.6
Δ Information cost	-19.6	-43.0	-11.5	-9.9
Δ Welfare ex. info (no taste shock)	11.6	43.9	380.1	185.5

Notes: Columns 1 and 2 show counterfactual simulations in which the set of plans offered is restricted by eliminating plans with mean utility below the 10th/25th percentile. Columns 3 and 4 show counterfactual simulations in which out-of-pocket costs are capped at \$5,000/\$10,000.

In addition to restricting the choice set, we note that there are other policies that could steer consumers away from the largest mistakes while still allowing individuals sufficient choice given idiosyncratic preferences. For instance, consumers could be shown a suggested set of plans or be given a targeted default. Like restricting the choice set, these policies could also potentially increase welfare when information is costly.

5.3 Out-of-Pocket Cost Cap Counterfactual

In order to examine how cost sharing interacts with endogenous information acquisition, we examine counterfactual simulations in which we impose a cap on out-of-pocket payments. This policy has been proposed for Medicare Part D and has already been implemented in other health insurance settings. Currently, Medicare Part D enrollees who have out-of-pocket costs above the catastrophic threshold can still be liable for substantial costs.[38] Imposing an out-of-pocket cap not only makes it less likely for individuals to accidentally choose an expensive plan, but also reduces the variance in out-of-pocket costs across plans, reducing the stakes as in the previous counterfactual.[39]

[38] As of 2019, the catastrophic threshold is \$5,100. Once enrollees have drug costs above the catastrophic threshold, they pay either 5 percent of total drug costs or \$3.40 (\$8.50) for each generic (brand name) drug.

[39] Note, however, that our analysis does not take into account potential changes in utilization and negotiated drug prices.

We find that endogenous information model implies higher welfare gains from capping out-of-pocket costs than other models, especially the differential weight model.[40] While a cap on out-of-pocket costs has a direct benefit for consumer by reducing cost, there are two additional reasons why the policy generates additional welfare gains in the presence of endogenous information frictions. First, individuals are less likely to "accidentally" choose a plan with high out-of-pocket costs when the cap is binding. Second, imposing the out-of-pocket cap also substantially reduces information acquisition costs. In other words, there is less risk of choosing a plan with very high out-of-pocket costs, individuals conduct less research. This implies that the welfare gains from an out-of-pocket cap accrue, in part, to individuals with spending below the cap.

The counterfactual effect of a cap on out-of-pocket cost is summarized in Table 6. Imposing a \$15,000 cap generates an increase in welfare of \$193 per enrollee after accounting for the change in information cost. Imposing a \$5,000 cap generates welfare gains of \$386.

Capping out-of-pocket costs would imply less revenue for insurers. While we do not model insurer premiums, we also consider a simple policy in which the decrease in out-of-pocket cost for individuals above the cap is offset with an increase in premiums that is the same for all plans, making the policy revenue-neutral. We find that under the endogenous information model this policy is still welfare increasing, while alternative models imply a decrease in welfare.[41]

Overall, these results highlight that a cap on out-of-pocket costs can help mitigate the welfare costs due to information frictions. More generally, evaluation of cost sharing policies should take into account the effect on the incentive to research plans and implications for consumer accidentally choosing plans with high out-of-pocket cost.

[40] See Figure A-5 for detailed results.

[41] See Figure A-6 in the Online Appendix for detailed results.

5.4 Discussion and Robustness

In this section, we discuss implications of our modeling assumptions and summarize robustness results.

Heterogeneous Preferences

Our baseline specification includes insurer fixed effects interacted with major chronic conditions to flexibly capture unobserved preferences. As a robustness exercise, we examine results without insurer fixed effects and their interactions with chronic conditions. We estimate that the mean elasticity is nearly the same as our main specification. The mean marginal cost of information is also very close at 2.3 (compared to 2.5 in the main specification). This implies that unobserved preferences for insurers are not driving the estimate after controlling for observable plan characteristics. We also estimate a specification that includes the outliers with very high stakes that are removed from the main sample and also find that the estimates for the information cost and preference parameters are similar to our baseline estimates.[42] While this suggests that the model is capturing the main sources of preference heterogeneity, it is possible that more complicated unobserved preference heterogeneity could affect the information cost estimates.[43]

Priors and Stakes

Another key assumption of the model is the form of individuals' prior since it determines information acquisition. In our baseline model, we assume that individuals start with a prior that out-of-pocket cost is uncorrelated with the premium of a plan, implying a common prior about the initially unobserved part of utility across options. One concern is that individuals believe high premium plans are more likely to result in low OOP cost prior to doing research.

We consider three alternative assumptions on the prior to examine sensitivity to

[42]The full set of estimates is available in Table A-6 in the Online Appendix.

[43]A related concern is that some individuals may be liquidity constrained and this could affect relative demand for premiums and out-of-pocket costs. However, the timing of premium payments and out-of-pocket payments is similar.

Table 7
Estimates for Demand Model with Endogenous Information Acquisition
Robustness to Alternative Definition of Prior

	(1)		(2)		(3)	
			Prior Variance Based on Group Average			
	Heterogenous Prior Mean		Homogenous Prior Mean		Heterogeneous Prior Mean Based on Observed Attributes	
	Estimate	SE	Estimate	SE	Estimate	SE
Price Sensitivity (β^{α})						
Constant	−2.0318	(0.0199)	−2.0053	(0.0198)	−1.9202	(0.0210)
Income	−0.0001	(0.0005)	−0.0006	(0.0005)	0.0002	(0.0005)
Other Plan Characteristics						
Previous insurer	6.5732	(0.0618)	6.9208	(0.0798)	6.9859	(0.0817)
Risk	−0.0594	(0.0029)	−0.0656	(0.0035)	−0.0721	(0.0039)
Star rating	2.0222	(0.1017)	2.7167	(0.1429)	2.9856	(0.1535)
Mean price sensitivity	−0.1311		−0.1346		−0.1466	
Mean marginal cost of information	3.2413		3.1758		3.7474	
Full Information Counterfactual						
OOP	594		596		593	
Premium	646		651		650	
LL	-50,735.50		-50,151.07		-50,410.02	
Observations	1,035,319		1,035,319		1,035,319	

Notes: Shows estimates from demand model with endogenous information with alternative assumptions about the prior. In Specification 1, prior variance is defined as the average variance in the individual's choice set, as in the baseline specification. However, prior mean is determined by population average over plan by year. In Specification 2 and 3, prior variance is defined as the average variance for similar individuals. For Specification 3, we regress out-of-pocket cost on plan premium, deductible, generic coverage, coverage in the gap, and cost sharing. We then assume individuals understand this relationship, informing their prior. Premium and out-of-pocket cost are in hundreds of dollars. Continuous individual characteristics (income, education, age, and age squared) are demeaned. Standard errors in parentheses.

this issue. In our first alternative specification, we allow heterogeneous prior means for options by assuming that individuals initially know the average cost of each plan across individuals that are similar to them but not their own out-of-pocket cost. In the second and third alternative specifications, we explore an alternative measure of the stakes. Instead of assuming that individuals know the variance of the unknown part of the utilities in their own choice set as in our baseline, we assume individuals initially know the average variance for similar individuals. In the second specification, the prior mean is assumed to be homogenous. In the third specification, we assume that individuals use easily observable characteristics—plan premium, deductible, generic coverage, coverage in the gap, and cost sharing—to predict out-of-pocket cost for each plan, while maintaining the assumption that the prior variance is given by the group average.

Table 7 presents the summary of the estimates from these alternative specifications. We find that the mean price sensitivity for these alternative specifications is slightly higher than the baseline specification. The marginal cost of information is also higher, especially when we allow prior means to depend on observed characteristics in specification (3). This is reasonable given that individuals are assumed to have more information before paying an information acquisition cost in this specification. Nevertheless, even under different assumptions about priors, counterfactual choices under full information in terms of the average costs of chosen options are almost identical to the baseline case. Given that the alternative specifications produce marginal cost estimates that are larger than our baseline estimates, we interpret our results as providing lower bounds for the welfare cost of information frictions and the effects of the policies that we consider.

It is also possible that individuals have a biased prior. While it is difficult to identify a biased prior in our setting, future work could use additional data sources or surveys to provide more insight into the nature of individuals' priors and the implications for information acquisition. In addition, there is concern that research about one plan could inform an individual about other similar plans. This could be accommodated in the model by allowing individuals to have a more complicated prior in which unobserved utility is correlated across plans. While this is beyond the scope of our paper, future work could also explore this issue.

Moral Hazard

Following Abaluck and Gruber (2016), we assume that individuals could substitute to the cheapest alternative within each class and Generic Code Number (a classification defined by ingredients, strength, dosage, and route of administration), but assume no other forms of moral hazard.[44] This assumption is relevant for at least three reasons. First, it allows us to calculate what each individual's realized out-of-pocket costs would have been for each plan in her choice set following Abaluck and Gruber

[44] Ho et al. (2017) also assume no moral hazard in a similar setting. Studies that focus on the effect of the introduction of Part D on drug utilization generally find a minimal to modest effect (see, for example, Basu et al. (2010)). The more relevant dimension of moral hazard for our counterfactuals is how individuals would adjust drug usage if they chose a different plan than observed in the data. To our knowledge, there is limited work examining this. An exception is Abaluck and Gruber (2011) who argue that this effect is small.

(2016). Second, we are able to hold drug consumption constant with individuals' plan choice in our counterfactual simulations. To the extent that individuals adjust their drug spending with their plan choice, our estimate of the welfare effect of the out-of-pocket cost cap would be an underestimate. As the cap directs individuals to a more generous plan, they might increase drug consumption, further adding to the welfare gains from the cap and affecting insurer costs. Given that typical estimates of the elasticity of demand for prescription drugs are low, we believe this effect would be relatively modest.[45]

6 Conclusion

We develop a micro-founded yet tractable demand model based on the rational inattention framework that can be applied to settings in which some product attributes are costly to observe or understand. Consistent with the model, we find evidence that individuals acquire more information as the stakes increase in the market for prescription drug insurance.

We propose and implement a feasible estimation strategy for our empirical model. Estimates imply that the welfare effects of information frictions are substantial. Among policy makers, there is a concern about the complexity of insurance choice and how to regulate plan features. Standard demand models provide little insight into how markets for complex products can be simplified or standardized. With this in mind, we use the model to examine how insurance regulation affects information acquisition and welfare. We find that simplifying insurance choice through restricting available products or capping out-of-pocket costs can improve welfare given the difficulty in researching plans.

Our empirical model of endogenous information frictions can be applied to other markets in which there are complex characteristics that are costly to research. A key requirement for the identification of our model is that there are product attributes that consumers value equally under full information but consumers initially observe only a subset of those attributes. For example, this is the case with various fees

[45] The typical estimates of the elasticity of demand for prescription drug fall in the range of -0.1 to -0.6.

that are difficult to observe in the market for tickets, food delivery, mortgages, and other financial products. Alternatively, variation in the marginal cost of information could be used to identify the model in other settings featuring characteristics that are costly to observe, such as the nutritional content of food or school quality. The model can potentially rationalize choice inconsistencies, choice overload, and consumer inertia that might arise in these settings. Furthermore, it can inform how consumer protection laws should be designed in these markets by, for instance, regulating or standardizing product offerings.

An important caveat of the analysis is that we focus only on the demand-side effects. The partial equilibrium analysis is useful for clarifying the role of endogenous information frictions holding a plan's premium and benefit design fixed. However, endogenous information acquisition is also likely important for examining the competitive effects of policies aimed at simplifying choice. Given that demand for premiums is more elastic, insurers have more incentive to compete on premiums rather than out-of-pocket costs. Moreover, there are implications for other dimensions of insurer competition, such as the number and complexity of plan offerings. Future work should examine how endogenous information acquisition affects competition over product characteristics that are difficult for consumer to observe, as well as firms' equilibrium responses on product positioning and complexity.

References

Abaluck, Jason and Abi Adams-Prassl, "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, 2021, *136* (3), 1611–1663.

_ **and Jonathan Gruber**, "Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program," *American Economic Review*, 2011, *101* (4), 1180–1210.

_ **and** _ , "Evolving choice inconsistencies in choice of prescription drug insurance," *American Economic Review*, 2016, *106* (8), 2145–84.

Altman, DE, J Benson, R Blendon, M Brodie, and C Deane, "Seniors and the Medicare Prescription Drug Benefit," *Kaiser Family Foundation Publication*, 2006, *7604*.

Basu, Anirban, Wesley Yin, and G Caleb Alexander, "Impact of Medicare Part D on Medicare–Medicaid dual-eligible beneficiaries' prescription utilization and expenditures," *Health services research*, 2010, *45* (1), 133–151.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor, "Choose to Lose: Health Plan Choices from a Menu with Dominated Option," *The Quarterly Journal of Economics*, 2017, *132* (3), 1319–1372.

Bhattacharya, Vivek and Greg Howard, "Rational Inattention in the Infield," Working Paper 2020.

Bundorf, M Kate, Jonathan Levin, and Neale Mahoney, "Pricing and welfare in health plan choice," *American Economic Review*, 2012, *102* (7), 3214–48.

_ , Maria Polyakova, Cheryl Stults, Amy Meehan, Roman Klimke, Ting Pun, Albert Solomon Chan, and Ming Tai-Seale, "Machine-Based Expert Recommendations And Insurance Choices Among Medicare Part D Enrollees," *Health Affairs*, 2019, *38* (3), 482–490.

Cabrales, Antonio, Olivier Gossner, and Roberto Serrano, "Entropy and the value of information for investors," *American Economic Review*, 2013, *103* (1), 360–77.

Cardell, N Scott, "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 1997, *13* (2), 185–213.

Coughlin, Maura, "Insurance choice with non-monetary plan attributes: Limited consideration in Medicare Part D," Technical Report, Cornell University Working Paper 2019.

Cummings, Janet R, Thomas Rice, and Yaniv Hanoch, "Who thinks that Part D is too complicated? Survey results on the Medicare prescription drug benefit," *Medical Care Research and Review*, 2009, *66* (1), 97–115.

Ericson, Keith M Marzilli and Amanda Starc, "How product standardization affects choice: Evidence from the Massachusetts Health Insurance Exchange," *Journal of Health Economics*, 2016, *50*, 71–85.

Fosgerau, Mogens, Emerson Melo, Andre De Palma, and Matthew Shum, "Discrete choice and rational inattention: A general equivalence result," *International economic review*, 2020, *61* (4), 1569–1589.

Galichon, Alfred, "On the representation of the nested logit model," *Econometric Theory*, 2022, *38* (2), 370–380.

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

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

Heiss, Florian, Adam Leive, Daniel McFadden, and Joachim Winter, "Plan selection in Medicare Part D: Evidence from administrative data," *Journal of Health Economics*, 2013, *32* (6), 1325–1344.

_ , Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou, "Inattention and switching costs as sources of inertia in medicare part d," Technical Report, National Bureau of Economic Research 2016.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton, "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 2017, *48* (4), 877–905.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest, "Empirical search and consideration sets," in "Handbook of the Economics of Marketing," Vol. 1, Elsevier, 2019, pp. 193–257.

Joo, Joonhwi, "Rational inattention as an empirical framework for discrete choice and consumer-welfare evaluation," *Journal of Marketing Research*, 2023, *60* (2), 278–298.

Ketcham, Jonathan D, Claudio Lucarelli, and Christopher A Powers, "Paying attention or paying too much in Medicare Part D," *American economic review*, 2015, *105* (1), 204–33.

Kling, Jeffrey R, Sendhil Mullainathan, Eldar Shafir, Lee C Vermeulen, and Marian V Wrobel, "Comparison friction: Experimental evidence from Medicare drug plans," *The Quarterly Journal of Economics*, 2012, *127* (1), 199–235.

Lucarelli, Claudio, Jeffrey Prince, and Kosali Simon, "The welfare impact of reducing choice in Medicare Part D: A comparison of two regulation strategies," *International Economic Review*, 2012, *53* (4), 1155–1177.

Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt, "Rational inattention: A disciplined behavioral model," Technical Report, Mimeo, New York City 2018.

Matějka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 2015, *105* (1), 272–98.

Polyakova, Maria, "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 2016, *8* (3), 165–95.

Sims, Christopher A, "Implications of Rational Inattention," *Journal of monetary Economics*, 2003, *50* (3), 665–690.

Train, Kenneth, "Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples," *Journal of Choice Modelling*, 2015, *16* (C), 15–22.

ONLINE APPENDIX

A Model Details

A-1 Model Derivation

Utility is given by

$$u_{ij} = \underbrace{\alpha v_{ij} + \beta X_j^u}_{\text{Initially Unknown}} + \underbrace{\alpha p_j + \theta X_j^k + \epsilon_{ij}}_{\text{Known}} \tag{A-1}$$

and individuals have prior G_i and marginal cost of information acquisition λ.

As in Matějka and McKay (2015), initial choice probabilities before individuals obtain information, $P_1^0, .., P_N^0$, are determined by integrating over the prior given cost of information λ:

$$\max_{P_{i1}^0,..,P_{iJ}^0} \int_{\boldsymbol{\xi_i}} \lambda \log \Sigma_{j=1}^J P_{ij}^0 \exp\left[(\xi_{ij} + \alpha p_j + \theta X_j^k + \epsilon_{ij})/\lambda\right] G(d\boldsymbol{\xi_i})$$
$$\text{s.t.} \quad \sum_{j\in\mathcal{J}} P_{ij}^0 = 1,\, P_{ij}^0 \geq 0\ \forall j \tag{A-2}$$

We start by deriving a closed-form expression for $P_1^0, .., P_N^0$. Note that $\lambda \log \sum_j e^{v_j/\lambda} = \lambda \mathbb{E}_e\left[\max_j(v_j/\lambda + e_j)\right] - \lambda\gamma^e$ where $e_j \overset{iid}{\sim} EV1$ and γ^e is Euler's constant (Small and Rosen 1981). Applying this we have

$$\int_{\boldsymbol{\xi_i}} \lambda \log \Sigma_j e^{(\xi_{ij}+\alpha p_j+\theta X_j^k+\epsilon_{ij})/\lambda+\log(P_{ij}^0)} G(d\boldsymbol{\xi_i}) \tag{A-3}$$

$$= \lambda \mathbb{E}_{\xi,e}\left[\max_j((\alpha p_{ij} + \theta X_j^k + \epsilon_{ij} + \xi_{ij})/\lambda + \log(P_{ij}^0) + e_{ij})\right] - \lambda\gamma^e \tag{A-4}$$

$$= \lambda \mathbb{E}_{\xi,e}\left[\max_j((\alpha_i p_{ij} + \theta X_j^k + \epsilon_{ij})/\lambda + \log(P_{ij}^0) + \xi_{ij}/\lambda + e_{ij})\right] - \lambda\gamma^e \tag{A-5}$$

$$= \lambda \mathbb{E}_{\xi',e}\left[\max_j((\alpha_i p_{ij} + \theta X_j^k + \epsilon_{ij})/\lambda + \log(P_{ij}^0) + \xi_{ij}^0/\lambda + \xi_{ij}'/\lambda + e_{ij})\right] - \lambda\gamma^e \tag{A-6}$$

where ξ_{ijt}' has mean zero and variance σ_{it}^2. The last line follows from the fact that $\mathbb{E}[\xi_{ij}] = \xi_{ij}^0$.

Note that the joint error is $\xi_{ij}'/\lambda + e_j$. Given that e_j is distributed EV1, $Var[e_j] = \frac{\pi^2}{6}$ so

$$Var_j[\xi_{ij}'/\lambda + e_j] = \frac{\sigma_{it}^2}{\lambda^2} + \frac{\pi^2}{6}.$$

We define the joint error as $\ell(\sigma_i, \lambda)e_{ij}^{'} \equiv \xi_{ij}'/\lambda + e_j$ where $Var_j[e_{ij}^{'}] = \frac{\pi^2}{6}$. Therefore,

$$\begin{aligned} Var_j[\ell(\sigma_i, \lambda)e_{ij}^{'}] &= \frac{\sigma_i^2}{\lambda^2} + \frac{\pi^2}{6} \\ \ell(\sigma_i, \lambda)^2 &= \frac{6\sigma_i^2}{\pi^2\lambda^2} + 1 \end{aligned}$$

Then, equation (A-6) can be rewritten as

$$\lambda \mathbb{E}_{e'}\left[\max_j((\alpha p_{ij} + \xi_{ij}^0 + \theta X_j^k + \epsilon_{ij})/\lambda + \log(P_{ij}^0) + \xi_{ij}^0/\lambda + \ell(\sigma_i, \lambda)e_{ij}^{'})\right] - \lambda\gamma^e. \tag{A-7}$$

Note $\mathbb{E}[e_{ij}^{'}] = \frac{\gamma^e}{\ell(\sigma_i,\lambda)}$. Let $e_{ij}^{''} \equiv e_{ij}^{'} + \gamma^e \frac{\ell(\sigma_i,\lambda)-1}{\ell(\sigma_i,\lambda)}$ and assume $e_{ij}^{''}$ is distributed EV1 so $\mathbb{E}[e_{ij}^{''}] = \gamma^e$ and $Var[e_{ij}^{''}] = \frac{\pi^2}{6}$. This implies that the distribution of ξ_{ij}' follows the distribution as in Cardell (1997) and Galichon (2022). Equation (A-7) can then be expressed as

$$\lambda \mathbb{E}_{e'}\left[\max_j((\alpha p_{ij} + \xi_{ij}^0 + \theta X_j^k + \epsilon_{ij})/\lambda + \log(P_{ij}^0) + \xi_{ij}^0/\lambda + \ell(\sigma_i, \lambda)e_{ij}^{''})\right] - \lambda\ell(\sigma_i, \lambda)\gamma^e. \tag{A-8}$$

Now we can again apply the formula from Small and Rosen (1981), this time in reverse. In particular, note that $\mathbb{E}_e\left[\max_j(v_j + \ell e_j)\right] = \ell \log \sum_j e^{v_j/\ell} + \ell\gamma^e$ where e_j is

EV1. This implies that equation (A-8) can be expressed as

$$\lambda\ell(\sigma_i,\lambda)\log\sum_j e^{(\alpha p_{ij}+\xi^0_{ij}+\theta X^k_j+\epsilon_{ij})/\lambda+\log(P^0_{ij})+\xi^0_{ij}/\lambda)/\ell(\sigma_i,\lambda)}. \tag{A-9}$$

Now the maximization problem in equation (A-2) can be rewritten as

$$\max_{P^0_{i1},..,P^0_{iJ}} \Sigma_{j\in\mathcal{J}}\exp[(\alpha p_{ij}+\xi^0_{ij}+\theta X^k_j+\epsilon_{ij})/\ell(\sigma_i,\lambda)\lambda+\log(P^0_{ij})/\ell(\sigma_i,\lambda)] \text{ s.t. } \sum_{j\in\mathcal{J}}P^0_{ij}=1,\, P^0_{ij}\geq 0\ \forall j$$

In the maximization problem we have ignored terms that do not affect the solution. From solving this maximization problem, we can derive a closed-form expression for P^0_{ijt} as

$$P^0_{ij}=\frac{\exp\left[(\alpha p_j+\xi^0_{ij}+\theta X^k_j+\epsilon_{ij})/(\lambda\ell(\sigma_i,\lambda)-\lambda)\right]}{\sum_{k\in\mathcal{J}}\exp\left[(\alpha p_k+\xi^0_{ik}+\theta X^k_k+\epsilon_{ik})/(\lambda\ell(\sigma_i,\lambda)-\lambda)\right]}.$$

With an expression for P^0_{ij} in hand, we can now derive an expression for choice probabilities after information acquisition. Based on Theorem 1 in Matějka and McKay (2015), choice probabilities can be written as

$$P_{ij}=\int_{\boldsymbol{\epsilon_i}}\frac{\exp\left[(\alpha v_{ij}+\beta X^u_j+\alpha_i p_j+\theta X^k_j+\epsilon_{ij})/\lambda+\log(P^0_{ij})\right]}{\sum_{k\in\mathcal{J}}\exp\left[(\alpha v_{ik}+\beta X^u_k+\alpha p_{kt}+\theta X^k_k+\epsilon_{ik})/\lambda+\log(P^0_{ik})\right]}G_i(\boldsymbol{\epsilon_i})$$

where $G_i(\boldsymbol{\epsilon_i})$ is the CDF of the taste shock. Therefore, the problem is now as if individuals maximize utility given by

$$\mathbb{E}[u_{ij}]=(\alpha v_{ij}+\beta X^u_j+\alpha p_j+\theta X^k_j+\epsilon_{ij})/\lambda+\log(P^0_{ij})+e_{ij}$$

where ϵ_{ij} is an iid taste shock and e_{ijt} is an iid EV1 error causes by incorrect beliefs (with variance $\pi^2/6$). Substituting the expression for P^0_{ij}, this becomes

$$\mathbb{E}[u_j]=(\alpha v_{ij}+\beta X^u_j+\alpha p_j+\theta X^k_j+\epsilon_{ij})/\lambda+(\alpha p_j+\xi^0_{ij}+\theta X^k_j+\epsilon_{ij})/(\lambda\ell(\sigma_i,\lambda)-\lambda)+e_{ij} \tag{A-10}$$

where $\log\left[\sum_{k=1}^{N}\exp\left[(\alpha p_k+\xi_{ik}^0+\theta X_k^k+\epsilon_{ik})/(\lambda\ell(\sigma_i,\lambda)-\lambda)\right]\right]$ is a constant that is the same for every option, and therefore does not affect choice probabilities. We can simplify equation (A-10) to

$$
\begin{aligned}
\mathbb{E}[u_{ij}] &= (\alpha v_{ij}+\beta X_j^u+\alpha p_j+\theta X_j^k)/\lambda+(\alpha p_j+\xi_{ij}^0+\theta X_j^k)/(\lambda\ell(\sigma_i,\lambda)-\lambda)+\epsilon_{ij}/(\lambda\ell(\sigma_i,\lambda)-\lambda)+\epsilon_{ij}/\lambda+e_{ij}\\
&= \frac{\alpha v_{ij}+\beta X_j^u+\alpha p_j+\theta X_j^k}{\lambda}+\frac{\alpha p_j+\xi_{ij}^0+\theta X_j^k}{\lambda(\ell(\sigma_i,\lambda)-1)}+\frac{\ell(\sigma_i,\lambda)}{\lambda(\ell(\sigma_i,\lambda)-1)}\epsilon_{ij}+e_{ij}\\
&= \frac{\alpha v_{ij}+\beta X_j^u}{\lambda}+\frac{(\ell(\sigma_i,\lambda)-1)\left(\alpha p_j+\theta X_j^k\right)}{\lambda(l_i-1)}+\frac{\alpha p_j+\xi_{ij}^0+\theta X_j^k}{\lambda(\ell(\sigma_i,\lambda)-1)}+\frac{\ell(\sigma_i,\lambda)}{\lambda(l_i-1)}\epsilon_{ij}+e_{ij}\\
&= \frac{\alpha v_{ij}+\beta X_j^u}{\lambda}+\frac{\alpha\ell(\sigma_i,\lambda)p_j+\xi_{ij}^0+\theta\ell(\sigma_i,\lambda)X_j^k}{\lambda(\ell(\sigma_i,\lambda)-1)}+\frac{\ell(\sigma_i,\lambda)}{\lambda(\ell(\sigma_i,\lambda)-1)}\epsilon_{ij}+e_{ij} \qquad \text{(A-11)}
\end{aligned}
$$

Define the joint error as $k_i e_{ij}^{'}\equiv\frac{\ell(\sigma_i,\lambda)}{\lambda(\ell(\sigma_i,\lambda)-1)}\epsilon_{ij}+e_{ij}$ where $Var[e_{ij}^{'}]=\frac{\pi^2}{6}$. Again, we assume that the distribution of the taste shock is such that the joint error is distributed extreme value type 1. Therefore,

$$
\begin{aligned}
Var[k_i e_{ij}^{'}] &= \frac{\ell(\sigma_i,\lambda)^2}{\lambda^2(\ell(\sigma_i,\lambda)-1)^2}\frac{\pi^2}{6}+\frac{\pi^2}{6}\\
\Rightarrow k_i^2 &= \frac{\ell(\sigma_i,\lambda)^2}{\lambda^2(\ell(\sigma_i,\lambda)-1)^2}+1
\end{aligned}
$$

The expected utility in equation (A-11) can be then be rewritten as

$$
\frac{\alpha v_{ij}+\beta X_j^u}{k_i\lambda}+\frac{\alpha\ell(\sigma_i,\lambda)p_j+\xi_{ij}^0+\theta\ell(\sigma_i,\lambda)X_j^k}{k_i\lambda(\ell(\sigma_i,\lambda)-1)}+e_{ij}^{'}.
$$

Note that the error has been renormalized. Therefore, the choice probabilities are

$$
P_{ij}=\frac{\exp\left[\frac{\alpha v_{ij}+\beta X_j^u}{k_i\lambda}+\frac{\alpha\ell(\sigma_i,\lambda)p_j+\xi_{ij}^0+\theta\ell_{it}X_j^k}{k_i\lambda(\ell(\sigma_i,\lambda)-1)}\right]}{\sum_{k\in\mathcal{J}}\exp\left[\frac{\alpha v_{ik}+\beta X_k^u}{k_i\lambda}+\frac{\alpha\ell(\sigma_i,\lambda)p_k+\xi_{ij}^0+\theta\ell(\sigma_i,\lambda)X_k^k}{k_i\lambda(\ell(\sigma_i,\lambda)-1)}\right]}.
$$

The elasticity of demand with respect to the known component of cost, p_j, is then given by

$$
\begin{aligned}
e^p &= \frac{\partial P_{ij}}{\partial p_j}\frac{p_j+v_{ij}}{P_{ij}}\\
&= \frac{\partial V_{ij}}{\partial p_j}P_{ij}(1-P_{ij})\frac{p_j+v_{ij}}{P_{ij}}
\end{aligned}
$$

$$= \alpha_i \frac{\ell_{it}}{k_{it}\lambda(\ell_{it}-1)}(1-P_{ij})(p_j+v_{ij}), \tag{A-12}$$

while the elasticity of demand with respect to initially unknown component of cost, v_{ij}, is given by

$$\begin{aligned} e^v &= \frac{\partial P_{ij}}{\partial v_{ij}} \frac{p_j+v_{ij}}{P_{ij}} \\ &= \frac{\partial V_{ij}}{\partial v_{ij}} P_{ij}(1-P_{ij}) \frac{p_j+v_{ij}}{P_{ij}} \\ &= \alpha_i \frac{1}{k_{it}\lambda}(1-P_{ij})(p_j+v_{ij}) \end{aligned} \tag{A-13}$$

The above elasticities can be interpreted as the percent change in demand due to a one percent change in cost due to p_j and v_{ij} respectively. In the context of insurance choice, this implies individuals will be more sensitive to premiums then out-of-pocket cost when information is costly.

A-2 Basic Model for Simulation

We can consider the simple case with no idiosyncratic taste shock in which utility is given by $u_{ij} = -p_j - v_{ij}$ where v_{ij} requires costly information acquisition. Given the distributional assumption on the prior, P_{ij}^0 is then given by

$$P_{ij}^0 = \frac{e^{-p_j/(\lambda\ell(\sigma_i,\lambda)-\lambda)}}{\sum_k e^{-p_k/(\lambda\ell(\sigma_i,\lambda)-\lambda)}}$$

where

$$\ell(\sigma_i,\lambda) \equiv \left(\frac{6\sigma_i^2}{\pi^2\lambda^2}+1\right)^{\frac{1}{2}}.$$

It is as if the agent maximizes expected utility

$$\mathbb{E}[u_{ij}] = (-p_j - v_{ij})/\lambda + \log(P_{ij}^0) + e_{ij}$$

where e_{ij} is an iid EV1 error causes by incorrect beliefs. Substituting the expression for P_{ij}^0, expected utility is

$$\mathbb{E}[u_{ij}] = (-p_j - v_{ij})/\lambda + p_j/(\lambda\ell(\sigma_i,\lambda)-\lambda) + \log(\sum_k e^{-p_k/(\lambda\ell(\sigma_i,\lambda)-\lambda)}) + e_{ij}$$

where $\log(\sum_k e^{-p_k/(\lambda\ell(\sigma_i,\lambda)-\lambda)})$ is the same for every option, and therefore can be ignored. This yields closed-form choice probabilities given by

$$P_{ij} = \frac{e^{(-p_j\ell(\sigma_i,\lambda)/(\ell(\sigma_i,\lambda)-1)-v_{ij})/\lambda}}{\sum_k e^{(-p_k\ell(\sigma_i,\lambda)/(\ell(\sigma_i,\lambda)-1)-v_{ik})/\lambda}}. \tag{A-14}$$

The above expression implies that individuals respond differentially to an equivalent change in p_j and v_{ij}. In particular, the elasticity of demand with respect to a change in cost due to p_j is given by

$$e^p = \frac{\ell(\sigma_i,\lambda)}{\lambda(\ell(\sigma_i,\lambda)-1)}(1-P_{ij})(p_j+v_{ij}), \tag{A-15}$$

while the elasticity of demand with respect to a change in cost due to v_j is given by

$$e^v = \frac{1}{\lambda}(1-P_{ij})(p_j+v_{ij}). \tag{A-16}$$

A-3 Empirical Model Likelihood function

Given the empirical model presented in section 4, choice probabilities are given by

$$P_{ijt} = \frac{\exp\left[a(\sigma_{it},\lambda_{it})\left(\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2\widetilde{\sigma}^2_{ijt}\right) + b(\sigma_{it},\lambda_{it})\left(\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}\right)\right]}{\sum_{k\in\mathcal{J}}\exp\left[a(\sigma_{it},\lambda_{it})\left(\alpha_i v_{ikt} + \beta_1 X^u_{kt} + \beta_2\widetilde{\sigma}^2_{ikt}\right) + b(\sigma_{it},\lambda_{it})\left(\alpha_i p_{kt} + \beta_3 X^k_{kt} + \zeta_{b(k)d(it)}\right)\right]} \tag{A-17}$$

Let the set of parameters by $\Phi = \{\alpha,\lambda,\beta,\zeta\}$. The log-likelihood function is given by

$$\mathcal{L}(\Phi) = \sum_i\sum_t\left(\sum_{j\in\mathcal{J}_{it}} I(y_{it}=j)\tilde{\nu}_{ijt}(\Phi) - \log\left(\sum_{j\in\mathcal{J}_{it}}\exp\tilde{\nu}_{ijt}(\Phi)\right)\right) \tag{A-18}$$

where

$$\tilde{\nu}_{ijt}(\Phi) = a(\sigma_{it}(\Phi),\lambda_{it})\left(\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2\widetilde{\sigma}^2_{ijt}\right) + b(\sigma_{it}(\Phi),\lambda_{it})\left(\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}\right). \tag{A-19}$$

Note that $\sigma_{it}(\Phi) = Var\left[\alpha v_{i1t} + \beta X^u_{1t}, \alpha v_{i2t} + \beta X^u_{2t}, \ldots, \alpha v_{iJt} + \beta X^u_{Jt}\right]$ is a function of model parameters.

A-4 Derivation of Welfare

We denote individual i's expected utility from alternative j given beliefs after information acquisition as $\tilde{u}_{ijt}$. The difference between the realized utility and the expected utility given information acquisition is denoted d_{ijt}. Then, the realized utility can be written as

$$u_{ijt} = \tilde{u}_{ijt} + d_{ijt}$$

Denoting j^* as the option in $\mathcal{J}$ that maximizes the individual's belief utility, consumer surplus under rational inattention can be expressed as

$$\begin{aligned} CS^{RI} &= \frac{1}{-\alpha_i}\mathbb{E}[\tilde{u}_{ij^*t} + d_{ij^*t}] \\ &= \frac{1}{-\alpha_i}\mathbb{E}[\max_j \tilde{u}_{ijt}] + \frac{1}{-\alpha_i}\sum_j P_{ijt}d_{ijt} \\ &= \frac{1}{-\alpha_i}\log\sum_j \exp[\tilde{\nu}_{ijt}] + \frac{1}{-\alpha_i}\sum_j P_{ijt}[\nu_{ijt} - \tilde{\nu}_{ijt}] \end{aligned}$$

where ν_{ijt} and $\tilde{\nu}_{ijt}$ are given by equation (17) and equation (18).

The cost function can be expressed in terms of the initial choice probabilities before individuals acquire information and the final choice probabilities

$$\hat{C}_{it} = \frac{\lambda_{it}}{-\alpha_i}\int_\epsilon \left(-\sum_{j\in\mathcal{J}_{it}} P^0_{ijt}(\epsilon)\log P^0_{ijt}(\epsilon) + \int_\xi \left(\sum_{j\in\mathcal{J}_{it}} P_{ijt}(\xi,\epsilon)\log P_{ijt}(\xi,\epsilon) \right) G_i(d\xi) \right) M(d\epsilon) \tag{A-20}$$

where $P^0_{ijt}(\epsilon)$ is the initial choice probability before information acquisition given ϵ, $P_{ijt}(\xi,\epsilon)$ is the choice probability after information acquisition given (ξ,ϵ), $G_i(\xi)$ is the distribution of the prior, and $M(\epsilon)$ is the distribution of the taste shock. In practice, the entropy of posterior beliefs can be evaluated using simulation methods by drawing from distribution $G_i(\xi)$ and $M(\epsilon)$ and averaging over the draws.

B Details on Data Construction

The sample selection criteria follows Abaluck and Gruber (2016). We drop individuals that are eligible for low-income subsidies, those with employer coverage, individuals who move during the year, those with enrolled in multiple plans, those that are enrolled for less than a full year, and those enrolled in plans with less than 100 enrollees in the state. Furthermore, we limit the sample to active switchers. Active switchers are defined as new enrollees in addition to individuals that were previously enrolled in a plan that is no longer available.

In order to construct expected out-of-pocket costs, we employ the Medicare Part D calculator from Abaluck and Gruber (2016). The calculator uses observed claims for an individual to construct out-of-pocket costs for all plans in the individual's choice set. While we follow the approach of Abaluck and Gruber (2016) closely, one difference is that our sample allows us to use data on plan formularies rather than reconstruct formularies from observed claims. The formulary data, which is provided by CMS, provides information about the tier of each drug and if the drug is covered at all. We combine this with information on plan characteristics that are constant for all plans in a given year such as the catastrophic threshold.

For each plan, an individual's claims are put into the calculator in chronological order and the copay and coinsurance are calculated given the plan formulary and Medicare Part D benefit design. Following Abaluck and Gruber (2016) we allow individuals to substitute to lower cost drugs, where drugs are defined by their ingredients, strength, dosage, and route of administration. To construct the rational expectations measure of expected out-of-pocket costs, the calculator defines 1,000 groups based on prior year's total expenditure, quantity of branded drugs in days, and quantity of generic drugs in days as in Abaluck and Gruber (2011). When prior year claims are not available, the calculator uses the beginning of the current year. We then consider the average and variance of individuals in the same group to get expected out-of-pocket costs and plan variance respectively. Abaluck and Gruber (2016) find that their calculator is able to accurately predict out-of-pocket costs for individuals' chosen plans and is robust to alternative specifications.

C Details of Alternative Models without Endogenous Information

In order to examine the implications of the endogenous information model, it is useful to compare the results to alternative empirical models of insurance demand that do not have endogenous information. In this section, we present that details of these alternative models.

Standard logit model

Canonical models of insurance often assume that individuals have full information about the distribution of out-of-pocket cost.[46] We start by estimating a standard logit model assuming that individuals have full information about both premiums and expected out-of-pocket cost. Therefore, individuals treat both premium and expected out-of-pocket cost in the same way, i.e. they have the same coefficient. The endogenous information model nests this model when the marginal cost of information is zero. In this case, utility takes the form

$$u_{ijt} = \alpha_i \underbrace{(v_{ijt} + p_{jt})}_{\text{Total Cost}} + \beta_1 \widetilde{\sigma}^2_{ijt} + \beta_2 X_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt}. \tag{A-21}$$

As in the baseline endogenous information model, $\widetilde{\sigma}^2_{ijt}$ is the riskiness of the plan, i.e. variance of out-of-pocket costs, X_{jt} is plan quality, and $\zeta_{b(j)d(it)}$ are plan fixed effects. In all of the above models, the coefficient on cost, α_i, is assumed to be a function of individual observable characteristics (income, education, age, age squared, female, and an indicator for rural). The idiosycratic error, ϵ_{ijt}, is assumed to follow a EV1 distribution.

Coverage characteristics model

A common approach in the empirical literature on insurance demand is to assume that utility is a function of premium and coverage characteristics rather than expected out-of-pocket cost. See, for instance, Bundorf et al. (2012), Handel (2013), and Polyakova (2016). Decarolis et al. (2020), Polyakova (2016), Ericson and Starc (2016), and Tebaldi (2017). A related approach uses plan fixed effects to absorb differences in deductible, coinsurance, or other coverage characteristics. In particular, we assume utility takes the form

$$u_{ijt} = \alpha_i p_{jt} + \beta_1 C_{jt} + \beta_2 \widetilde{\sigma}^2_{ijt} + \beta_3 X_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt} \tag{A-22}$$

where C_{jt} are coverage characteristics including deductible, cost sharing, generic coverage, and coverage in the gap. Assumptions about $\widetilde{\sigma}^2_{ijt}$, X_{jt}, α_i, $\zeta_{b(j)d(it)}$, and ϵ_{ijt} are the same as the previous model.

[46] See, for instance, review by Einav et al. (2010).

Differential weight model

Finally, we consider a model in which there is a different coefficient on premium and expected out-of-pocket cost. This approach, used by Abaluck and Gruber (2011) and Abaluck and Gruber (2016), assumes that the coefficients are fixed when considering counterfactual policies. Ho et al. (2017) and Heiss et al. (2016) use a similar approach. For this model, we assume utility is given by

$$u_{ijt} = \alpha_i p_{jt} + \beta_1 v_{ijt} + \beta_2 \widetilde{\sigma}^2_{ijt} + \beta_3 X_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt}. \tag{A-23}$$

We maintain assumptions regarding $\widetilde{\sigma}^2_{ijt}$, X_{jt}, α_i, $\zeta_{b(j)d(it)}$, and ϵ_{ijt}. One interpretation of this model is that the difference between α_i and β_1 reflects *exogenous* information frictions. Unlike the endogenous information model presented in the previous section, there is no scope for the stakes to affect information acquisition.

Results from Alternative Models

We estimate the models via MLE and present the parameter estimates in Table A-1.

To evaluate the fit of the alternative models, we simulate baseline choice probabilities from each model and use the simulated data to estimate the probability of choosing the lowest-cost option based on equation (12) and weights on premium and expected out-of-pocket cost based on equation (11). Figure A-1 Panel c and d show the fit of the alternative models.

Table A-2 shows results from counterfactual experiments under the alternative models.

Table A-1
Estimates for Alternative Models of Insurance Demand
without Endogenous Information

	Standard Logit		Coverage Characteristics		Differential Weight	
Total cost	-0.0773^{***}	(0.0034)				
Total cost × Income	0.0003^{***}	(0.0000)				
Risk	0.0027^{**}	(0.0011)	-0.0009	(0.0011)	0.0007	(0.0011)
Premium			-0.1705^{***}	(0.0052)	-0.1130^{***}	(0.0042)
Premium × Income			0.0004^{***}	(0.0000)	0.0000	(0.0000)
Deductible			-0.0051^{***}	(0.0001)		
Generic coverage			-0.8841^{***}	(0.0266)		
Coverage in gap			0.3227^{***}	(0.0266)		
Cost sharing			0.5176^{***}	(0.0734)		
OOP					-0.0211^{***}	(0.0015)
Other controls for plan characteristic	Yes		Yes		Yes	
Insurer Fixed Effects × Chronic Conditions	Yes		Yes		Yes	
Log Likelihood	-51,940		-96,649		-50,772	

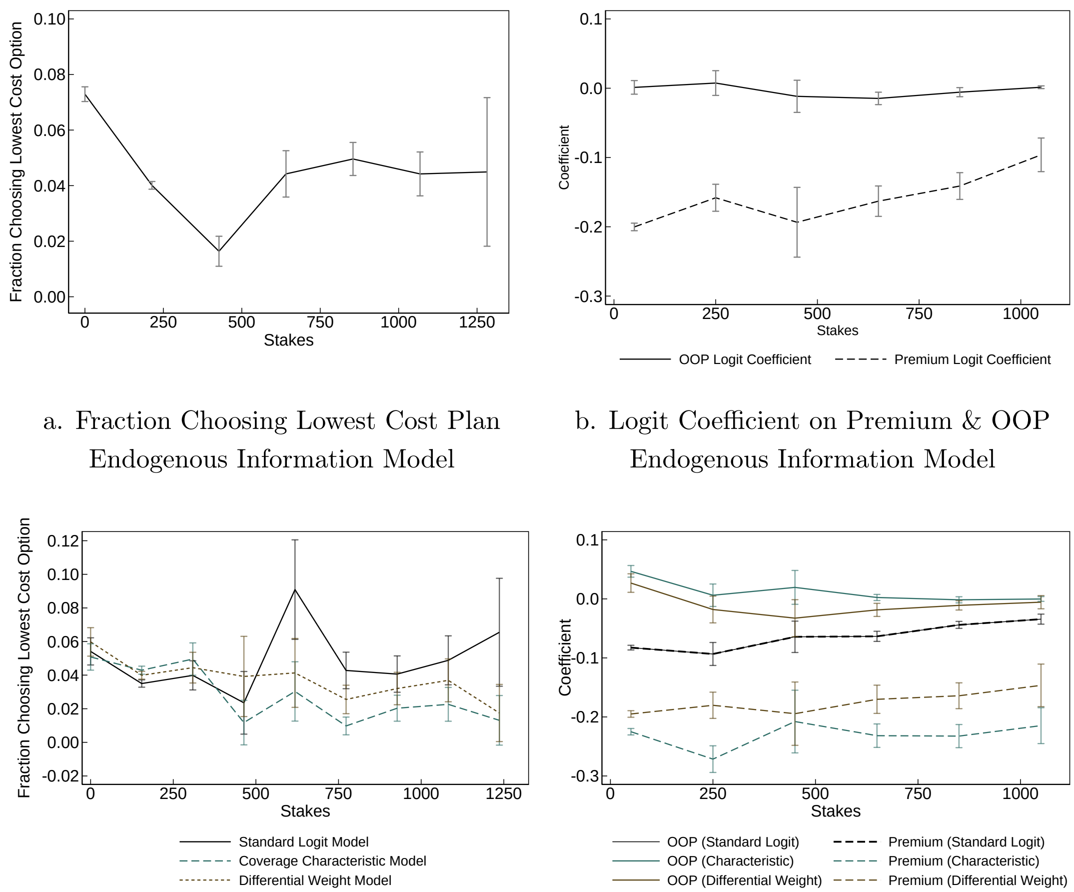
Notes: The details of each model are presented in Appendix C. Premium and out-of-pocket cost are in hundreds of dollars. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-2
Counterfactual Spending and Welfare for Restricted Choice Set and Out-of-Pocket Cap from Alternative Demand Models

	Restricted Choice Set		Out-of-Pocket Cap	
	10th Percentile Cutoff	25th Percentile Cutoff	\$5,000 Cap	\$15,000 Cap
Standard logit model				
Δ Premium	-0.2	-0.3	-21.2	-13.0
Δ Out-of-pocket cost	-0.2	-0.9	-318.1	-137.7
Δ Spending	-0.4	-1.2	-339.3	-150.6
Δ Welfare	-4.3	-18.8	350.7	166.5
Coverage characteristics model				
Δ Premium	0.1	0.2	0.0	0.0
Δ Out-of-pocket cost	-0.2	-0.7	-410.5	-215.7
Δ Spending	-0.1	-0.5	-410.5	-215.7
Δ Welfare	-2.1	-8.2	0.0	0.0
Differential weight model				
Δ Premium	-0.1	-0.2	-7.5	-4.7
Δ Out-of-pocket cost	-0.1	-0.5	-379.6	-186.4
Δ Spending	-0.2	-0.7	-387.1	-191.1
Δ Welfare	-1.5	-6.8	73.0	36.9

Notes: Counterfactual simulations from alternative models described in Appendix C. Restricted choice counterfactual removes plans with average utility below cutoff based on estimates from endogenous information model. Out-of-pocket cap counterfactual imposes limit on out-of-pocket cost of all plans and then simulates plan choice.

Figure A-1
Fit of Endogenous Information Model and Alternative Models



a. Fraction Choosing Lowest Cost Plan
Endogenous Information Model

b. Logit Coefficient on Premium & OOP
Endogenous Information Model

c. Fraction Choosing Lowest Cost Plan
Alternative Models

d. Logit Coefficient on Premium & OOP
Alternative Models

Notes: Left charts show mean fraction of individuals choosing lowest cost option. Standard error bars show 95% confidence interval for the mean. Right charts show logit coefficient on annual out-of-pocket cost and annual premium interacted with indicators for the stakes. These can be compared to Figure 2 and Figure 3. For further description see Section 2.

D Robustness Results for Motivating Evidence

In this section, we present detailed results from robustness checks for our analysis in Section 3.2. In Table A-3, we examine the relationship between the stakes and the probability of choosing the lowest cost plan while using the perfect-foresight measure of out-of-pocket costs. This measure is constructed based on each individual's realized utilization of out-of-pocket costs to address the concern that there can be measurement error with our baseline measure based on rational expectations. Table A-4 shows the relationship between the stakes and choice quality remains even more stronger when using the restricted sample of new enrollees. In Figure A-2, we explore two alternative measures of choice quality: the fraction of individuals choosing a plan in the lowest decile and quintile of out-of-pocket costs among the plans in their choice set. We also consider quality measures based on plan riskiness and quality in Figure A-3. To the extent that these plan characteristics are also initially hard to observe, we would expect a similar relationship. Across all of these alternative outcomes, we find the evidence of a U-shaped relationship between the stakes and choice quality.

Table A-3
Non-Monotonic Effect of Stakes on Insurance Choice
Robustness Check with Perfect Foresight Assumption

	(1)	(2)	(3)	(4)	(5)	(6)
Stakes (100s)	-0.0220^{***}	-0.0213^{***}	-0.0016	-0.0204^{***}		
	(0.0023)	(0.0025)	(0.0017)	(0.0025)		
Stakes Squared	0.0019^{***}	0.0018^{***}	0.0003^{*}	0.0018^{***}		
	(0.0002)	(0.0002)	(0.0001)	(0.0002)		
Stakes quintile 2					-0.0444^{***}	-0.0015
					(0.0034)	(0.0027)
Stakes quintile 3					-0.0536^{***}	-0.0065^{*}
					(0.0042)	(0.0034)
Stakes quintile 4					-0.0539^{***}	-0.0089^{***}
					(0.0036)	(0.0030)
Stakes quintile 5					-0.0474^{***}	0.0017
					(0.0033)	(0.0036)
Individual FEs	No	No	Yes	No	No	Yes
Year FEs	No	No	Yes	Yes	No	Yes
Market FEs	No	No	No	Yes	No	No
Controls for Plan Characteristics & Number of Plans	No	Yes	Yes	Yes	Yes	Yes
Implied minimum	573.5	582.7	300.9	581.9		
Adjusted R2	0.007	0.009	0.269	0.011	0.016	0.269
Observations	199,783	193,745	183,402	193,745	193,745	183,402

Notes: Estimates from linear probability model where dependent variable is the indicator variable for whether the individual chooses the lowest cost plan, where lowest cost plan is defined using a perfect foresight assumption. Standard errors clustered at the market level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-4
Non-Monotonic Effect of Stakes on Choice of Lowest Cost Insurance Plan
Robustness Check with First-Time Enrollees

	(1)	(2)	(3)	(4)
Stakes (100s)	-0.0728^{***}	-0.0724^{***}	-0.0726^{***}	-0.0719^{***}
	(0.0060)	(0.0076)	(0.0076)	(0.0076)
Stakes Squared	0.0060^{***}	0.0061^{***}	0.0061^{***}	0.0060^{***}
	(0.0006)	(0.0008)	(0.0008)	(0.0008)
Year FEs	No	No	Yes	Yes
Market FEs	No	No	No	Yes
Controls for Plan Characteristics & Number of Plans	No	Yes	Yes	Yes
Implied minimum	605.2	592.9	592.7	594.1
Adjusted R2	0.044	0.066	0.070	0.074
Observations	99,031	95,271	95,271	95,271

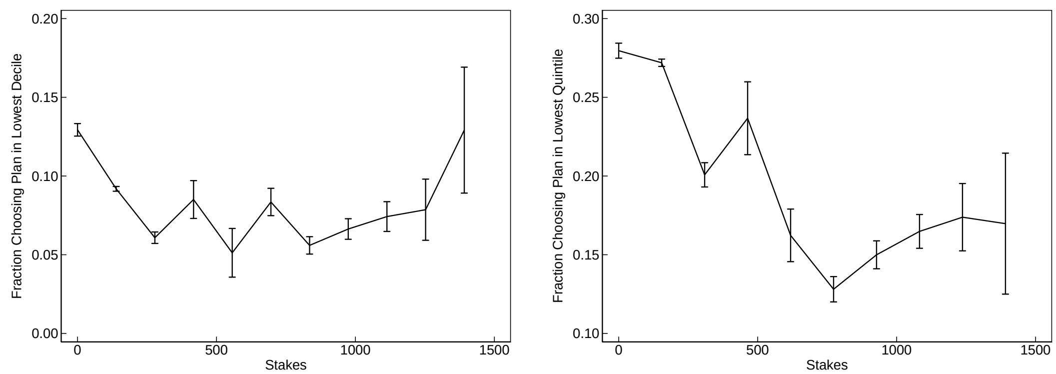
Notes: Estimates from linear probability model where dependent variable is the indicator variable for whether the individual chooses the lowest cost plan. Standard errors clustered at the market level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-5
Interaction of Stakes and Price Coefficient in Standard Logit Model
Robustness Check with Perfect Foresight Assumption

	(1)	(2)	(3)	(4)	(5)	(6)
Premium (100s)	-0.234^{***}	-0.279^{***}	-0.492^{***}	-0.294^{***}	-0.489^{***}	-0.486^{***}
	(0.003)	(0.003)	(0.021)	(0.003)	(0.021)	(0.022)
Premium × Indiv. avg stakes				0.019^{***}	0.018^{***}	0.017^{***}
				(0.001)	(0.001)	(0.001)
Premium × Stakes		0.020^{***}	0.018^{***}	0.008^{***}	0.008^{***}	
		(0.001)	(0.001)	(0.001)	(0.001)	
Premium × Stakes × $\mathbb{1}$(Stakes > 0)						0.005^{***}
						(0.001)
Premium × Stakes × $\mathbb{1}$(Stakes < 0)						0.013^{***}
						(0.001)
Out-of-Pocket Cost (100s)	-0.023^{***}	-0.020^{***}	-0.057^{***}	-0.013^{***}	-0.049^{**}	-0.046^{**}
	(0.002)	(0.005)	(0.019)	(0.005)	(0.019)	(0.019)
OOP × Indiv. avg stakes				0.003^{***}	0.002^{***}	0.002^{***}
				(0.001)	(0.001)	(0.001)
OOP × Stakes		0.003^{***}	0.003^{***}	0.001^{**}	0.001^{*}	
		(0.000)	(0.000)	(0.000)	(0.000)	
OOP × Stakes × $\mathbb{1}$(Stakes > 0)						0.000
						(0.000)
OOP × Stakes × $\mathbb{1}$(Stakes < 0)						0.001^{***}
						(0.000)
Premium × X_i	No	No	Yes	No	Yes	Yes
OOP × X_i	No	No	Yes	No	Yes	Yes
Log Likelihood	-114,144	-113,804	-113,329	-113,652	-113,196	-113,179
Observations	1,025,674	1,025,674	1,025,674	1,025,674	1,025,674	1,025,674

Notes: Stakes in hundreds of dollars. All specifications include controls for risk aversion (OOP variance), plan quality rating, deductible, generic coverage, coverage in the donut hole, and cost sharing. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

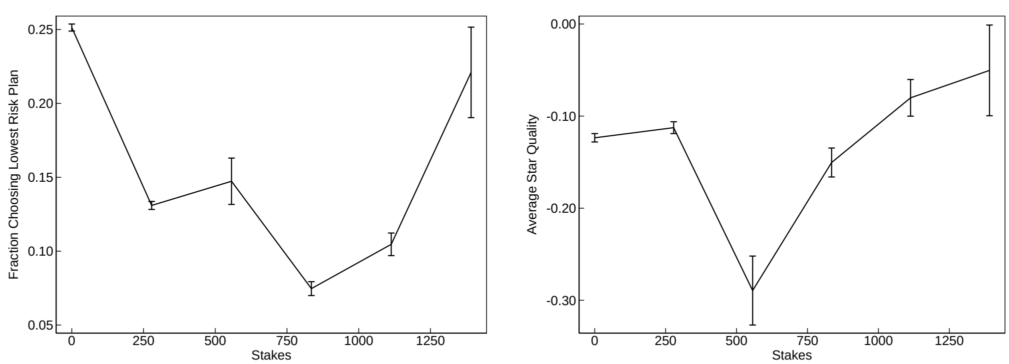
Figure A-2
Alternative Measures of Probability of Choosing Low Cost Plan by Stakes



a. Plan in Lowest Decile b. Plan in Lowest Quintile

Notes: For average percentile rank, higher percentile rank indicates lower cost choice. Standard error bars show 95% confidence interval for the mean.

Figure A-3
Alternative Measures of Choice Quality by Stakes



a. Fraction Choosing Lowest Risk Plan b. Average Plan Quality

Notes: Plan quality measured by Medicare star ratings. Standard error bars show 95% confidence interval for the mean.

E Results from Alternative Specifications and Additional Counterfactuals

Table A-6 presents results from alternative specifications to our baseline version. In the first column of the table, we consider excluding insurer fixed effects. In the second column, we consider include outliers with extreme values of the stakes to our main sample.

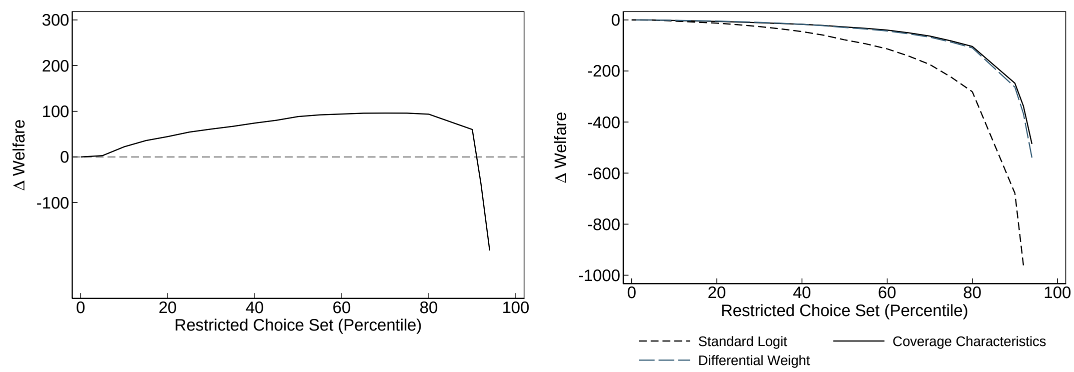
Table A-6
Estimates for Demand Model with Endogenous Information Acquisition
Alternative Specifications

	(1)		(2)	
	No Insurer Fixed Effects		Including Outliers	
	Estimate	SE	Estimate	SE
Price Sensitivity (β^{α})				
Constant	−2.1185	(0.0186)	−2.1560	(0.0209)
Income	−0.0008	(0.0005)	−0.0011	(0.0005)
Other Plan Characteristics				
Previous insurer	6.2254	(0.0506)	6.5487	(0.0707)
Risk	−0.0436	(0.0024)	−0.0542	(0.0032)
Star rating	1.5402	(0.0849)	1.9178	(0.1294)
Marginal cost of information (β^{λ})				
Constant	2.9757	(0.1636)	2.9856	(0.1437)
Zip Income	−0.0002	(0.0011)	0.0000	(0.0009)
Zip Education	−0.0007	(0.0023)	−0.0012	(0.0020)
Age	0.6377	(0.0957)	0.4214	(0.0714)
Age^2	−0.0038	(0.0006)	−0.0025	(0.0004)
Female	−0.0113	(0.0447)	0.0119	(0.0393)
Part D Experience	−0.4511	(0.0338)	−0.3762	(0.0282)
Rural	0.2124	(0.0588)	0.2103	(0.0514)
Has alzheimers	0.0935	(0.0732)	0.0630	(0.0644)
Has lung disease	0.1712	(0.0704)	0.1058	(0.0616)
Has kidney disease	−0.0694	(0.0571)	−0.0958	(0.0512)
Has heart failure	0.0851	(0.0624)	0.0769	(0.0564)
Has depression	0.0244	(0.0659)	−0.0002	(0.0579)
Has diabetes	0.0957	(0.0504)	0.0510	(0.0449)
Has other chronic condition	0.0304	(0.0511)	−0.0349	(0.0441)
Mean price sensitivity	−0.1203		−0.1159	
Mean marginal cost of information	2.2975		3.1415	
LL	-54,452.83		-50,850.09	
Observations	1,035,319		1,021,782	

Notes: Specification 1 does not include insurer fixed effects. Specification 2 includes individuals with outlier stakes, which are not included in the baseline specification. Premium and out-of-pocket cost are in hundreds of dollars. Continuous individual characteristics (income, education, age, and age squared) are demeaned. Standard errors in parentheses.

We conduct a counterfactual that restricts the choice set by offering the personalized list of plans optimal to each individual. This contrast with our baseline specification in which the choice set is personalized by age bins. The welfare gains are even larger in this case as shown in Figure A-4.

Figure A-4
Counterfactual Welfare Effects of Restricted Choice Set

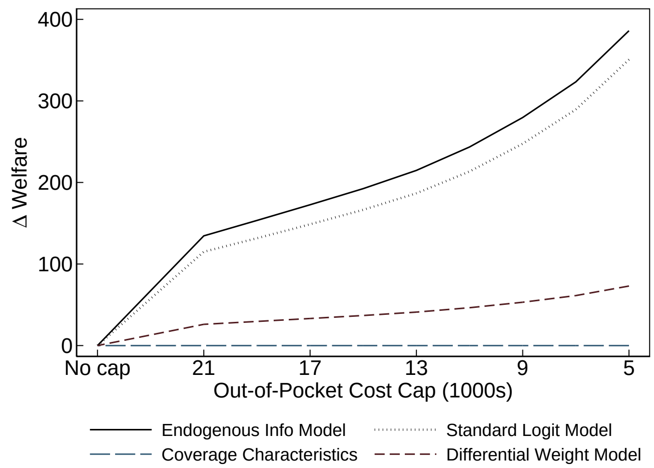


a. Endogenous Information Model

b. Alternative Models

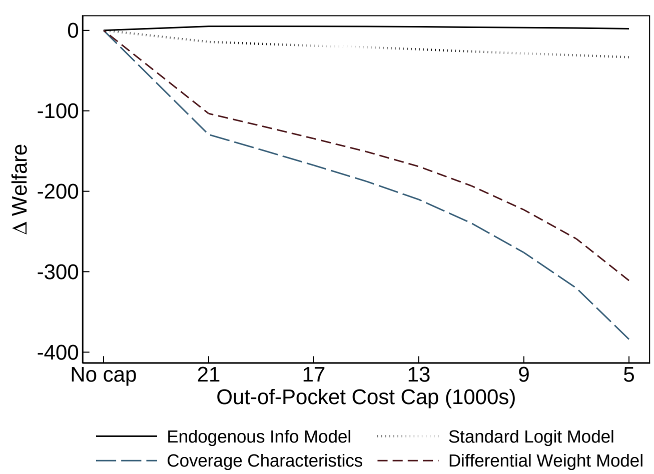
Notes: Chart shows counterfactual average change in welfare per enrollee from removing plans with mean utility below a given percentile where average utlity is computed for each individual. Counterfactual estimates from model with endogenous information acquisition are contrasted with counterfactual welfare estimates from commonly used models of plan demand.

Figure A-5
Counterfactual Welfare Effects of Out-of-Pocket Cost Cap



Notes: Chart shows counterfactual change in welfare from capping out-of-pocket cost at different levels. Counterfactual estimates from model with endogenous information acquisition is contrasted with counterfactual estimates from alternative models without endogenous information.

Figure A-6
Counterfactual Welfare for Out-of-Pocket Cost Cap
When Adjusting Premiums so Policy is Revenue Neutral



Notes: Chart shows counterfactual change in welfare from capping out-of-pocket cost at different levels while increasing premiums such that the policy is revenue neutral. Counterfactual estimates from model with endogenous information acquisition is contrasted with counterfactual estimates from alternative demand models without endogenous information.

F Identification

For simplicity, consider the baseline model we estimate in which individuals hold common priors for all options. Furthermore, we abstract from the product fixed effects, $\zeta_{b(j)d(it)}$. The choice probabilities are given by

$$P_{ijt} = \frac{\exp\left[a(\sigma_{it},\lambda_{it})\left(\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma}^2_{ijt}\right) + b(\sigma_{it},\lambda_{it})\left(\alpha_i p_{jt} + \beta_3 X^k_{jt}\right)\right]}{\sum_{k\in\mathcal{J}} \exp\left[a(\sigma_{it},\lambda_{it})\left(\alpha_i v_{ikt} + \beta_1 X^u_{kt} + \beta_2 \widetilde{\sigma}^2_{ikt}\right) + b(\sigma_{it},\lambda_{it})\left(\alpha_i p_{kt} + \beta_3 X^k_{kt}\right)\right]}. \tag{A-24}$$

where $a(\sigma_{it},\lambda_{it})$ and $b(\sigma_{it},\lambda_{it})$ are defined in Section 2. The key assumptions that lead to these choice probabilities are a) individuals have risk preferences approximated by CARA utility with normally distributed out-of-pocket costs; and b) there is an additive taste shock with distribution $M(\epsilon_{ijt})$ and c) the distribution of the priors is $G_i(\xi_{ij})$.

We can redefine the coefficients in equation (A-24) and rewrite the choice probabilities as:

$$P_{ijt} = \frac{\exp\left[\rho^0_{it} v_{ijt} + \rho^1_{it} X^u_{jt} + \rho^2_{it} \widetilde{\sigma}^2_{ijt} + \rho^3_{it} p_{jt} + \rho^4_{it} X^k_{jt}\right]}{\sum_{k\in\mathcal{J}} \exp\left[\rho^0_{it} v_{ikt} + \rho^1_{it} X^u_{kt} + \rho^2_{it} \widetilde{\sigma}^2_{ikt} + \rho^3_{it} p_{kt} + \rho^4_{it} X^k_{kt}\right]}$$

where $\rho^0_{it} = \alpha_i a(\sigma_{it},\lambda_{it})$, $\rho^1_{it} = \beta_1 a(\sigma_{it},\lambda_{it})$, $\rho^2_{it} = \beta_2 a(\sigma_{it},\lambda_{it})$, $\rho^3_{it} = \alpha_i b(\sigma_{it},\lambda_{it})$, and $\rho^4_{it} = \beta_3 b(\sigma_{it},\lambda_{it})$. Identification of parameters $\boldsymbol{\rho_i} = \{\rho^0_{it},\rho^1_{it},\rho^2_{it},\rho^3_{it},\rho^4_{it}\}$ is then standard and comes from variation in individuals' choice sets across markets.[47] If individuals are more sensitive to premiums than out-of-pocket cost, the coefficient on the premium, ρ^3_{it}, will differ from the coefficient on the out-of-pocket cost, ρ^0_{it}. Dividing the coefficient on the premium by the coefficient on out-of-pocket cost, we obtain

$$\frac{\rho^3_{it}}{\rho^0_{it}} = \frac{\ell_{it}(\sigma_{it},\lambda_{it})}{\ell_{it}(\sigma_{it},\lambda_{it})-1} = \frac{\left(6\sigma^2_{it} + \pi^2\lambda^2_{it}\right)^{\frac{1}{2}}}{\left(6\sigma^2_{it} + \pi^2\lambda^2_{it}\right)^{\frac{1}{2}} - \pi\lambda_{it}} \tag{A-25}$$

Hence, given the variance of the prior belief, σ^2_{it}, the ratio $\frac{\rho^3_{it}}{\rho^0_{it}}$ pins down the information cost parameter λ_{it}. Based on the estimates of λ_{it} and $\boldsymbol{\rho_{it}}$, we can then obtain the price coefficient α_i and other preference parameters $(\beta_1,\beta_2,\beta_3)$.

Alternatively, one could estimate $\boldsymbol{\rho_i} = \{\rho^0_{it},\rho^1_{it},\rho^2_{it},\rho^3_{it},\rho^4_{it}\}$ directly by estimating a logit model, ideally including interactions to allow the coefficients to vary by individual characteristics and stakes. In this case, λ_{it} can be recovered from the "reduced-form" parameters

[47]For example, Abaluck and Gruber (2011) estimates these parameters in a standard logit model. The same identification argument applies.

using

$$\lambda_{it} = \frac{\sqrt{6}(\frac{\rho^3_{it}}{\rho^0_{it}} - 1)\sigma_{it}}{\pi(2\frac{\rho^3_{it}}{\rho^0_{it}} - 1)^{\frac{1}{2}}}.$$

G Monte Carlo Analysis to Assess Sensitivity to Distributional Assumptions

We conduct a Monte Carlo exercise as part of our robustness analysis. In particular, we examine whether estimates are sensitive to the distributional assumption on the prior of out-of-pocket costs that is used in deriving the closed-form expression of choice probabilities. Using the model presented in appendix A-2, we simulate premiums and out-of-pocket costs by drawing from a normal distribution. Table A-7 lists parameter values chosen for the simulation.

Table A-7
Parameter Values for a Monte Carlo Simulation

Number of choice situations (N)	{1000,5000}
Number of options	3
Cost of information (λ)	10
Variance of out-of-pocket costs	15
Variance of premiums	10

We compute choice probabilities based on two different assumptions about the prior. In the first case, we assume a normally distributed prior that coincides with the true distribution of out-of-pocket costs. In this case, we can compute initial choice probabilities by numerically solving equation (A-2) based on simulated maximum likelihood. In the second case, we assume that a non-standard prior that gives rise to a closed-form expression for choice probabilities as described in Online Appendix A. Then, we can compute initial choice probabilities based on equation (A-14). We draw choices based on these two sets of choice probabilities and estimate the cost of information using maximum likelihood.

We simulate 1000 and 5000 choice situations under the two sets of assumptions and repeat each simulation 50 times. Table A-8 shows results from the simulations. The distributional assumption on the prior does not have a significant effect on the estimate of the information cost (λ). The mean squared error is 0.016 under the normal prior and 0.037 under the alternative non-standard distribution for the sample size of 5000. Given that

Table A-8
Monte Carlo Results

	$N = 1000$			
	Estimate		MSE	
True value	Normal	Non-standard	Normal	Non-standard
10	10.087 (0.314)	9.973 (0.497)	0.104	0.243
	$N = 5000$			
	Estimate		MSE	
True value	Normal	Non-standard	Normal	Non-standard
10	9.990 (0.129)	9.990 (0.193)	0.016	0.037

Notes: Standard errors are in parentheses.

the misspecified model is quite accurate, this implies that the distributional assumption is relatively innocuous. At the same time, the use of the closed-form expression dramatically reduces the computational burden. When using simulated MLE with the normal prior, the Monte Carlo exercise with the sample size of 5000 takes nearly 6 hours on 56 cores. With the closed-form expression, the computational time is reduced to 5 seconds.

Appendix References

Abaluck, Jason and Jonathan Gruber, "Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program," *American Economic Review*, 2011, *101* (4), 1180–1210.

_ **and** _ , "Evolving choice inconsistencies in choice of prescription drug insurance," *American Economic Review*, 2016, *106* (8), 2145–84.

Bundorf, M Kate, Jonathan Levin, and Neale Mahoney, "Pricing and welfare in health plan choice," *American Economic Review*, 2012, *102* (7), 3214–48.

Cardell, N Scott, "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 1997, *13* (2), 185–213.

Decarolis, Francesco, Maria Polyakova, and Stephen P Ryan, "Subsidy design in privately provided social insurance: Lessons from Medicare Part D," *Journal of Political Economy*, 2020, *128* (5), 1712–1752.

Einav, Liran, Amy Finkelstein, and Jonathan Levin, "Beyond testing: Empirical models of insurance markets," *Annu. Rev. Econ.*, 2010, *2* (1), 311–336.

Ericson, Keith M Marzilli and Amanda Starc, "How product standardization affects choice: Evidence from the Massachusetts Health Insurance Exchange," *Journal of Health Economics*, 2016, *50*, 71–85.

Galichon, Alfred, "On the representation of the nested logit model," *Econometric Theory*, 2022, *38* (2), 370–380.

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

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou, "Inattention and switching costs as sources of inertia in medicare part d," Technical Report, National Bureau of Economic Research 2016.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton, "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 2017, *48* (4), 877–905.

Matějka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 2015, *105* (1), 272–98.

Polyakova, Maria, "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 2016, *8* (3), 165–95.

Small, Kenneth A. and Harvey S. Rosen, "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 1981, *49* (1), 105–130.

Tebaldi, Pietro, "Estimating Equilibrium in Health Insurance Exchanges:Price Competition and Subsidy Design under the ACA," Working Paper 2017.