Endogenous Information and Simplifying Insurance Choice*

Zach Y. Brown
University of Michigan
& NBER

Jihye Jeon
Boston University

January 2022

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. In contrast to standard demand models, counterfactual simulations imply that simplifying choice by reducing the number of plans can raise welfare. Capping out-of-pocket costs also generates larger welfare gains than standard models. The empirical model can be applied to other settings to examine regulation of complex products.

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

*We thank Ying Fan, 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 Boston University, ASSA, IIOC, University of Michigan, SITE, NBER Summer Institute, Boston IO, Yale, MIT, Harvard, SMU, WashU, CeMENT Workshop, NBER IO, and Princeton. 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. 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. Individuals may choose to spend significant time and effort researching each plan's potential out-of-pocket costs given their risk. Depending on each individual's incentive to do research, some may choose dominated options, with important implications for regulation of insurance markets.

We develop a tractable micro-founded framework for examining demand when individuals choose how much information to acquire. We focus on insurance choice, although the model can be applied more broadly. For instance, similar issues arise in consumer finance markets. 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, also known as Medicare Part D, we find evidence consistent with our model's predictions, motivating an empirical model incorporating costly information acquisition. The model also compliments recent work documenting choice frictions in insurance markets and other markets featuring complex products. We show that the welfare implications of the model are quite different than commonly used empirical models of insurance demand.

The model builds on theoretical work incorporating rational inattention in discrete choice models (Matějka and McKay 2015). While there is a growing theoretical literature on rational inattention, there are two key issues making it challenging to incorporate in an empirical model. First, it is difficult to separately identify heterogeneous preferences and information frictions. Leveraging results from Matějka and McKay (2015), we develop a model to account for the case in which some characteristics are always observed, which is key for identification in our empirical setting. Second, a fundamental challenge of the rational inattention framework is that the complexity of the model generally makes estimation infeasible. We derive a novel analytical solution for choice probabilities that incorporates preference heterogeneity, allowing for a feasible estimation strategy. Our paper is the first to use this method to estimate a model based on rational inattention. We argue this approach can be used more generally in order to provide empirical insight into consumer protection interventions in complex markets.

We begin by presenting a simple theoretical framework. In the model, individuals decide how much to research their options given their prior beliefs and easy-to-observe information such as plan premiums. The more research individuals do, the more accurate their beliefs will tend to be about out-of-pocket costs or other difficult-to-observe characteristics. They then choose an insurance plan to maximize expected utility given their resulting beliefs.

A key implication of the model is that the amount of information acquired by individuals depends on the stakes. Individuals with small consequences from choosing the wrong plan, such as healthy individuals expecting few claims, acquire less information than individuals with large consequences. When the stakes are very low, individuals tend to make the correct choice simply by comparing easy-to-observe characteristics, despite the small amount of information acquisition. When the stakes are very high, individuals have a strong incentive to acquire information, also resulting in choices that tend to be correct ex-post. Therefore, it is those facing moderate stakes that are least likely to choose the utility-maximizing option given their beliefs, implying a non-monotonic relationship between the stakes and the quality of choices that individuals make. In addition, the model implies that, when choosing a plan, the relative weight that individuals appear to place on premiums versus out-of-pocket cost depends on the stakes. As the stakes increase, individuals appear to be more sensitive to out-of-pocket cost since they acquire more information and therefore have more accurate beliefs.

We start by documenting descriptive evidence consistent with the model. We leverage administrative data from Medicare Part D. Focusing on individuals that are forced to make an active choice, i.e. new enrollees and those who had a previous plan that was discontinued, 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 the individual is expecting to be in the Medicare Part D coverage gap, the individual makes choices that are consistent with having acquired more information. In particular, the individual's demand is more elastic with respect to out-of-pocket cost in these years. These results are difficult to rationalize with standard models of insurance demand that have been previously used in the literature.

Motivated by the reduced-form evidence, we develop an empirical model that generalizes the standard discrete choice model by incorporating rational inattention over a subset of product characteristics. By incorporating heterogeneous preferences, including a taste shock, the model allows for the fact that individuals may not always choose the plan with the lowest cost or highest quality even if they have full information. Using the model, we 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 those with previous experience choosing a plans.

Empirical results imply that endogenous information frictions play an important role in our setting. The marginal cost of information is especially high for older enrollees and those with little prior experience choosing Medicare Part D plans. If individuals had full information, they would choose plans that had somewhat higher premiums in exchange for significantly lower out-of-pocket costs. Average annual premiums would increase from $570 to $642 but annual out-of-pocket costs would decline from $713 to $601. In addition, information frictions also cause individuals to choose plans with suboptimal quality and risk, implying that welfare effects are larger than the savings. Estimates imply that full information would generate annual welfare gains of $285 per individual excluding information acquisition costs. The average annual information acquisition costs are $127 per enrollee for those making active choices. Given heterogeneity in the unit cost of acquiring information and the incentives to acquire information, there is large variation in the total cost of information that individuals incur.

We examine the implications of the model for restricting the choice set of individuals. Policy makers often set minimum standards for insurance plans, implicitly limiting available plans.[1] In standard demand models, restricting the choice set strictly decreases welfare,

[1] This is also closely related to standardization of health exchanges. See Ericson and Starc (2016).

which seems at odds with individuals' strong desire for a reduced and simplified choice set as documented in existing surveys.[2] By contrast, the model presented in this paper implies that a policy maker can increase consumer welfare by showing individuals only a selected subset of plans. This simplifies choice, reducing the amount of research that individuals need to do and reducing the probability of choosing plans that an individual would not have chosen if they had been fully informed. We simulate demand with endogenous information under a counterfactual policy in which individuals are shown a subset of plans based on their age. Removing a quarter of the plans with the lowest 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.

Another potential way to increase consumer welfare is to limit out-of-pocket costs in order to ensure that individuals do not accidentally choose plans with high cost. This issue is particularly important given that cost sharing has been growing for health insurance plans, including Medicare prescription drug plans.[3] This has motivated policy makers to propose caps on out-of-pocket spending in Medicare Part D. To provide insight into the role of endogenous information, we examine counterfactual demand when out-of-pocket costs are capped in the model. We find that imposing the cap has two effects: it lowers information acquisition costs and reduces the probability that individuals "accidentally" choose high out-of-pocket cost plans. For these reasons, a $5,000 out-of-pocket cap increases welfare more than what is 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 protecting consumers in markets featuring complex choices.

[2]For example, Altman et al. (2006) conduct a survey and find 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 simplifying the new benefit by reducing the number of available plans and 60% agree with the statement that Medicare should select a handful of plans that meet certain standards, so seniors have an easier time choosing.

[3]In contrast, premiums from Medicare Part D have remained fairly constant. As a result, cost sharing defined as out-of-pocket cost as a fraction of gross drug cost has increased from 0.29 to 0.32 from 2010 to 2015.

1.1 Related Literature

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. This result is further generalized by Fosgerau et al. (2020). Other theoretical work incorporating rational inattention in a discrete choice framework includes Caplin et al. (2016) and Caplin et al. (2019). In this framework, decision makers choose how much and what type of information to acquire. Given a cost of acquiring information, individuals optimally learn about the payoff structure of various options. There is limited work testing the rational inattention framework in real-world settings.[4] There is also very limited work incorporating the rational inattention framework into structural models. One exception is recent work by Joo (2020) who uses the rational inattention framework to examine the effect of a new product introduction. Unlike previous work, we develop a tractable model allowing for both observed and initially unobserved characteristics, which is a key feature of many consumer finance markets. To our knowledge, we are the first to test and empirically estimate a model of rational inattention in an insurance context.

Our work is related to the large literature on choice frictions in health insurance markets. There is an influential literature documenting that individuals choose dominated health insurance plans, often overpaying significantly (e.g. Abaluck and Gruber 2011; Heiss et al. 2013; Bhargava et al. 2017). It has been argued that this is due to the complexity of health insurance plans and the fact that individuals are not using all available information. For instance, Handel et al. (2019) survey individuals choosing health insurance and find that they do not fully understand the insurance plans, making it difficult to choose correctly. Handel and Kolstad (2015) examine the resulting implications for regulation of insurance markets under perfect competition. There is also evidence that consumers respond to easy-to-use information (Kling et al. 2012; Bundorf et al. 2019). There are many other papers that assess the rationality of individual choices and document the presence of information frictions specifically for Part D markets. In particular, our paper compliments Ketcham et al. (2012), Ketcham et al. (2015), and Keane et al. (2019). Some papers argue inattention may be a driving force of inertia in insurance plan choice (e.g. Handel 2013; Polyakova 2016; Heiss et al. 2016; Ho et al. 2017). Papers that evaluate policies that reduce choice in Part

[4] Bhattacharya and Howard (2020) examine rational inattention in a professional sports context. There is also work testing rational inattention in laboratory experiments (e.g. Dean and Neligh 2019).

D include Lucarelli et al. (2012) and Ketcham et al. (2019). In recent work, Abaluck and Gruber (2020) finds that individuals make better decisions with smaller choice sets. Coughlin (2019) explores a consideration set model of insurance choice in which non-monetary plan attributes determine the set of plans that an individual considers, and monetary plan attributes determine the individual's expected utility for options in her consideration set.

In contrast to this literature, we focus on the role of endogenous information. We show that it is important to account for the fact that individuals choose how much information to acquire and allow information to endogenously adjust in counterfactual simulations. We contrast the implications of our model with commonly used demand models in the literature. In important cases, welfare effects of policies can have the opposite sign depending on the underlying model. For these reasons, we argue that testing and developing a micro-founded model incorporating endogenous information is important for providing insight into policy. Furthermore, the model developed in this paper can be applied to other settings involving complex products where individuals decide how much to research product characteristics.

Finally, our approach is related to the literature on consumer search (Stigler 1961; Diamond 1971). The search framework has been incorporated into empirical demand models and applied to a variety of markets (for instance, see Hortaçsu and Syverson (2004), Hong and Shum (2006), De Los Santos et al. (2012), Seiler (2013)). In particular, Honka (2014) and Honka and Chintagunta (2017) estimates a model of car insurance demand incorporating search costs. Cebul et al. (2011) analyze the role of search frictions when employers choose health insurance plan for their employees. In search models, individuals generally start with full information about one option and then pay a cost to become fully informed about other options in their choice set. One implication is that consumers will, at a minimum, have full information about the option they choose. In contrast to standard search models, individuals may choose to acquire partial information about any of the options in their choice set. This is consistent with the evidence that individuals are often not fully informed about health insurance plans, including their chosen option (Handel et al. 2019). In addition, our model focuses on the case in which some characteristics are easier to observe than others. We argue that this is a key feature of insurance markets—premiums are easy to compare but out-of-pocket cost are not. In general, search models are well suited to situations with a large number of simple options while the rational inattention approach is useful for analyzing markets with complicated product attributes. As noted by Matějka and McKay (2012), these models can have quite different implications.

The remainder of this paper is as follows. Section 2 presents the basic framework. Section 3 discusses background and data. Section 4 presents reduced-form evidence consistent with the model. Section 5 presents an empirical framework and Section 6 presents counterfactual results. Section 7 concludes.

2 Theoretical Framework

In this section, we present a basic version of the discrete choice model in which individuals minimize expected total cost when part of the cost, i.e. out-of-pocket costs, are initially unobserved unless individuals acquire costly information. We leverage theoretical results linking the rational inattention framework with discrete choice models (Matějka and McKay 2015; Fosgerau et al. 2020). This literature focuses 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. In addition, we show that, under relatively innocuous assumptions, one can derive a straightforward expression for choice probabilities.

To fix ideas, we start by abstracting from risk aversion and preferences over non-cost characteristics. The results from the simple theoretical framework motivate our reduced-form analysis in the following section. In Section 5, we build on the framework and present a richer empirical model that accounts for individual risk aversion and heterogeneous preferences, including an idiosyncratic taste shocks.

Individual i chooses between N alternatives indexed by j. Each alternative has cost p_j, which is initially observed, and v_{ij}, which is initially unobserved unless the individual acquires costly information. The vector of payoffs, $\mathbf{u}_i \in \mathbb{R}^N$, is determined by the vector of observed cost, $\mathbf{p} \in \mathbb{R}^N$, and initially unobserved cost, $\mathbf{v_i} \in \mathbb{R}^N$. Specifically,

$$u_{ij} = \underbrace{-p_j}_{\substack{\text{Initially}\\\text{Observed}\\\text{Cost}}} \underbrace{-v_{ij}}_{\substack{\text{Initially}\\\text{Unobserved}\\\text{Cost}}} \tag{1}$$

In the case of insurance choice, p_j is the premium and v_{ij} is expected out-of-pocket costs. 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.

Following Matějka and McKay (2015), we can consider the decision problem having two stages. In the first stage, individuals have a prior and rationally choose how much information to acquire about v_{ij}, forming posterior beliefs about the total cost of each option. In the second stage, individuals maximize expected utility given beliefs that were formed in the first stage.

We start with the second stage decision. After acquiring the chosen amount of information, the individual has beliefs $B_i \in \Delta(\mathbb{R}^N)$ about the expected payoff of each option where the set of all probability distributions is given by $\Delta\mathbb{R}^N$. The individual chooses the option that maximized utility given beliefs:

$$V(B) = \max_{j \in J} \left[-p_j - \mathbb{E}_{B_i}[v_{ij}] \right].$$

In the first stage, the individual chooses what signals to receive based on the expected payoff, the cost of information, and the prior. The individual's potential information acquisition strategies are unconstrained—any information about any of the options can be acquired in any manner, subject to the cost of information. In particular, individuals may wish to become partially informed about options, i.e. receive vector of signals, $\mathbf{s}_i$, with limited information content.

The information strategy can be expressed as a joint distribution of signals and payoffs, $F(\mathbf{s}_i, \mathbf{v}_i) \in \Delta(\mathbb{R}^{2N})$. Given the individual's prior, G_i, the individual chooses the conditional distribution $F(\mathbf{s}_i|\mathbf{v}_i)$. This results in posterior belief $F(\mathbf{v}_i|\mathbf{s}_i)$.

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

$$c(F) = \lambda \left(H(G_i) - \mathbb{E}_{\mathbf{s}_i}[H(F(\cdot|\mathbf{s}_i))] \right) \tag{2}$$

where $H(F)$ 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]

The individual chooses an information acquisition strategy that solves

$$\max_{F(\mathbf{s}_i,\mathbf{v}_i)\in\Delta(\mathbb{R}^{2N})} \int_{\mathbf{v}_i}\int_{\mathbf{s}_i} V(F(\cdot\,|\mathbf{s}_i))F(d\mathbf{s}_i|\mathbf{v}_i)G(d\mathbf{v}_i) - c(F) \tag{3}$$

$$\text{s.t.} \quad \int_{\mathbf{s}_i} F(d\mathbf{s}_i,\mathbf{v}_i) = G(\mathbf{v}_i)\ \forall \mathbf{v}_i \in \mathbb{R}$$

Matějka and McKay (2015) show that the optimal strategy results in choice probabilities that are closely related to the multinomial logit model, reflecting both the true payoffs and prior beliefs.

$$P_{ij} = \frac{P^0_{ij}e^{(-p_j-v_{ij})/\lambda}}{\sum_{j=1}^N P^0_{ij}e^{(-p_j-v_{ij})/\lambda}} \tag{4}$$

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

$$\max_{P^0_{i1},..,P^0_{iN}} \int_{\mathbf{v}} \lambda \log \Sigma_j P^0_{ij} e^{(-p_j-v_{ij})/\lambda} G(d\mathbf{v}) \text{ s.t. } \sum_j P^0_{ij} = 1,\ P^0_{ij} \geq 0\ \forall j. \tag{5}$$

The choice probabilities in equation (4) imply that it is as if individuals maximize utility given by

$$\tilde{u}_{ij} = \underbrace{-p_j - v_{ij}}_{\text{Actual Utility}} + \underbrace{\lambda \log P^0_{ij}}_{\text{Contribution of Prior}} + \underbrace{\lambda e_{ij}}_{\text{Belief Error}} \tag{6}$$

where e_{ij} is distributed EV1. The distribution of the belief error is not an assumption, but rather a natural consequence of the rational inattention framework.

We assume that individuals have a common prior on out-of-pocket cost for all of the options in their choice set. The variance of this common prior is given by σ^2, a key parameter we describe in greater detail below. In section Section 5, we generalize the model to account for heterogeneous prior mean across options.

Although this model provides a simple analytical structure compared to the original continuous-action formulation of the rational inattention model, it still poses a significant challenge for empirical application. First, equation (5) makes it difficult to interpret demand

[5]For example, the cost function is consistent with an individual asking a series of yes-no questions with a fixed cost per question. The individual is completely free to choose what questions to ask and is free to focus on some specific aspect of the environment while ignoring other aspects. 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.

in this framework. Second, estimating the empirical model presented in Section 5 would be infeasible given the high dimensional integration involved in solving for P^0_{ij}.

We develop a tractable model of demand with endogenous information. We do this by assuming that the distribution of the prior, $G(\mathbf{v})$, follows the conjugate to the EV1 distribution.[6] This leads to choice probabilities that take a relatively simple form given by

$$P_{ij} = \frac{e^{(-p_j\ell/(\ell-1)-v_{ij})/\lambda}}{\sum_k e^{(-p_k\ell/(\ell-1)-v_{ik})/\lambda}} \tag{7}$$

where $\ell^2 \equiv \frac{6\sigma^2}{\pi^2\lambda^2}+1$. We present the derivation of equation (7) and the discussion of our distributional assumption in Appendix A. In Appendix H, 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.

Given the above expression for choice probabilities, expected utility can be expressed as

$$\tilde{u}_{ij} = -p_j - \underbrace{\frac{(\ell-1)}{\ell}\left(v_{ij} - \lambda e_{ij}\right)}_{\mathbb{E}_{B_i}[v_{ij}]}. \tag{8}$$

The expected out-of-pocket cost, $\mathbb{E}_{B_i}[v_{ij}]$, depends on both the variance of an individual's prior and the cost of information. Alternatively, the error term can be normalized and expected utility can be written

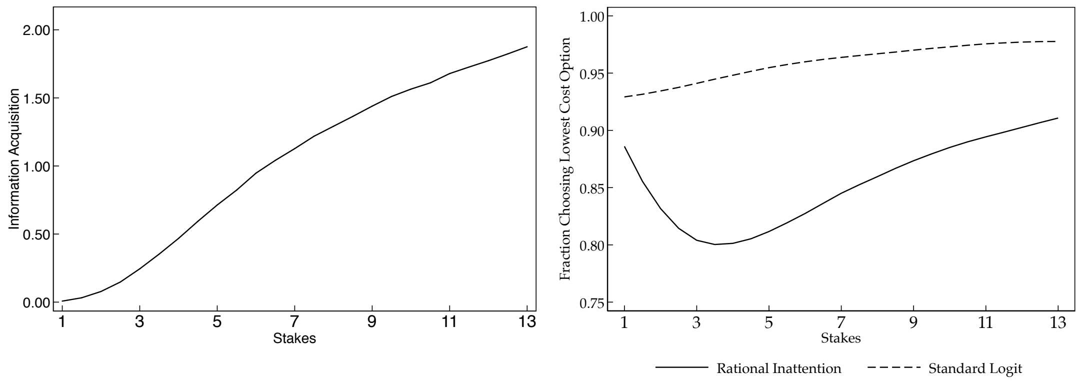
$$\tilde{u}'_{ij} = -\underbrace{\frac{\ell}{\lambda(\ell-1)}}_{\substack{\text{Premium}\\ \text{Weight}}} p_j - \underbrace{\frac{1}{\lambda}}_{\substack{\text{OOP}\\ \text{Weight}}} v_{ij} + \underbrace{e_{ij}}_{\substack{\text{Normalized}\\ \text{Belief Error}}}. \tag{9}$$

Even though payoffs are deterministic in this simple version of the model, it is as if choices are the result of a random utility model. Rather than a taste shock, the idiosyncratic error is due to endogenous information frictions. Note that e_{ij} is normalized and has scale parameter 1 and is distributed iid EV1.

As can be seen in equation (7), the model implies that choices depend on the standard deviation of the prior, σ, which we interpret as a measure of the stakes. When individuals

[6] This implies that when v 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) for details about this distribution. This distribution is also an integral part of the nested logit demand system. See discussion in Berry (1994).

Figure 1
Predicated Information Acquisition and Fraction Choosing Lowest Cost Plan by Stakes



a. Information Acquisition by Stakes

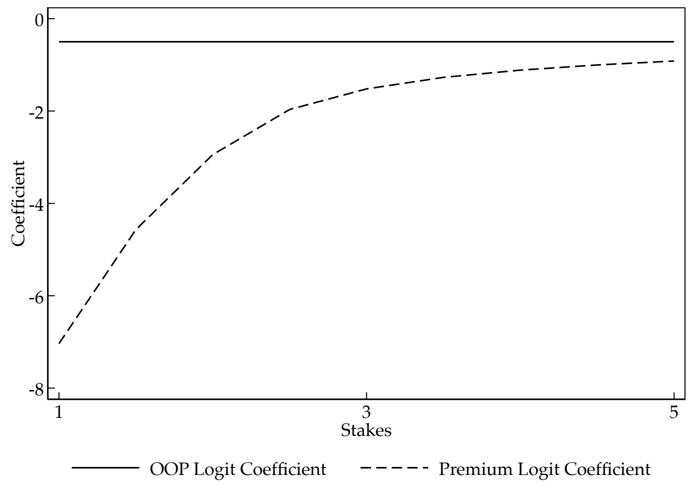
b. Fraction Choosing Lowest Cost Plan by Stakes

Notes: Charts show total information acquisition cost and mean fraction of individuals choosing lowest cost option from simulations varying the stakes, i.e. the standard deviation of the prior. Simulation assumes 3 options, $\lambda = 2$, out-of-pocket cost standard deviation of 10, premium standard deviation of 4, and prior standard deviation for out-of-pocket costs (σ) of 10.

have a less precise prior, i.e. when σ 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. This is depicted graphically in Figure 1 Panel a. In the figure, information acquisition is simulated for different values of the stakes using equation (2).

Endogenous information acquisition has important implications for choice quality and overspending. Figure 1 Panel b 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.

Figure 2
Predicted Logit Coefficient on Premium and Out-of-Pocket cost by Stakes



Notes: Chart shows implied logit coefficient on annual out-of-pocket cost and annual premium from simulations based on endogenous information model with 3 options, $\lambda = 2$, out-of-pocket cost standard deviation of 10, premium standard deviation of 4, and prior standard deviation for out-of-pocket costs (σ) of 10.

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, as in equation (1), the stakes will have no effect on choices. In a logit demand model with a taste shock, there is a monotonic relationship between stakes and probability of choosing the least expensive plan. As the stakes grow, the taste shock becomes less important, generating a positive relationship. This can be seen in Figure 1 Panel b.

Moreover, the model has stark predictions for the effective weight that decision makers place on p_j and v_{ij}. Under full information, a change in p_j affects choices the same as an equivalent change in v_{ij}, i.e. the elasticity of demand is the same for premium and expected out-of-pocket cost. However, in the demand model with endogenous information acquisition where v_{ij} is initially unobserved, the weight that individuals appear to place on characteristics is endogenous and differs for p_j and v_{ij}. In equation (9), the coefficient on v_{ij} is solely a function of the marginal cost of information; however the coefficient on p_j depends on both the marginal cost of information and the stakes. As shown in Figure 2, the magnitude of the the coefficient on p_j decreases when the stakes increase. As individuals acquire more information about v_{ij}, the weight on p_j and v_{ij} converge. Consequently, elasticity of demand with respect to premium differs from elasticity of demand with respect to expected out-of-pocket cost.[7] These elasticities converge as the stakes increase or the marginal cost of information decreases.

Finally, the endogenous information model can be contrasted with alternative models featuring behavior that is not rational. Handel and Schwartzstein (2018) note that the failure of individuals to use all available information could reflect information frictions, as in the rational inattention framework, or some other psychological distortion or mental gap. Although there are many possible models of psychological distortions, it is not clear why these distortions would be a function of the stakes. For instance, if individuals have a behavioral bias in which they always ignore out-of-pocket cost, the weight that individuals place on out-of-pocket costs and premiums should not converge as the stakes increase and one would not expect the relationships seen in Figure 1 and Figure 2.

In Section 4 we examine the predictions of the model in the context of Medicare prescription drug insurance choice. In particular, we ask whether choices are affected by the stakes in a manner consistent with the model presented in this section, motivating a structural model that accounts for endogenous information.

3 Background and Data

Many markets feature opaque product characteristics that are complicated to understand. The difficulty in comparing cost across options is especially relevant for insurance, including health insurance, car insurance, and life insurance. For our application, we focus on Medicare prescription drug insurance, known as Medicare part D. When individuals choose a Medicare prescription drug plan, it is 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, potentially requiring costly effort. First, individuals must know their likely drug usage over the coming year, including dosage and frequency. Then 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

[7] The elasticities are derived in Appendix A.

with complex health care needs to research their Medicare plans.[8]

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.[9] 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. To the extent that the PlanFinder aids consumer choice, we would expect the cost of information to be lower in the market for Medicare Part D plans relative to other insurance markets.

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. The large sample size allows us to construct more precise estimates of expected out-of-pocket costs. We focus on the period starting in 2010 since this is the period in which we have detailed drug formulary data. This allows us to more accurately construct out-of-pocket costs.

In the context of our model, we wish to construct a measure of expected out-of-pocket costs that reflects the beliefs of individuals as the marginal cost of information goes to zero (or information acquisition goes to infinity). Following Abaluck and Gruber (2016), we construct two measures of out-of-pocket cost. The primary measure, based on the rational expectations assumption, is constructed by binning individuals into groups based on similarity and then constructing out-of-pocket costs for each individual for each plan in their choice set by applying the plan's formulary and cost sharing rules to observed drug utilization in the chosen plan. As in Abaluck and Gruber (2016), we allow for substitution to equivalent drugs in less expensive tiers. Then out-of-pocket costs for each plan are averaged across individuals in the group to obtain an estimate of expected out-of-pocket cost. 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 Appendix B.

Abaluck and Gruber (2016) validate their Part D calculator and show that estimated

[8]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.

[9]See, for instance, McGarry et al. (2018).

[10]See Altman et al. (2006) and Cummings et al. (2009).

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	

expected out-of-pocket costs for chosen plans are quite close to actual out-of-pocket costs. Nevertheless, there is concern about measurement error, and therefore we also construct an alternative measure of out-of-pocket costs based on a perfect foresight assumption also following Abaluck and Gruber (2016). The perfect foresight measure assumes that, with full information, individuals would know their future utilization exactly. Therefore, each individual's realized claims is used to construct out-of-pocket costs. This approach abstracts from moral hazard.

The previous literature has documented the importance of consumer inertia in plan choice (e.g. Handel 2013; Polyakova 2016; Ho et al. 2017). Following 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. 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.[11] Importantly, these individuals have not previous chosen any of the plans in their choice set, implying that they are unlikely to start with information specific to certain plans. 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. Individuals forced to make an active choice constitute 22.0 percent of the sample. 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.[12] We use a 5 percent sample for the reduced-form analysis, which includes 90,200 individuals and 206,891 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 reduced-form analysis. The claims data contain information on age and sex of each individual. We also construct indicators for the most common chronic conditions. In addition, we use individuals' zip code to merge on 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.[13]

Table 1 also shows the calculated out-of-pocket costs for the two measures, rational expectations (RE) and perfect foresight (PF). 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.

We now turn to the definition of the stakes used in the empirical analysis. Individuals may understand the variance of v_{ij} across alternatives, forming the basis of their prior G. For example, those that currently take 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. Therefore, they know the stakes are high. 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.[14] Therefore, the stakes are low when expected out-of-pocket costs are similar across

[11] Appendix Table A-2 compares demographics of all Medicare Part D enrollees, active choice makers, and new enrollees.

[12] 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.

[13] See Appendix Table A-2.

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

plans, perhaps because the individual expects to have few claims or individuals know that plans have similar coverage. In this case, the model predicts that individuals acquire little information. Conversely, individuals have more incentive to acquire information when out-of-pocket costs differ widely across plans since there is more scope for accidentally choosing an expensive plan.

This measure of the stakes is significantly correlated with health, including whether a patient has a chronic condition.[15] 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.

Individuals face mean stakes of $204. Figure A-1 shows that there is significant variation in the stakes across choice situations. Panel b shows that individuals also face significant variation in the stakes across years, either due to changes in health status or changes in the plans available.

4 Motivating Evidence

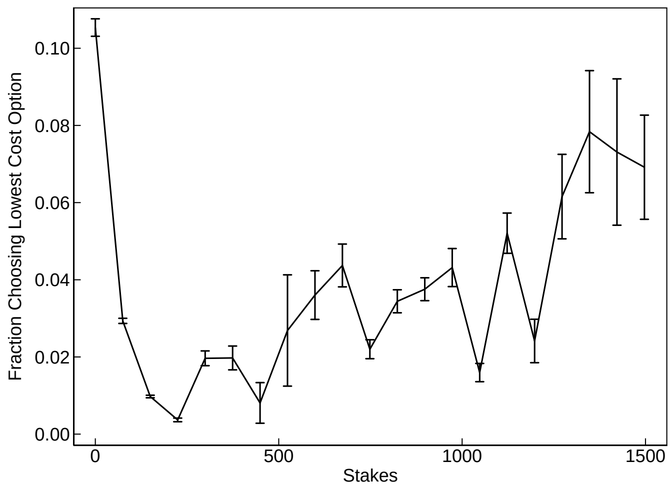
Motivated by the results of the model in Section 2, we now examine how insurance plan choice is affected by the stakes. We use individual-level data on Medicare prescription drug plan choice and exploit within-individual variation, i.e. the same individual who makes an active plan choice facing different stakes.

Stakes and Overspending

We start by examining the relationship between the fraction of individuals choosing the lowest cost plan and the stakes. Figure 3 shows that there is a non-monotonic relationship. As in Figure 1 Panel b, the relationship is U-shaped. When individuals face very low stakes and rely heavily on easily observed characteristics such as the premium, individuals are more likely to make optimal choices despite the low research efforts, since there is little variation in difficult-to-observe characteristics across options. Individuals are also more likely to make optimal choices when stakes are very high such that there are high incentives to acquire information. We interpret this as initial evidence in support of the model. However,

[15] See Appendix Table A-1.

Figure 3
Fraction Choosing Lowest Cost Plan by Stakes



Notes: Chart shows mean fraction of individuals choosing lowest cost option. Standard error bars show 95% confidence interval for the mean.

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 to examine the relationship between stakes and choice quality by focusing on the sample of individuals making active choices multiple times in the sample period. For individual i in year t, we estimate the following linear probability model

$$y_{it} = \beta_0 + \alpha_1 Stakes_{it} + \alpha_2 Stakes_{it}^2 + \beta X_{it} + \gamma_i + \theta_t + \varepsilon_{it} \tag{10}$$

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. In addition, we control for within-plan out-of-pocket cost variance to account for risk-aversion. By including individual fixed effects, identification of α_1 and α_2 exploits within-individual variation in the stakes. Year fixed effects control for changes in plans offered over the period. The dependent variable, y_{it}, is an indicator for whether individual i chose the option with the lowest total cost, multiplied by 100, 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. The primary hypothesis is that there is a U-shaped relationship

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

	(1)	(2)	(3)	(4)	(5)	(6)
Stakes (100s)	-2.415^{***}	-2.340^{***}	-0.466^{***}	-2.144^{***}		
	(0.054)	(0.056)	(0.079)	(0.057)		
Stakes Squared	0.217^{***}	0.211^{***}	0.062^{***}	0.195^{***}		
	(0.005)	(0.005)	(0.007)	(0.005)		
Stakes quintile 2					-5.286^{***}	-1.066^{***}
					(0.114)	(0.150)
Stakes quintile 3					-6.074^{***}	-1.856^{***}
					(0.110)	(0.146)
Stakes quintile 4					-6.418^{***}	-2.370^{***}
					(0.117)	(0.163)
Stakes quintile 5					-5.102^{***}	-1.063^{***}
					(0.114)	(0.184)
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	557.3	555.3	375.9	550.6		
Adjusted R2	0.010	0.013	0.304	0.016	0.026	0.305
Observations	206,891	200,526	189,870	193,745	200,526	189,870

Notes: Estimates from linear probability model where dependent variable is the indicator variable for whether the individual chooses the lowest cost plan. Standard errors in parentheses, multiplied by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 $\alpha_2 > 0$, implying a U-shaped relationship. 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 column 5 and column 6, we allow 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. In Table 2 Column 4, we include market fixed effects in order to examine the effect of within-market variation in the stakes and also find a U-shaped relationship.[16]

One concern is that there could be measurement error stemming from the fact that each individual's out-of-pocket costs are predicted based on the average of similar individuals. However, measurement error is likely to attenuate the relationship between stakes and choice quality rather than result in a U-shaped relationship. As an additional robustness check, we use each individual's actual utilization to predict out-of-pocket costs, i.e. assume perfect foresight. The regression results are presented in Appendix Table A-3. All of the specifications also imply a U-shaped relationship, although the standard errors are slightly larger. Finally, there is concern that new enrollees may have different behavior. In Appendix Table A-4, we find that the U-shaped relationship is even more pronounced for first-time enrollees.

The fraction of individuals choosing the lowest cost plan is only one measure of choice quality. In Appendix Figure A-2 we examine the fraction of individuals choosing a plan in the lowest decile and lowest quintile of the plans in their choice set. We also examine the average percentile rank of individuals' chosen plan. In addition, to the extent that plan riskiness and quality are also initially unobserved unless individuals conduct costly research, we would expect a similar U-shaped relationship between these outcomes and stakes. We examine these outcomes in Appendix Figure A-3. Across all of these alternative outcomes, we find evidence of a U-shaped relationship between the stakes and choice quality.

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. The model is "reduced-form" in the sense that we do not incorporate the cost of information. In Section 5, we estimate a demand model that is directly based on the rational inattention framework.

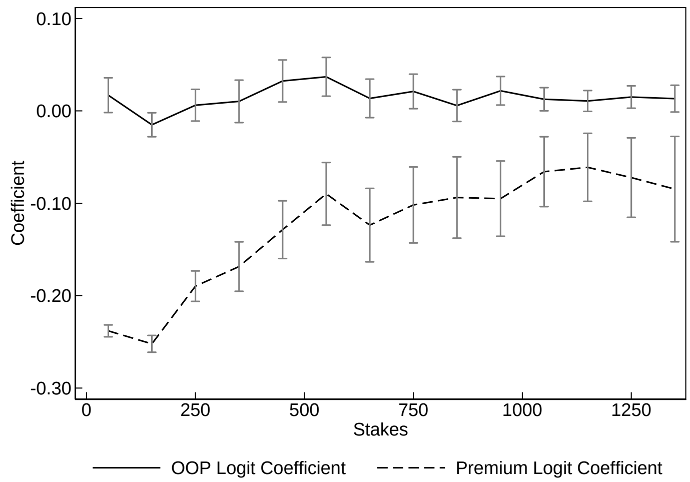
We start by considering the following specification for observable utility of plan j

$$\nu_{ijt} = \alpha_1 p_{jt} + \alpha_2 p_{jt} Stakes_{it} + \gamma_1 v_{ijt} + \gamma_2 v_{jt} Stakes_{it} + \theta \tilde{\sigma}^2_{ijt} + \beta X_{ijt} \tag{11}$$

The specification controls for risk aversion by including $\tilde{\sigma}^2_{ijt}$, the variance of out-of-pocket

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

Figure 4
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 (OOP variance), plan quality rating, deductible, generic coverage, coverage in the donut hole, and cost sharing. Standard error bars show 95% confidence interval.

costs for plan j.[17]. We also include other plan characteristics, X_{ijt}. Given additive i.i.d. EV1 error, choice probabilities are $P_{ijt} = \exp[\nu_{ijt}] / (\sum_k \exp[\nu_{ikt}])$.

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 4 presents the results in graphical form by interacting stake bins with coefficients on premium and out-of-pocket cost.[18] 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

[17]This can be derived by considering a first-order Taylor series expansion when individuals have CARA utility.

[18]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 $\times$ Indiv. avg stakes				0.019^{***}	0.017^{***}	0.017^{***}
				(0.001)	(0.001)	(0.001)
Premium $\times$ Stakes		0.020^{***}	0.017^{***}	0.008^{***}	0.007^{***}	
		(0.001)	(0.001)	(0.001)	(0.001)	
Premium $\times$ Stakes $\times$ $\mathbb{1}(\Delta > 0)$						0.005^{***}
						(0.001)
Premium $\times$ Stakes $\times$ $\mathbb{1}(\Delta < 0)$						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 $\times$ Indiv. avg stakes				0.000	0.000	0.000
				(0.000)	(0.000)	(0.000)
OOP $\times$ Stakes		-0.000	-0.000	-0.000	-0.001^{**}	
		(0.000)	(0.000)	(0.000)	(0.000)	
OOP $\times$ Stakes $\times$ $\mathbb{1}(\Delta > 0)$						-0.001^{**}
						(0.000)
OOP $\times$ Stakes $\times$ $\mathbb{1}(\Delta < 0)$						-0.000
						(0.000)
Premium $\times$ Z_i	No	No	Yes	No	Yes	Yes
OOP $\times$ 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: 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 2.

The results using the specification described in equation (11) 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 the same.

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 coefficient on $p_{jt} \times Stakes_{it}$ and $v_{jt} \times Stakes_{it}$. The results, with and without the interaction of observable characteristics, are presented in Table 3 Column 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. The interaction of out-of-pocket cost and within-individual stakes remains small in magnitude. This provides additional evidence in support of the endogenous information model.

As an additional robustness check, we allow for additional heterogeneity in preferences by including a random-coefficient on premium and out-of-pocket cost. The results, which are very similar to the baseline specification, are presented in Appendix Table A-5. We also examine the results using the alternative definition of out-of-pocket cost, i.e. assuming perfect foresight. Results, presented in Appendix Table A-6, are also qualitatively similar.

When exploiting within-individual variation, we are relying on individuals getting older, thus getting more experience in choosing a plan in addition to becoming sicker and facing higher stakes on average. To address the concern that it is primarily the experience that is helping the individuals making an improved choice, rather than more research due to facing higher stakes, we conduct various robustness checks. First, we examine the relationship between stakes and choice quality, while restricting the sample to first-time enrollees only (see Table A-4) and the first choice made by each individual (see Table **??**), respectively, and find that the results are robust. Additionally, in column (6) of Table 3, we estimate separate coefficients when the stakes increase versus decrease. We find qualitatively similar results for an increase versus a decrease in the stakes. 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 implies that there is an important relationship between the stakes and choice quality. This relationship still holds controlling for individuals fixed effects which control for unobserved time-invariant characteristics. 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 based the rational inattention framework. The empirical model allows us to quantify welfare effects of information frictions and provide insight into the implications of endogenous information for simplifying choice.

5 Empirical Model

In this section, we develop an empirical model of demand with endogenous information acquisition that incorporates preferences over non-pecuniary plan characteristics including plan quality and riskiness. In addition, we generalize the simple model in Section 2 to incorporate an idiosyncratic taste shock. Incorporation of a taste shock is important for capturing unobserved preferences which could explain why individuals do not choose cost-minimizing plans. In this way, the model seeks to identify the degree to which individuals choose expensive plans due to preferences over non-price characteristics versus information frictions.

Consider individual i choosing plan $j \in \mathcal{J}_{it}$ in year t where the choice set is defined by $\mathcal{J}_{it}$. Following the previous literature, we map a CARA utility function into a conditional logit model while adding preferences over non-cost characteristics, writing the indirect utility as

$$u_{ijt} = \underbrace{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}_{\text{Initially Unknown}} + \underbrace{\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt}}_{\text{Known}} \tag{12}$$

where p_{jt} is the premium, v_{ijt} is expected out-of-pocket cost assuming rational expectations, and $\widetilde{\sigma^2_{ijt}}$ is the within-plan variance of out-of-pocket cost.[19]

As in Section 2, a key assumption is that p_{jt} is initially observed while v_{ijt} is initially unknown unless individuals choose to acquire costly information. The model can accommodate plan characteristics that are initially unknown, X^u_{jt}, as well as plan characteristics that are initially known, X^k_{jt}. We assume that plan quality, as measured by star ratings, is initially unobserved. Plan risk is also difficult to observe; it also requires knowing all contract terms. For this reason, we assume that $\widetilde{\sigma^2_{ijt}}$ is initially unknown and also requires costly information acquisition. Since the unobserved part of utility includes plan quality and risk, the prior beliefs about these characteristics will also affect the stakes and, therefore,

[19]We start with CARA utility, $-\exp(-\gamma(W - C_{ijt}))$, where cost, C_{ijt}, is normally distributed. In particular, let $C_{ijt} \sim N(p_{jt} + v_{ijt}, \widetilde{\sigma^2_{ijt}})$. Following Abaluck and Gruber (2009), this can be mapped into a conditional logit model.

incentives to conduct research. 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. Finally, we 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.[20] 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. This helps address concerns that individuals with specific diagnoses may have heterogeneous preferences for plan characteristics other than cost.

The idiosyncratic taste shock ϵ_{ijt} is assumed to be i.i.d. with variance normalized to $\pi^2/6$, as in standard logit models. We assume that the taste shock follows the conjugate to the EV1 distribution, the same distribution as the prior. This allows us to derive a novel formulation of the rational inattention model with unobserved heterogeneity, allowing for feasible estimation.[21] As in a standard model, the taste shock is assumed to be known by the decision maker, but not to the econometrician.[22] The magnitude of parameter α_i can be interpreted as the marginal utility per dollar when individuals are fully informed.

Let $\xi_{ijt} \equiv \alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}$ be the component of utility that is initially unknown to the individual but can be observed with costly information acquisition. We assume that individuals have prior mean ξ^0_{ijt}, which may differ across options. As before, we measure prior variance as

$$\sigma^2_{it} = Var_j \left[\xi_{ijt}\right] \tag{13}$$

which is common to all options in an individual's choice set.[23] The prior distribution for

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

[21]The key distributional assumption is described in greater detail in Appendix A-2. The logit and probit model, which assume EV1 and normally distributed errors respectively, often yield nearly identical estimates. Similarly, we argue that the model's assumption regarding the distribution of the idiosyncratic shock is relatively innocuous. Also see related Monte Carlo simulation in Appendix H.

[22]The model can also accommodate a taste shock that is initially unobserved unless individuals acquire costly information. To the extent that the taste shock reflects factors such as an individual's preference for a specific insurer company, we expect these characteristics to be easily observable by the individual.

[23]This is analogous to the search literature in which individuals are assumed to know the distribution of prices

each option are assumed to be independent. Let this multivariate distribution have CDF given by $G_{it}(\xi_{ijt})$. As in Section 2, this distribution is assumed to be the conjugate to the EV1 distribution.

This emits a closed-form-solution for initial choice probabilities, which in turn allows us to derive an expression for choice probabilities after information acquisition:

$$P_{ijt} = \frac{\exp\left[\frac{\alpha_i v_{ijt}+\beta_1 X^u_{jt}+\beta_2\widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt}+\ell_{it}\beta_3 X^k_{jt}+\ell_{it}\zeta_{b(j)d(it)}+\xi^0_{ijt}}{k_{it}\lambda_{it}(\ell_{it}-1)}\right]}{\sum_{k\in\mathcal{J}_{it}}\exp\left[\frac{\alpha_i v_{ikt}+\beta_1 X^u_{kt}+\beta_2\widetilde{\sigma^2_{ikt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{kt}+\ell_{it}\beta_3 X^k_{kt}+\ell_{it}\zeta_{b(k)}+\xi^0_{ikt}}{k_{it}\lambda_{it}(\ell_{it}-1)}\right]}. \tag{14}$$

where

$$\ell^2_{it} \equiv \frac{6\sigma^2_{it}}{\pi^2\lambda^2_{it}} + 1, \quad k^2_{it} \equiv \frac{\ell^2_{it} + \lambda^2_{it}(\ell_{it}-1)^2}{\lambda^2_{it}(\ell_{it}-1)^2} \tag{15}$$

The derivation for equation (14) is found in Appendix A-2.

Given choice probabilities, the choice problem is as if the individual maximizes utility given by

$$\tilde{u}_{ijt} = \frac{1}{k_{it}\lambda_{it}(\ell_{it}-1)}(\ell_{it}\alpha_i p_j + \xi^0_{ijt}) + \frac{1}{k_{it}\lambda_{it}}\alpha_i v_{ij} + \frac{1}{k_{it}\lambda_{it}}\beta_1 X^u_{jt} + \frac{\ell_{it}}{k_{it}\lambda_{it}(\ell_{it}-1)}(\beta_3 X^k_{jt} + \zeta_{b(j)d(it)}) + e_{ij} \tag{16}$$

Unlike the simple model in Section 2, the normalized idiosyncratic error, e_{ij}, now reflects the combined effect of the taste shock as well as heterogeneous beliefs. By construction, e_{ij} is distributed iid extreme value type 1 with scale parameter 1.

It is useful to consider the choice probabilities as the marginal cost of information goes to zero. This is given by

$$\lim_{\lambda_{it}\to 0} P_{ijt} = \frac{\exp\left[\alpha_i(v_{ijt}+p_{jt}) + \beta_1 X^u_{jt} + \beta_2\widetilde{\sigma^2_{ijt}} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}\right]}{\sum_{k\in\mathcal{J}_{it}}\exp\left[\alpha_i(v_{ikt}+p_{kt}) + \beta_1 X^u_{kt} + \beta_2\widetilde{\sigma^2_{ikt}} + \beta_3 X^k_{kt} + \zeta_{b(k)}\right]} \tag{17}$$

which are choice probabilities under full information. In other words, the model nests the standard logit model assuming rational expectations when the marginal cost of information is zero.

We now describe the specific assumptions we make regarding heterogeneity across indi-

in the market before they search.

viduals in the price coefficient, α_i, and the marginal cost of information, λ_{it}. We allow for observable heterogeneity in price sensitivity by assuming

$$\alpha_i = -\exp(\beta^{\alpha} Z_i) \tag{18}$$

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

$$\lambda_{it} = \exp(\beta^{\lambda 1} Z_i + \beta^{\lambda 2} W_{it}) \tag{19}$$

where W_{it} are time varying characteristics including the individual's health status and experience with Medicare Part D. 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 designs, making them equally complicated.

In the baseline specification, we assume that an individual has a common prior mean across options in her 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.[24]

We also consider specifications in which individuals start with additional information about plans, i.e. allow for a heterogeneous prior across choices. We consider a model in which ξ^0_{ijt} is determined by average out-of-pocket spending for the plan across all individuals in each year. In other words, individuals initially know the mean out-of-pocket cost for a plan but do not known their individual out-of-pocket cost until they conduct costly research. In addition, we consider specifications with alternative assumptions about the prior variance. We describe these specifications in greater detail in Appendix C.

5.1 Alternative Models of Insurance Demand 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

[24]Since choice probabilities only depend on differences in expected utility, the normalization of the prior is inconsequential.

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 are presented in Appendix D. Parameter estimates are in Table A-7.

5.2 Welfare

With costly information acquisition, individuals choose plans that maximize expected utility given beliefs, but do not necessarily maximize ex-post utility. Welfare must take into account the fact that there is a difference between the utility anticipated at the time of decision-making and the realized utility, 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

$$CS_{it}^{RI} = \frac{1}{|\alpha_i|} \log \sum_j e^{\tilde{v}_{ijt}} + \frac{1}{|\alpha_i|} \sum_j P_{ijt}[\nu_{ijt} - \tilde{v}_{ijt}] \tag{20}$$

where $\nu_{ijt} = \alpha_i v_{ijt} + \beta_1 X_{jt}^u + \beta_2 \widetilde{\sigma_{ijt}^2} + \alpha_i p_{jt} + \beta_3 X_{jt}^k$ is the true utility excluding the i.i.d. shock ϵ_{ijt} and $\tilde{v}_{ijt} = \frac{\alpha_i v_{ijt} + \beta_1 X_{jt}^u + \beta_2 \widetilde{\sigma_{ijt}^2}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt} + \xi_{ijt}^0 + \beta_3 \ell_{it} X_{jt}^k}{k_{it}\lambda_{it}(\ell_{it}-1)}$ is the expected utility given beliefs excluding the i.i.d. shock e'_{ijt}. The first term is the expected welfare calculated as if beliefs were correct. Note that $1/|\alpha_i|$ is the marginal utility of income. The second term adjusts for the fact that there may be a difference between beliefs and the true utility of each option. This term is the weighted average of the difference between anticipated consumer surplus and true consumer surplus where the weights are the probability of choosing each option as determined by equation (14).[27] Further detail is provided in Appendix A-3.

Following the assumptions of the rational inattention model in Section 2, the total cost of

[25]This general approach has been used by Bundorf et al. (2012), Handel (2013), Decarolis et al. (2020), Polyakova (2016), Ericson and Starc (2016), Tebaldi (2017), and others.

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

[27]Equation 20 follows from Allcott (2013) and Train (2015) who consider welfare in discrete choice models when beliefs, $\tilde{v}_{ijt}$, and actual utility differ and beliefs are observed. In contrast, $\tilde{v}_{ijt}$ is unobserved and endogenous in our setting.

information is determined by the mutual information

$$\hat{C}_{it} = \frac{\lambda_{it}}{|\alpha_i|} \mathbb{E}_{\epsilon} \left(H(G) - \mathbb{E}_s[H(F(\cdot|s))] \right) \tag{21}$$

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

$$\hat{C}_{it} = \frac{\lambda_{it}}{|\alpha_i|} \int_{\epsilon} \left(- \sum_{j \in \mathcal{J}_{it}} P^0_{ijt}(\epsilon) \log P^0_{ijt}(\epsilon) + \int_{\xi} \left(\sum_{j \in \mathcal{J}_{it}} P_{ijt}(\xi, \epsilon) \log P_{ijt}(\xi, \epsilon) \right) G(d\xi) \right) M(d\epsilon) \tag{22}$$

where $G(\xi)$ 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 $G(\xi)$ and $M(\epsilon)$ and averaging over the draws.

The welfare loss due to information frictions is then given by

$$\Delta CS_{it} = CS^{FullInfo}_{it} - CS^{RI}_{it} + \hat{C}_{it} \tag{23}$$

where $C^{FullInfo}_{it}$ is consumer surplus under full information.

5.3 Identification and Estimation

Our primary identification concern is separately identifying the marginal cost of information from preference heterogeneity. Specifically, we wish to separately identify preference parameters, including the coefficients on the price and other product characteristics, separately from the cost of acquiring information. In many applications, information frictions in which an individual receives an imprecise signal of product characteristics imply an error term that is observationally equivalent to a taste shock.

For identification, we leverage the fact that individuals observe premiums but do not initially observe out-of-pocket costs. In addition, we rely on the assumption that individuals are equally sensitive toward premiums and expected out-of-pocket costs under full information controlling for plan riskiness and other plan characteristics (see equation (17)). This assumption implies that under full information, a change in premiums and an equivalent change in expected out-of-pocket cost have the same effect on choice probabilities. If observed choices are equally sensitive to premiums and out-of-pocket costs, then we conclude that there are no information frictions and heterogeneous preferences are largely a result of the taste shock or preferences over non-price characteristics.

By contrast, our model implies that the gap between the premium and out-of-pocket cost coefficients depends both on the unit cost of information and the stakes as seen in equation (14). Furthermore, our empirical application allows the marginal cost of information to vary with individual characteristics such as income, age, education, part D experience, and chronic conditions, which introduces correlation between λ_{it} and the stakes. Therefore, the observed weights that individuals place on premiums and out-of-pocket costs and how those weights change as the stakes increase help identify both the average level of λ_{it} in the population and heterogeneity in λ_{it}.

Appendix G presents a more formal discussion of identification. We show how coefficients in our model map to coefficients in a standard discrete-choice model. In particular, if one can identify the coefficient on premium and a separate coefficient on the out-of-pocket cost in a standard discrete choice model, the ratio of those coefficients pins down the marginal cost of information for a given level of the stakes. In this way, the identification of our model is related to previous work estimating preference parameters for premiums and plan benefits in insurance markets.[28] 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 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. Finally, there is significant variation in which insurers operate in each of the 34 Medicare Part D markets, generating variation in the choice set across markets.

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

[28] See, for instance, discussion in Ho et al. (2017) and Polyakova (2016).

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 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. A related concern is that unhealthy individuals, such as those with chronic conditions, have different preferences for non-pecuniary characteristics of plans. In a separate specification, we explore a specification that include interactions between insurer fixed effects and an indicator for chronic conditions to allow for preferences for insurers to be correlated with health status. Finally, there is concern that individuals with previous experience with an insurer may have brand loyalty even if their previous plan in no longer available. For this reason, we also include an indicator variable for the previously chosen insurer.

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.

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.[29] The log-likelihood function is reported in Appendix A-2.

5.4 Empirical Model Estimates and Fit

The parameter estimates from the baseline demand model are presented in Table 4. Average price sensitivity parameter for individuals, 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 sensi-

[29]One challenge is that the log-likelihood is prone to numerical rounding errors when λ_{it} is large relative to σ_{it}^2 or vice versa, causing the log-likelihood to be non-finite. We address this by ensuring that estimation is robust to using increased numerical precision by employing Multiprecision Computing Toolbox for Matlab.

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: Premium and out-of-pocket cost are in hundreds of dollars. Continuous individual characteristics (income, education, age, and age squared) are demeaned. Insurer fixed effects are included. Standard errors in parentheses.

tive, 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.

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

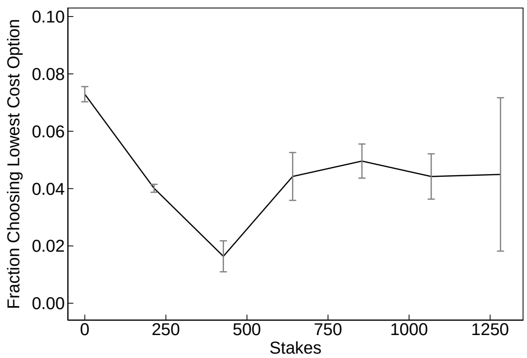
Individuals in more educated zip codes have lower marginal cost of information, consistent with the idea that it is easier for more educated individuals to research plans. However, this parameter is not statistically significant. Older individuals may have more difficulty researching plans. The coefficient 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, there is large variation in the marginal cost of information across individuals. Along with the variation in the stakes, this implies large differences in the total realized cost of information acquisition. The distribution of the marginal cost of information and total cost of information is shown in Figure A-8 in Appendix F. The results imply that a quarter of individuals spend less than $40 researching plans.

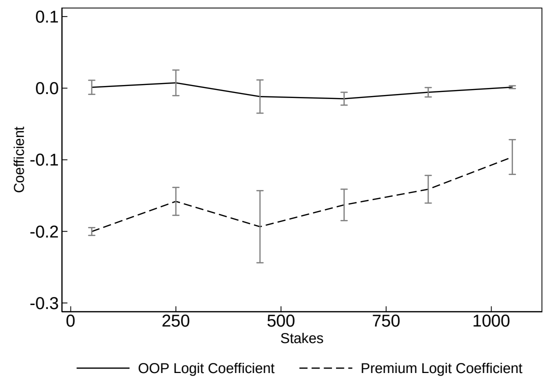
We evaluate model fit in a few ways. First, we use the baseline specification to simulate the probability of choosing the lowest cost plan and the weight that individuals appear to place on premium and out-of-pocket costs as a function of the stakes. The results can be compared to the descriptive analysis presented in Section 4. Panel a and b in Figure 5 show that the model can recover the patterns documented in Figure 3 and Figure 4 in the previous section. We also use the estimates from each of the three alternative models that do not allow for endogenous information. As seen in Panel c and d, these alternative models have difficulty rationalizing why choices change when the stakes change.

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 model accurately predicts the total cost, the out-of-pocket cost and premium both differ by over $50.

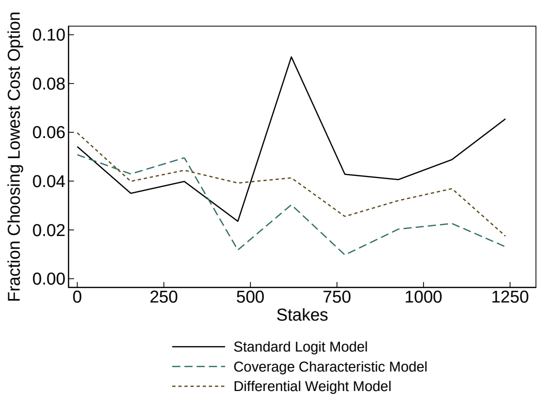
Figure 5
Fit of Endogenous Information Model and Alternative Models



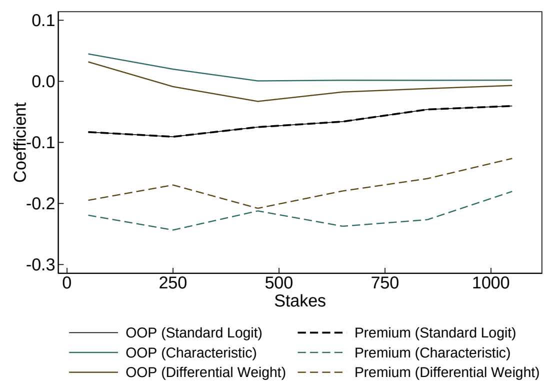
a. Fraction Choosing Lowest Cost Plan
Endogenous Information Model



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 3 and Figure 4. For further description see Section 2.

6 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 help highlight the role of endogenous information and implications for proposed policies aimed at simplifying choice.

6.1 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 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 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.[31]

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. potential drugs to take) 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 the expected cost of understanding the forms and adjusting to a new plan is high. 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.[32] As seen in Table 5, the implied welfare gains of full information are even

[31] 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.

[32] 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: Counterfactual simulations for endogenous information model use parameter estimates from specification 1 in Table 4. 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.

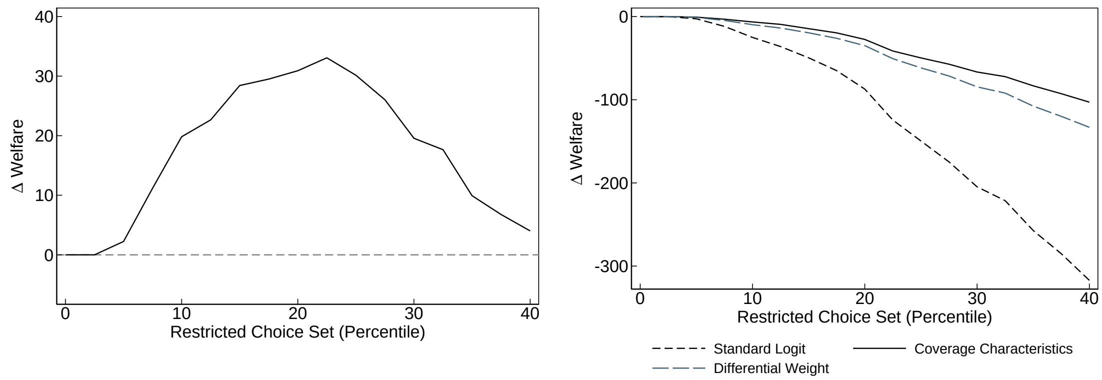
larger when the taste shock is excluded.

In the baseline case, the estimated elasticity of demand with respect to premiums is -1.5, however the elasticity with respect to out-of-pocket costs is only -0.4.[33] 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). We derive the expressions for elasticity in Appendix A-2. The large difference in elasticities reflects the importance of information frictions. Under full information, the elasticity of demand is -1.6, the same for both premiums and 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.

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

Figure 6
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 commonly used models of plan demand.

6.2 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.[34] 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.[35]

The change in welfare from restricting the choice set accounting for endogenous information is depicted in Figure 6 Panel a. As seen in the figure, there is a trade-off. 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

[34]To a certain extent, insurance regulators already do this through allocation policies that set minimum standards for plans. Handel and Schwartzstein (2018) list various allocation policies in health insurance and other markets.

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

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: Counterfactual simulations for endogenous information model use parameter estimates from specification 1 in Table 4. Restricted choice counterfactual removes plans with average utility below cutoff. Out-of-pocket cap counterfactual imposes limit on out-of-pocket cost of all plans and increases premiums such that the policy is revenue neutral.

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 indicating that individuals would prefer fewer options in the Part D market.

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. 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 6 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-10.

The welfare gains of restricting choices are even larger if a social planner can provide a personalized list of plans to each individuals.[36] This can be seen in Figure A-4 Panel a,

[36] I.e. if the social planner restricts the choice set based on representative utility given by $\alpha_i v_{ijt} + \beta_1 X^u_{jt} +$

which implies that welfare is maximized when more than 75% of plans are removed. Again, Figure A-4 Panel b shows that alternative models imply that that there are large welfare losses from this policy.

Theoretically, it may be possible to increase welfare by randomly removing options in some cases. In the context of Medicare Part D, we do not find that this is the case. These results, presented in Figure A-7, show that welfare generally decreases as options are randomly removed. However, as shown in panel b of Figure A-7, the decrease in welfare is much smaller than implied by the standard logit model.

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 with heterogeneous preferences to have choice. 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.

6.3 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.[37] Imposing an out-of-pocket cap effectively, 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.[38]

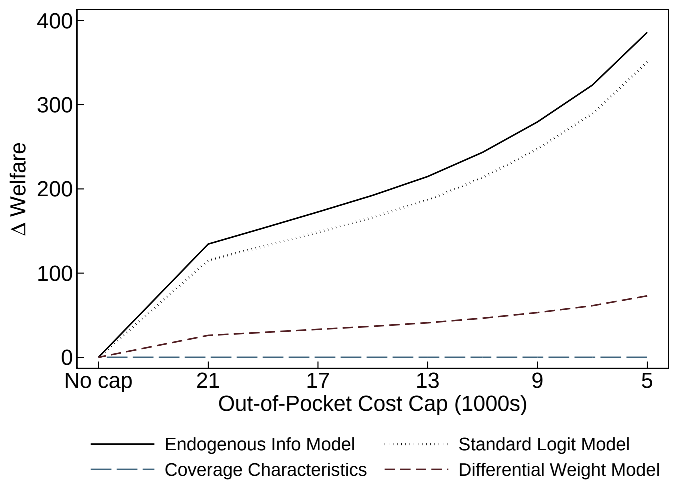
Figure 7 shows the change in welfare for different levels of an out-of-pocket cost cap. The endogenous information model implies higher welfare gains from capping out-of-pocket costs than other models, especially the differential weight model. The coverage characteristics model provides no insight into a cap given that demand is not directly a function of expected out-of-pocket costs. 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

$\beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}$.

[37]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.

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

Figure 7
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.

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. Under the endogenous information model, this policy is still welfare increasing (see Figure A-6). In contrast, alternative models imply a decrease in welfare. Results for these alternative models are presented in Table A-10.

Overall, these results highlight that a cap on out-of-pocket costs can 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.

6.4 Discussion and Robustness

Estimating a model of insurance demand with costly information requires a number of simplifying assumptions. First, identifying the cost of information requires assumptions on preferences for plan characteristics other than cost. We include 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 find that they are similar, implying that unobserved preferences for insurers are not driving the estimate after controlling for observable plan characteristics.[39] 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.[40]

Another key assumption of the model is the form of individuals' prior since it determines information acquisition. We examine estimates under alternative assumptions on the prior. Specification 1 in Table A-9 assumes that individuals initially know the average cost of the plan across all individuals but not their own out-of-pocket cost. Allowing for a different prior mean for different plans in this way yields similar parameter estimates. The mean price sensitivity is very similar to Specification 1, although the implied marginal cost of information is higher. We also consider alternative specifications in which we assume individuals initially know the average variance in out-of-pocket costs and other unknown plan characteristics for similar individuals, but not the variance of out-of-pocket costs across their own choice set. This alternative definition of the stakes also implies quite similar parameter estimates. This can be seen in Table A-9 Specification 2 and 3. Finally, it is possible that individuals may 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.

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

[39] See first column of Table A-8. Table A-8 also shows a specification including the individuals with very high stakes (these outliers are removed from the main sample). These results are also quite similar.

[40] 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.

ingredients, strength, dosage, and route of administration), but assume no other forms of moral hazard.[41] 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 (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.[42]

7 Conclusion

We develop a micro-founded yet tractable demand model 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 of demand directly based on the rational inattention framework. Estimates imply that the welfare effects of information frictions are substantial, especially when information acquisition costs are included. 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. We find that accounting for endogenous

[41] 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), Levy and Weir (2008), Lichtenberg and Sun (2007), and Ketcham and Simon (2008)). 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 (2009). Based on a stylized model they show that ignoring the fact that individuals adjust drug usage when choosing a different plan leads to an underestimation of the welfare effect of choice frictions. They also argue that the magnitude of this bias is small given that the elasticity of demand for prescription drugs is estimated to be small in the existing literature.

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

information is important when considering these policies, both because of the change in information acquisition costs and the resulting effect on insurance choice. The results are also broadly consistent with survey evidence that individuals do not fully understand insurance plans and would prefer policies that simplify plan choice.

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. There are a variety of settings that have this feature. For example, consumers face sticker prices versus add-on charges such as delivery fees and taxes, interest rates versus hidden fees in the mortgage market, and car prices and loan interest rates in the auto market. 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. Similar approaches could be used to provide insight into the welfare effects of targeted defaults.

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. Using the model, we find demand elasticity with respect to premiums is -1.5 while demand elasticity with respect to out-of-pocket costs is only -0.4, implying that insurers have more incentive to compete on premiums. 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 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.

— **and** — , "When Less is More: Improving Choices in Health Insurance Markets," Working Paper 2020.

Abaluck, Jason T and Jonathan Gruber, "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," Technical Report, National Bureau of Economic Research 2009.

Allcott, Hunt, "The welfare effects of misperceived product costs: Data and calibrations from the automobile market," *American Economic Journal: Economic Policy*, 2013, *5* (3), 30–66.

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.

Berry, Steven T, "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

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.

Caplin, Andrew, Mark Dean, and John Leahy, "Rational inattention and inference from market share data," Working Paper 2016.

— , — , and — , "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *The Review of Economic Studies*, 2019, *86* (3), 1061–1094.

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.

Cebul, Randall D, James B Rebitzer, Lowell J Taylor, and Mark E Votruba, "Unhealthy insurance markets: Search frictions and the cost and quality of health insurance," *American Economic Review*, 2011, *101* (5), 1842–71.

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.

Dean, Mark and Naathaniel Neligh, "Experimental Tests of Rational Inattention," Working Paper 2019.

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.

Diamond, Peter A., "A model of price adjustment," *Journal of Economic Theory*, June 1971, *3* (2), 156–168.

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.

Fosgerau, Mogens, Emerson Melo, André de Palma, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," *International Economic Review*, 2020.

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.

_ and Joshua Schwartzstein, "Frictions or mental gaps: what's behind the information we (don't) use and when do we care?," *Journal of Economic Perspectives*, 2018, *32* (1), 155–78.

_ , Jonathan T Kolstad, and Johannes Spinnewijn, "Information frictions and adverse selection: Policy interventions in health insurance markets," *Review of Economics and Statistics*, 2019, *101* (2), 326–340.

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.

Hong, Han and Matthew Shum, "Using price distributions to estimate search costs," *RAND Journal of Economics*, 06 2006, *37* (2), 257–275.

Honka, Elisabeth, "Quantifying search and switching costs in the US auto insurance industry," *RAND Journal of Economics*, December 2014, *45* (4), 847–884.

_ and Pradeep Chintagunta, "Simultaneous or sequential? search strategies in the us auto insurance industry," *Marketing Science*, 2017, *36* (1), 21–42.

Hortaçsu, Ali and Chad Syverson, "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *The Quarterly Journal of Economics*, 2004, *119* (2), 403–456.

Joo, Joonhwi, "Rational Inattention as an Empirical Framework: Application to the Welfare Effects of New-Product Introduction," Working Paper 2020.

Keane, Michael P, Jonathan D Ketcham, Nicolai V Kuminoff, and Timothy Neal, "Evaluating Consumers' Choices of Medicare Part D Plans: A Study in Behavioral Welfare Economics," Technical Report, National Bureau of Economic Research 2019.

Ketcham, Jonathan D and Kosali Simon, "Medicare Part D's effects on elderly drug costs and utilization," Technical Report, National Bureau of Economic Research 2008.

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

_ , _ , Eugenio J Miravete, and M Christopher Roebuck, "Sinking, swimming, or learning to swim in Medicare Part D," *American Economic Review*, 2012, *102* (6), 2639–73.

_ , Nicolai V Kuminoff, and Christopher A Powers, "Estimating the Heterogeneous Welfare Effects of Choice Architecture," *International Economic Review*, 2019, *60* (3), 1171–1208.

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.

Levy, Helen and David Weir, "The impact of Medicare Part D on drug utilization and out-of-pocket spending: Evidence from the Health and Retirement Study," *Unpublished Manuscript*, 2008.

Lichtenberg, Frank R and Shawn X Sun, "The impact of Medicare Part D on prescription drug use by the elderly," *Health Affairs*, 2007, *26* (6), 1735–1744.

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, "Simple Market Equilibria with Rationally Inattentive Consumers," *American Economic Review*, 2012, *102* (3), 24–29.

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

McGarry, Brian E, Nicole Maestas, and David C Grabowski, "Simplifying The Medicare Plan Finder Tool Could Help Older Adults Choose Lower-Cost Part D Plans," *Health Affairs*, 2018, *37* (8), 1290–1297.

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.

Santos, Babur De Los, Ali Hortaçsu, and Matthijs R. Wildenbeest, "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, October 2012, *102* (6), 2955–80.

Seiler, Stephan, "The impact of search costs on consumer behavior: A dynamic approach," *Quantitative Marketing and Economics*, 2013, *11* (2), 155–203.

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

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

Stigler, George J., "The Economics of Information," *Journal of Political Economy*, 1961, *69*, 213.

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

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 Derivation

A-1 Basic Model without Taste Shock

Before individuals obtain information, initial choice probabilities, $P_1^0, .., P_N^0$, are determined by integrating over the prior given cost of information λ:

$$\max_{P_1^0,..,P_N^0} \int_{\mathbf{v}} \lambda \log \Sigma_j P_j^0 e^{(-p_j - v_j)/\lambda} G(d\mathbf{v}) \text{ s.t. } \sum_j P_j^0 = 1, P_j^0 \geq 0 \ \forall j. \quad \text{(A-1)}$$

For simplicity, we suppress subscripts for individual i. We start by deriving a closed-form expression for $P_1^0, .., P_N^0$ under assumptions about the distribution of the prior.

First, note that $\log \sum_j e^{v_j/k} = \mathbb{E}_e \left[\max_j(v_j + ke_j)\right] + C$ where $e_j \overset{iid}{\sim} EV1$ and C is a constant (Small and Rosen 1981). Applying this we have

$$\begin{aligned}\int_{\mathbf{v}} \log \Sigma_j e^{(-p_j - v_j)/\lambda + \log(P_j^0)} G(d\mathbf{v}) &= \mathbb{E}_{v,e}\left[\max_j((-p_j - v_j)/\lambda + \log(P_j^0) + e_j)\right] + C \quad \text{(A-2)} \\ &= \mathbb{E}_{v,e}\left[\max_j(-p_j/\lambda + \log(P_j^0) - v_j/\lambda + e_j)\right] + C \\ &= \mathbb{E}_{e'}\left[\max_j(-p_j/\lambda + \log(P_j^0) + \ell e_j')\right] + C\end{aligned}$$

where $\ell e_j' \equiv -v_i/\lambda + e_j$ is the joint error and $\ell^2 \equiv (\frac{6\sigma^2}{\pi^2\lambda^2} + 1)$ so that e_j' is normalized to have variance $\pi^2/6$. We assume e_j' is distributed EV1. This implies that the distribution of the prior is the conjugate to the EV1 distribution. Details about this distribution can be found in Cardell (1997).

Therefore,

$$\mathbb{E}_{e'}\left[\max_j((-p_j)/(\ell\lambda) + \log(P_j^0)/\ell + e_j')\right] + C = \log \Sigma_j e^{(-p_j)/(\lambda l) + \log(P_j^0)/\ell} + C'.$$

Now the maximization problem in equation (A-1) becomes

$$\max_{P_1^0,\ldots,P_N^0} \Sigma_j e^{-p_j/\ell\lambda+\log(P_j^0)/\ell} \text{ s.t.} \quad \sum_j P_j^0 = 1,\, P_j^0 \geq 0\ \forall j \tag{A-3}$$

where we have ignored the constant since it is the same for every option.

Then, the first order condition with respect to P_i^0 is given by

$$\frac{\partial}{\partial P_i^0}\left(\sum_j (P_j^0)^{\frac{1}{\ell}} e^{\frac{-p_j}{\lambda\ell}} + \eta\left(1-\sum_j P_j^0\right)\right) = 0$$

where η is the Lagrange multiplier. Solving for this first order condition, we obtain an expression of P_i^0 as a function of η:

$$P_i^0 = \left(\ell\eta e^{\frac{p_i}{\lambda\ell}}\right)^{\frac{\ell}{1-\ell}}. \tag{A-4}$$

From the constraint, we can obtain an expression for η.

$$\sum_j P_j^0 = \sum_j \left(\ell\eta e^{\frac{p_j}{\lambda\ell}}\right)^{\frac{\ell}{1-\ell}} = 1$$

$$\eta = 1/\sum_j \left(\ell e^{\frac{p_j}{\lambda(1-\ell)}}\right)^{\frac{1-\ell}{\ell}} \tag{A-5}$$

Plugging equation (A-5) into A-4, we can obtain a closed-form expression for P_j^0.

$$P_j^0 = \frac{e^{-p_j/(\lambda\ell-\lambda)}}{\sum_k e^{-p_k/(\lambda\ell-\lambda)}}$$

With an expression for P_j^0 in hand, we can now derive an expression for choice probabilities after information acquisition. From equation (A-3) we can see that it is as if the agent maximizes the following utility

$$u_j = (-p_j - v_j)/\lambda + \log(P_j^0) + e_j$$

where e_j is an iid EV1 error causes by incorrect beliefs. Substituting the expression for P_j^0, the utility is

$$u_j = (-p_j - v_j)/\lambda + q_j/(\lambda\ell - \lambda) + \log(\sum_k e^{-p_k/(\lambda\ell-\lambda)}) + e_j$$

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

$$P_j = \frac{e^{(-p_j\ell/(\ell-1)-v_j)/\lambda}}{\sum_k e^{(-p_k\ell/(\ell-1)-v_k)/\lambda}}. \tag{A-6}$$

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

$$e^p = \frac{\ell}{\lambda(\ell-1)}(1-P_j)(p_j+v_j), \tag{A-7}$$

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

$$e^v = \frac{1}{\lambda}(1-P_j)(p_j+v_j). \tag{A-8}$$

A-2 Empirical Model with Taste Shock

For the case with taste shocks, expected choice probabilities before information acquisition, P^0_{ijt}, are determined by integrating over individuals' prior beliefs given the marginal cost of information λ_{it}. In particular, they are determined as the solutions to the following problem:

$$\max_{P^0_{i1t},..,P^0_{iNt}} \int_{\boldsymbol{\xi}} \lambda_{it} \log \Sigma_{j=1}^N P^0_{ij} \exp\left[(\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it}\right] G_{it}(d\boldsymbol{\xi})$$

$$\text{s.t.} \quad \sum_{j=1}^N P^0_{ijt} = 1,\ P^0_{ijt} \geq 0\ \forall j \tag{A-9}$$

Now we apply a similar approach as the previous section. Note that $\log \sum_j e^{v_j/k} = \mathbb{E}_e\left[\max_j(v_j + ke_j)\right] + C$ where $e_j \overset{iid}{\sim} EV1$ and C is a constant (Small and Rosen 1981). Applying this we have

$$\int_{\boldsymbol{\xi}} \log \Sigma_j e^{\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it} + \log(P^0_{ijt})} G_{it}(d\boldsymbol{\xi})$$

$$= \mathbb{E}_{\xi,e}\left[\max_j((\alpha_i p_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it} + \log(P^0_{ijt}) + e_{ijt})\right] + C$$

$$
\begin{aligned}
&= \mathbb{E}_{\xi,e}\left[\max_j((\alpha_i p_{ijt} + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \log(P_{ijt}^0) + \xi_{ijt}/\lambda_{it} + e_{ijt})\right] + C \\
&= \mathbb{E}_{\xi',e}\left[\max_j((\alpha_i p_{ijt} + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \log(P_{ijt}^0) + \xi_{ijt}^0/\lambda_{it} + \xi'_{ijt}/\lambda_{it} + e_{ijt})\right] + C
\end{aligned} \tag{A-10}
$$

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

Note that the joint error is $\xi'_{ijt}/\lambda_{it} + e_j$ which has the following variance.

$$
Var[\xi'_{ijt}/\lambda_{it} + e_j] = \frac{\sigma_{it}^2}{\lambda_{it}^2} + \frac{\pi^2}{6}.
$$

We define joint error as $\ell_{it} e'_{ijt} \equiv \xi'_{ijt}/\lambda_{it} + e_j$ where $Var[e'_{ijt}] = \frac{\pi^2}{6}$. Therefore,

$$
\begin{aligned}
Var[\ell_{it} e'_{ijt}] &= \frac{\sigma_{it}^2}{\lambda_{it}^2} + \frac{\pi^2}{6} \\
\ell_{it}^2 \frac{\pi^2}{6} &= \frac{\sigma_{it}^2}{\lambda_{it}^2} + \frac{\pi^2}{6} \\
\ell_{it}^2 &= \frac{6\sigma_{it}^2}{\pi^2 \lambda_{it}^2} + 1
\end{aligned}
$$

Then, equation (A-10) can be rewritten as

$$
\begin{aligned}
&\int_{\boldsymbol{\xi}} \log \Sigma_j e^{\alpha_i p_{jt} + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it} + \log(P_{ijt}^0)} G_{it}(d\boldsymbol{\xi}) \\
&\qquad = \mathbb{E}_{e'}\left[\max_j((\alpha_i p_{ijt} + \xi_{ijt}^0 + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \log(P_{ijt}^0) + \ell_{it} e'_{ijt})\right] + C
\end{aligned}
$$

As in the case without a taste shock, we assume that e'_{ijt} is distributed EV1, which implies that the distribution of ξ'_{ijt} follows the same distribution as the prior, the conjugate to the scaled EV1 distribution.

Note that

$$
\begin{aligned}
&\mathbb{E}_{e'}\left[\max_j((\alpha_i p_{ijt} + \xi_{ijt}^0 + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \log(P_{ijt}^0) + \ell_{it} e'_{ijt})\right] + C \\
&\qquad = \log \Sigma_{j=1}^N \exp[(\alpha_i p_{ijt} + \xi_{ijt}^0 + \beta_3 X_{jt}^k + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\ell_{it}\lambda_{it} + \log(P_{ijt}^0)/\ell_{it}]
\end{aligned}
$$

Now the maximization problem in equation (A-9) can be rewritten as

$$\max_{P^0_{i1t},\ldots,P^0_{iNt}} \Sigma_{j=1}^{N} \exp[(\alpha_i p_{ijt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\ell_{it}\lambda_{it} + \log(P^0_{ijt})/\ell_{it}]$$

$$\text{s.t.} \quad \sum_{j=1}^{N} P^0_{ijt} = 1,\ P^0_{ijt} \geq 0\ \forall j$$

From solving this maximization problem, we can derive a closed-form expression for P^0_{ijt} as

$$P^0_{ijt} = \frac{\exp\left[(\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt})/(\lambda_{it}\ell_{it} - \lambda_{it})\right]}{\sum_{k=1}^{N} \exp\left[(\alpha_i(p_{kt} + v^0_{ikt}) + \beta_3 X^k_{kt} + \zeta_{b(k)} + \epsilon_{ikt})/(\lambda_{it}\ell_{it} - \lambda_{it})\right]}.$$

With an expression for P^0_{ijt} 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

$$P_{ijt} = \frac{\exp\left[(\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \log(P^0_{ijt})\right]}{\sum_{k=1}^{N} \exp\left[(\alpha_i v_{ikt} + \beta_1 X^u_{kt} + \beta_2 \widetilde{\sigma^2_{ikt}} + \alpha_i p_{kt} + \beta_3 X^k_{kt} + \zeta_{b(k)} + \epsilon_{ikt})/\lambda_{it} + \log(P^0_{ikt})\right]}$$

Therefore, the problem is now as if individuals maximize utility given by

$$\tilde{u}_{ijt} = (\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \log(P^0_{ijt}) + e_{ijt}$$

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

$$\begin{aligned}\tilde{u}_{ijt} = (\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt})/\lambda_{it} + \\ (\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt})/(\lambda_{it}\ell_{it} - \lambda_{it}) + e_{ijt} \quad \text{(A-11)}\end{aligned}$$

where $\log\left[\sum_{k=1}^{N} \exp\left[(\alpha_i(p_{kt} + v^0_{ikt}) + \beta_3 X^k_{kt} + \zeta_{b(j)d(it)} + \epsilon_{ikt})/(\lambda_{it}\ell_{it} - \lambda_{it})\right]\right]$ is a constant that is the same for every option, and therefore does not affect choice probabilities. We can simplify equation (A-11) to

$$\begin{aligned}\tilde{u}_{ijt} &= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}}{\lambda_{it}} + \frac{\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}}{\lambda_{it}\ell_{it} - \lambda_{it}} + \frac{\epsilon_{ijt}}{\lambda_{it}\ell_{it} - \lambda_{it}} + \frac{\epsilon_{ijt}}{\lambda_{it}} + e_{ijt} \\ &= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}}{\lambda_{it}} + \frac{\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}}{\lambda_{it}(\ell_{it} - 1)} + \frac{\ell_{it}}{\lambda_{it}(\ell_{it} - 1)}\epsilon_{ijt} + e_{ijt}\end{aligned}$$

$$= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{\lambda_{it}} + \frac{(\ell_{it}-1)\left(\alpha_i p_{jt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}\right)}{\lambda_{it}(\ell_{it}-1)} + \frac{\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \zeta_{b(j)d(it)}}{\lambda_{it}(\ell_{it}-1)} + \frac{\ell_{it}}{\lambda_{it}(\ell_{it}-1)}\epsilon_{ijt} + e_{ij}$$

$$= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt} + \xi^0_{ijt} + \ell_{it}(\beta_3 X^k_{jt} + \zeta_{b(j)d(it)})}{\lambda_{it}(\ell_{it}-1)} + \frac{\ell_{it}}{\lambda_{it}(\ell_{it}-1)}\epsilon_{ijt} + e_{ijt} \tag{A-12}$$

We define the joint error as $k_{it}e^{'}_{ijt} \equiv \frac{\ell_{it}}{\lambda_{it}(\ell_{it}-1)}\epsilon_{ijt} + e_{ijt}$ where $Var[e^{'}_{ijt}] = \frac{\pi^2}{6}$. Again, we assume that the distribution of the taste shock is such that the joint error is distributed extreme value type 1. Therefore,

$$\begin{aligned} Var[k_{it}e^{'}_{ijt}] &= \frac{\ell^2_{it}}{\lambda^2_{it}(\ell_{it}-1)^2}\frac{\pi^2}{6} + \frac{\pi^2}{6} \\ k^2_{it}\frac{\pi^2}{6} &= \frac{\pi^2}{6}\left[\frac{\ell^2_{it}}{\lambda^2_{it}(\ell_{it}-1)^2} + 1\right] \\ k^2_{it} &= \frac{\ell^2_{it}}{\lambda^2_{it}(\ell_{it}-1)^2} + 1 \end{aligned}$$

The utility in equation (A-12) can be then rewritten as

$$\frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt} + \xi^0_{ijt} + \ell_{it}(\beta_3 X^k_{jt} + \zeta_{b(j)d(it)})}{k_{it}\lambda_{it}(\ell_{it}-1)} + e'_{ijt}.$$

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

$$P_{ijt} = \frac{\exp\left[\frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt} + \xi^0_{ijt} + \ell_{it}(\beta_3 X^k_{jt} + \zeta_{b(j)d(it)})}{k_{it}\lambda_{it}(\ell_{it}-1)}\right]}{\sum_{k=1}^N \exp\left[\frac{\alpha_i v_{ikt} + \beta_1 X^u_{kt} + \beta_2 \widetilde{\sigma^2_{ikt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{kt} + \xi^0_{ijt} + \ell_{it}(\beta_3 X^k_{kt} + \zeta_{b(k)})}{k_{it}\lambda_{it}(\ell_{it}-1)}\right]}.$$

The elasticity of demand with respect to premiums is then given by

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

while the elasticity of demand with respect to expected out-of-pocket cost is given by

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

$$= \alpha_i \frac{1}{k_{it}\lambda_i}(1 - P_{ij})(p_j + v_{ij}) \tag{A-14}$$

The above elasticities can be interpreted as the percent change in demand due to a one percent change in cost due to premiums and out-of-pocket costs respectively.

The log-likelihood function is given by

$$\mathcal{L}(\alpha_i, \lambda_{it}, \beta) = \sum_i \sum_t \left(\sum_{j \in \mathcal{J}_{it}} I(y_{it} = j) V_{ijt}(\alpha_i, \lambda_{it}, \beta) - \log \left(\sum_{j \in \mathcal{J}_{it}} \exp V_{ijt}(\alpha_i, \lambda_{it}, \beta) \right) \right) \tag{A-15}$$

where $V_{ijt}(\alpha_i, \lambda_{it}, \beta_1, \beta_3) = \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt} + \xi^0_{ijt} + \ell_{it}(\beta_3 X^k_{jt} + \zeta_{b(j)d(it)})}{k_{it}\lambda_{it}(\ell_{it} - 1)}$. Note that ℓ is also a function of model parameters since prior variance is a function of utility parameters.

A-3 Derivation of Welfare

We denote the utility individual i expects from alternative j given beliefs after information acquisition as $\tilde{u}_{ijt}$, which we call "belief utility". The difference between the true utility and the belief utility is denoted d_{ijt}. Then, the true 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

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

where $\nu_{ijt} = \alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt}$ is the true utility excluding the i.i.d. shock ϵ_{ijt} and $\tilde{\nu}_{ijt} = \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt} + \xi^0_{ijt} + \beta_3 \ell_{it} X^k_{jt}}{k_{it}\lambda_{it}(\ell_{it} - 1)}$ is the belief utility excluding the i.i.d. shock e'_{ijt}.

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 on Alternative Assumptions on the Prior

A key assumption is the nature of individual's prior. While the main specification assumes that individuals have a homogeneous prior across options in their choice set, we also consider an alternative assumption in which individuals have a different prior mean for each plan. We assume that the prior mean is determined by the average plan characteristics over the population for a given year. For example, individuals initially know a plan has high out-of-pocket cost for the average enrollee but do not know their own expected out-of-pocket cost until they conduct research.

We also consider alternative measures of the variance of individual's prior σ^2. In the benchmark case, analogously to the search literature, we assume that individuals know the variance of out-of-pocket costs in their choice set which determines σ^2. As an alternative assumption, we assume that individuals know the average variance of out-of-pocket costs across the choice sets of similar individuals but not the variance of their own choice set. Similar individuals are defined by dividing the sample into spending deciles and calculating the average within each decile. The results for this alternative specification can be found in Table A-9.

D 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.[43] 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. In particular, utility takes the form

$$u_{ijt} = \alpha_i \underbrace{(v_{ijt} + p_{jt})}_{\text{Total Cost}} + \beta_1 \widetilde{\sigma^2_{ijt}} + \beta_2 X_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt}. \tag{A-16}$$

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), 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

$$u_{ijt} = \alpha_i p_{jt} + \beta_1 C_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \beta_3 X_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt} \tag{A-17}$$

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

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

model.

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

$$u_{ijt} = \alpha_i p_{jt} + \beta_1 v_{ijt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \beta_3 X_{jt} + \zeta_{b(j)d(it)} + \epsilon_{ijt}. \tag{A-18}$$

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.

We estimate the models via MLE and present the parameter estimates in Table A-7.

E Appendix Tables

Table A-1
Correlation of Stakes with Individual Characteristics and Choice Set Characteristics

	All		Active Choice	
	Corr	p-value	Corr	p-value
Total spending	0.827	0.000	0.799	0.000
Age	0.034	0.000	0.025	0.000
Zip income	0.048	0.000	0.045	0.000
Zip education (%BA)	0.046	0.000	0.043	0.000
Alzheimers	0.133	0.000	0.119	0.000
Lung disease	0.129	0.000	0.128	0.000
Kidney disease	0.132	0.000	0.127	0.000
Heart failure	0.133	0.000	0.125	0.000
Depression	0.119	0.000	0.112	0.000
Diabetes	0.156	0.000	0.154	0.000
Other chronic condition	0.181	0.000	0.170	0.000
Has any chronic condition	0.204	0.000	0.195	0.000

Notes: Shows correlation coefficient between relevant variable and stakes.

Table A-2
Summary of Demographics for Active Choice Makers and All Enrollees

	All		Active Choice		New Enrollees	
	Mean	SD	Mean	SD	Mean	SD
Age	76.2	7.2	76.2	7.4	67.7	4.7
Female	0.605	0.489	0.602	0.489	0.550	0.498
Zip income (1,000s)	76.4	34.5	77.3	35.1	74.2	33.0
Zip education (pct BA)	29.5	17.0	29.9	17.1	28.0	16.5
Rural	0.079	0.269	0.074	0.262	0.083	0.276
Years enrolled in Part D	5.25	2.15	5.53	2.31	1.00	0.00
Alzheimers	0.084	0.278	0.086	0.281	0.035	0.184
Lung disease	0.100	0.300	0.101	0.302	0.102	0.303
Kidney disease	0.149	0.356	0.157	0.364	0.108	0.310
Heart failure	0.131	0.338	0.132	0.339	0.098	0.297
Depression	0.112	0.316	0.118	0.322	0.132	0.338
Diabetes	0.263	0.440	0.268	0.443	0.284	0.451
Other chronic condition	0.293	0.455	0.303	0.460	0.279	0.449
Observations	370,580		214,522		10,052	

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)	-2.264^{***}	-2.215^{***}	-0.163^{*}	-0.205^{**}	-0.227^{**}	-2.127^{***}
	(0.060)	(0.062)	(0.087)	(0.091)	(0.091)	(0.062)
Stakes Squared	0.197^{***}	0.190^{***}	0.032^{***}	0.030^{***}	0.032^{***}	0.183^{***}
	(0.005)	(0.006)	(0.007)	(0.008)	(0.008)	(0.006)
Plan Characteristic Controls	No	Yes	No	Yes	Yes	Yes
Individual FEs	No	No	Yes	Yes	Yes	No
Year FEs	No	No	No	No	Yes	Yes
Market FEs	No	No	No	No	No	Yes
Implied minimum	575.3	582.7	1.0	339.8	352.1	581.7
Adjusted R2	0.007	0.009	0.344	0.273	0.273	0.012
Observations	206,891	200,526	206,891	189,870	189,870	200,526

Notes: Dependent variable is percent choosing lowest cost plan, where lowest cost plan is defined using a perfect foresight assumption. Standard errors 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)	-7.365^{***}	-7.341^{***}	-7.357^{***}	-7.293^{***}
	(0.108)	(0.113)	(0.113)	(0.113)
Stakes Squared	0.606^{***}	0.616^{***}	0.618^{***}	0.611^{***}
	(0.010)	(0.011)	(0.011)	(0.011)
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	608.0	595.7	595.3	596.6
Adjusted R2	0.045	0.068	0.072	0.075
Observations	102,455	98,541	98,541	98,541

Notes: Estimates from linear probability model where dependent variable is the indicator variable for whether the individual chooses the lowest cost plan. Standard errors 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 Random Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
Premium (100s)	-0.309^{***}	-0.348^{***}	-0.621^{***}	-0.376^{***}	-0.638^{***}	-0.637^{***}
	(0.004)	(0.004)	(0.030)	(0.004)	(0.030)	(0.030)
sd	0.211^{***}	0.205^{***}	0.205^{***}	0.206^{***}	0.206^{***}	0.205^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Premium × Indiv. avg stakes				0.023^{***}	0.021^{***}	0.021^{***}
				(0.001)	(0.002)	(0.002)
Premium × Stakes		0.023^{***}	0.022^{***}	0.009^{***}	0.010^{***}	
		(0.001)	(0.001)	(0.001)	(0.001)	
Premium × Stakes × $\mathbb{1}(\Delta > 0)$						0.008^{***}
						(0.002)
Premium × Stakes × $\mathbb{1}(\Delta < 0)$						0.012^{***}
						(0.002)
Out-of-Pocket Cost (100s)	0.048^{***}	0.063^{***}	0.161^{***}	0.082^{***}	0.175^{***}	0.178^{***}
	(0.005)	(0.007)	(0.035)	(0.008)	(0.035)	(0.035)
sd	-0.107^{***}	-0.108^{***}	-0.109^{***}	0.099^{***}	-0.101^{***}	-0.103^{***}
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
OOP × Indiv. avg stakes				-0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)
OOP × Stakes		0.001	0.001	-0.001	-0.001	
		(0.001)	(0.001)	(0.001)	(0.001)	
OOP × Stakes × $\mathbb{1}(\Delta > 0)$						-0.000
						(0.001)
OOP × Stakes × $\mathbb{1}(\Delta < 0)$						-0.001
						(0.001)
Premium × Z_i	No	No	Yes	No	Yes	Yes
OOP × Z_i	No	No	Yes	No	Yes	Yes
Log Likelihood	-112,168	-111,912	-111,493	-111,781	-111,384	-111,380
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-6
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 × $\mathbb{1}(\text{Stakes} > 0)$						0.005^{***}
						(0.001)
Premium × Stakes × $\mathbb{1}(\text{Stakes} < 0)$						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 × $\mathbb{1}(\text{Stakes} > 0)$						0.000
						(0.000)
OOP × Stakes × $\mathbb{1}(\text{Stakes} < 0)$						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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-7
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 D. 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-8
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	54452.83		50850.09	
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.

Table A-9
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		Heterogenous Prior Mean Plan Average OOP	
	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)
Marginal cost of information (β^{λ})						
Constant	2.7906	(0.1261)	2.2063	(0.0770)	2.2308	(0.0798)
Zip Income	0.0006	(0.0010)	0.0004	(0.0007)	0.0010	(0.0007)
Zip Education	−0.0019	(0.0020)	−0.0025	(0.0014)	−0.0031	(0.0015)
Age	0.3816	(0.0631)	0.2426	(0.0434)	0.1998	(0.0418)
Age^2	−0.0023	(0.0004)	−0.0014	(0.0003)	−0.0012	(0.0003)
Female	0.0106	(0.0389)	0.0086	(0.0277)	0.0103	(0.0284)
Part D Experience	−0.3265	(0.0197)	−0.1963	(0.0113)	−0.1701	(0.0104)
Rural	0.1753	(0.0505)	0.1121	(0.0362)	0.0836	(0.0367)
Has alzheimers	0.0876	(0.0650)	0.0523	(0.0488)	0.0596	(0.0503)
Has lung disease	0.1392	(0.0622)	0.0601	(0.0442)	0.0769	(0.0459)
Has kidney disease	−0.0961	(0.0504)	−0.0996	(0.0370)	−0.0921	(0.0379)
Has heart failure	0.0552	(0.0560)	0.0273	(0.0413)	0.0221	(0.0425)
Has depression	−0.0371	(0.0563)	−0.0493	(0.0406)	−0.0474	(0.0415)
Has diabetes	0.0504	(0.0442)	−0.0255	(0.0317)	−0.0037	(0.0327)
Has other chronic condition	−0.0072	(0.0435)	−0.0413	(0.0309)	−0.0321	(0.0317)
Mean price sensitivity	−0.1311		−0.1346		−0.1466	
Mean marginal cost of information	3.2413		3.1758		3.7474	
LL	50,735.50		50,151.07		50,410.02	
Observations	1,035,319		1,035,319		1,035,319	

Notes: 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. 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.

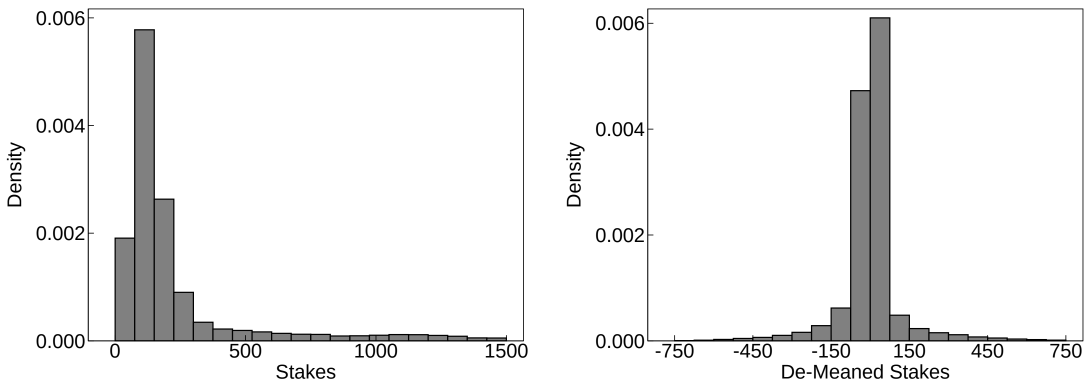
Table A-10
Counterfactual Spending and Welfare for Restricted Choice Set and Out-of-Pocket Cap from Alternative Demand Models

	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 D. 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.

F Appendix Figures

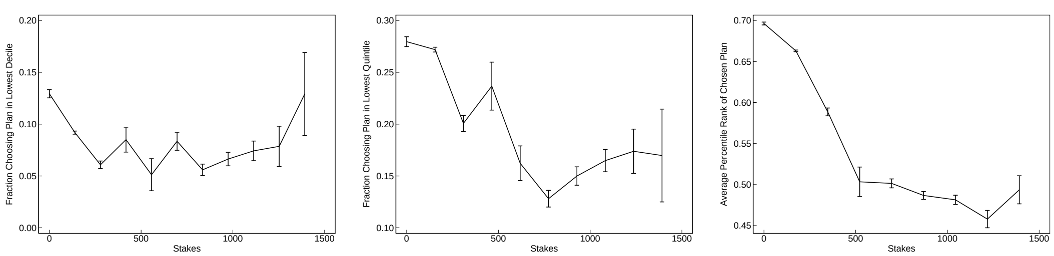
Figure A-1
Variation in Stakes



a. Variation in stakes   b. Within-individual variation in stakes

Notes: Stakes are defined as the standard deviation in annual out-of-pocket cost across plans in an individual's choice set.

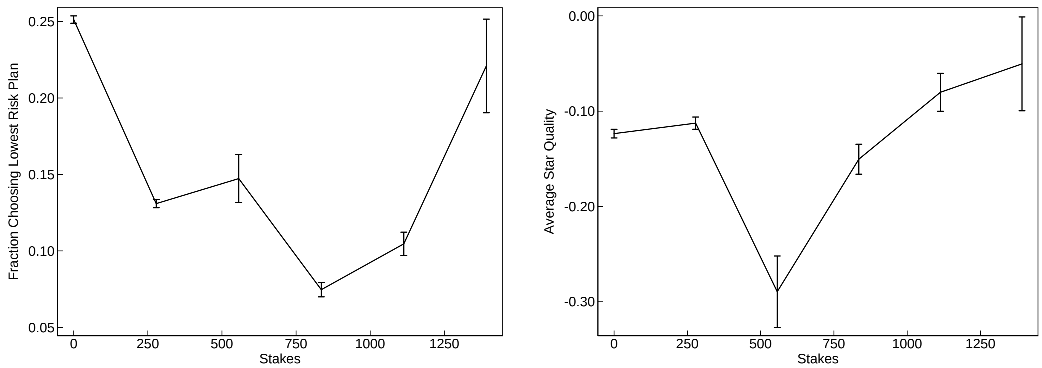
Figure A-2
Alternative Measures of Probability of Choosing Low Cost Plan by
Stakes



a. Plan in Lowest Decile   b. Plan in Lowest Quintile   c. Average Percentile Rank

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

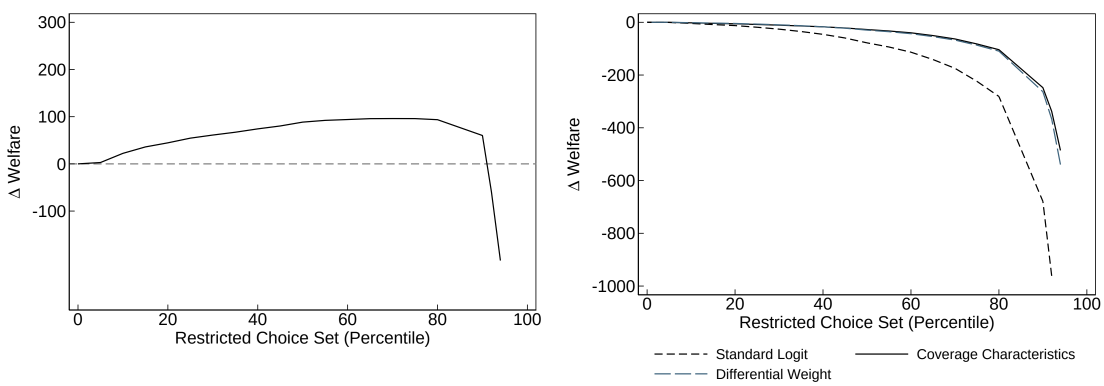
Figure A-3
Alternative Measures of Choice Quality by Stakes



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.

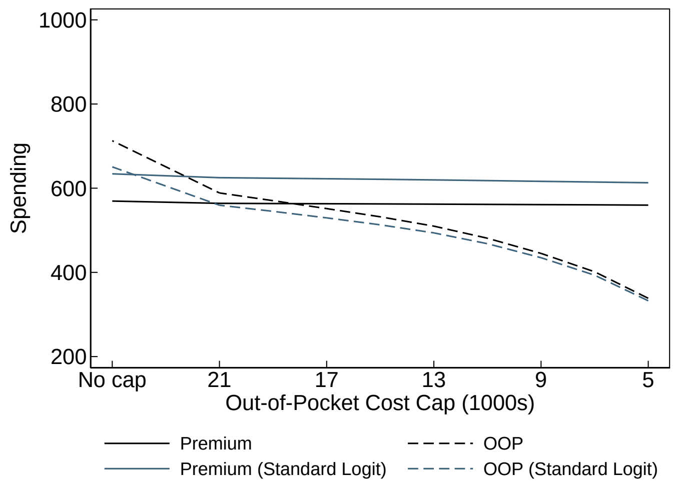
Figure A-4
Counterfactual Welfare Effects of Restricted Choice Set



a. Endogenous Information Model b. Alternative Models

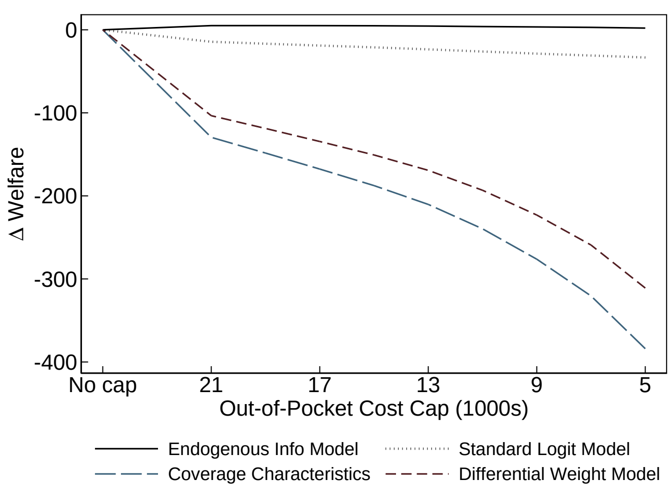
Notes: Chart shows counterfactual average change in welfare per enrollee from removing plans with mean utility below a given percentile where average utlity is computed for each individual. Counterfactual estimates from model with endogenous information acquisition are contrasted with counterfactual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

Figure A-5
Counterfactual Spending for Out-of-Pocket Cost Cap



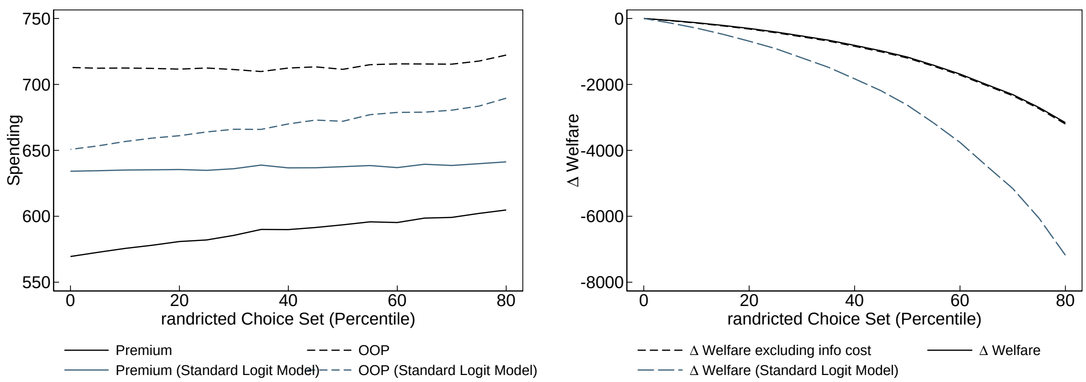
Notes: Chart shows counterfactual change in spending from capping out-of-pocket cost at different levels. Counterfactual estimates from model with endogenous information acquisition is contrasted with counterfactual estimates from alternative demand models without endogenous information.

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.

Figure A-7
Counterfactual Analysis of Randomly Removing Options from Choice Set

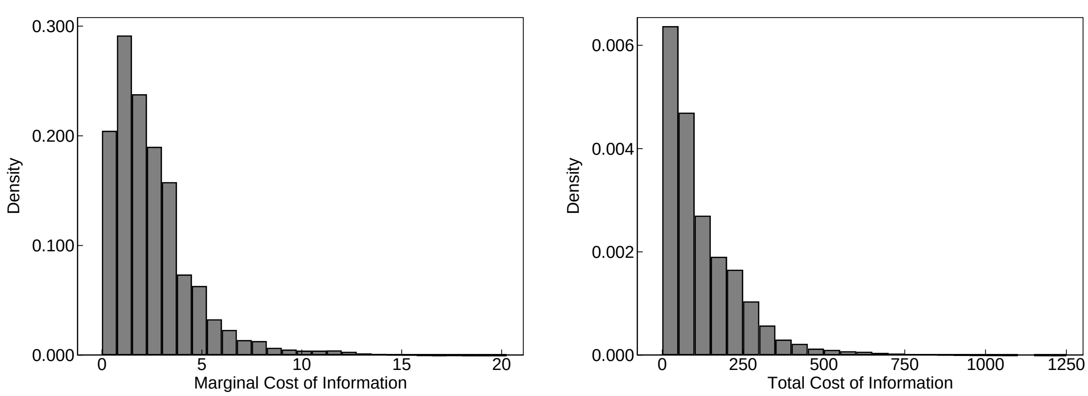


a. Spending

b. Change in Welfare

Notes: Chart shows counterfactual spending and change in welfare from randomly removing a percentage of plans. Counterfactual estimates from model with endogenous information acquisition using welfare as calculated in Section 5.2 is contrasted with welfare excluding taste shocks.

Figure A-8
Distribution of Cost of Information



a. Marginal Cost of Information

b. Total Cost of Information

Notes: Left chart shows histogram of λ_{it}, the marginal cost of information. Right chart shows histogram of the total cost of information, $\hat{C}_{it}$, given by equation (22).

G Identification

For simplicity, consider the baseline model we estimate in which individuals hold common priors for all options. The choice probabilities are given by

$$P_{ijt} = \frac{\exp\left[\frac{\alpha_i v_{ijt}+\beta_1 X^u_{jt}+\beta_2\widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{jt}+\ell_{it}\beta_3 X^k_{jt}}{k_{it}\lambda_{it}(\ell_{it}-1)}\right]}{\sum_{k\in\mathcal{J}_{it}}\exp\left[\frac{\alpha_i v_{ikt}+\beta_1 X^u_{kt}+\beta_2\widetilde{\sigma^2_{ikt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i \ell_{it} p_{kt}+\ell_{it}\beta_3 X^k_{kt}}{k_{it}\lambda_{it}(\ell_{it}-1)}\right]}. \tag{A-19}$$

where

$$\ell^2_{it} \equiv \frac{6\sigma^2_{it}}{\pi^2\lambda^2_{it}}+1, \;\; k^2_{it} \equiv \frac{\ell^2_{it}+\lambda^2_{it}(\ell_{it}-1)^2}{\lambda^2_{it}(\ell_{it}-1)^2}. \tag{A-20}$$

We can redefine the coefficients in equation (A-19) and rewrite the choice probabilities as:

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

where $\rho^0_{it} = \frac{\alpha_i}{k_{it}\lambda_{it}}, \rho^1_{it} = \frac{\beta_1}{k_{it}\lambda_{it}}, \rho^2_{it} = \frac{\beta_2}{k_{it}\lambda_{it}}, \rho^3_{it} = \frac{\alpha_i\ell_{it}}{k_{it}\lambda_{it}(\ell_{it}-1)}$, and $\rho^4_{it} = \frac{\ell_{it}\beta_3}{k_{it}\lambda_{it}(\ell_{it}-1)}$. 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.[44] 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 the following expression.

$$\frac{\rho^3_{it}}{\rho^0_{it}} = \frac{\frac{\alpha_i\ell_{it}}{k_{it}\lambda_{it}(\ell_{it}-1)}}{\frac{\alpha_i}{k_{it}\lambda_{it}}} = \frac{\ell_{it}}{\ell_{it}-1} = \frac{6\sigma^2_{it}+\pi^2\lambda^2_{it}}{6\sigma^2_{it}}$$

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 $(\beta_1, \beta_2, \beta_3)$.

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

H 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 (see appendix A-1 for the derivation and assumptions). We simulate premiums and out-of-pocket costs by drawing from a normal distribution. Table A-11 lists parameter values chosen for the simulation.

Table A-11
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 A-1 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 appendix A-1. Then, we can compute initial choice probabilities based on equation (A-6). We draw choices based on these two sets of choice probabilities and estimate the cost of information using maximum likelihood.

Table A-12
Monte Carlo Results

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

Notes: Standard errors are in parentheses.

We simulate 1000 and 5000 choice situations under the two sets of assumptions and repeat each simulation 50 times. Table A-12 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 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.