Endogenous Information Acquisition and Insurance Choice*

Zach Y. Brown
University of Michigan

Jihye Jeon
Boston University

November 2019

Abstract

Insurance contracts are complicated and individuals may choose how much time and effort to spend understanding and comparing plans. Building on the rational inattention literature, we develop a parsimonious demand model in which individuals choose how much to research difficult to observe characteristics, affecting the accuracy of their beliefs and subsequent choices. The model predicts that individuals acquire more information when the stakes are higher. Using prescription drug insurance data, we show that the model provides an explanation for behavior that is inconsistent with standard demand models. We estimate an empirical model of insurance demand and find that the marginal cost of acquiring information is higher for older enrollees and those with less previous experience choosing a plan. Counterfactual analysis sheds light on the welfare losses due to information frictions and how policy makers can restrict plan choice or decrease cost sharing to simplify decision-making and raise welfare.

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

*We thank Francesco Decarolis, Kate Ho, Ying Fan, Marc Rysman, Andre Veiga, and numerous seminar participants for helpful comments and suggestions. Zach Brown is grateful for 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

When individuals are deciding between complicated alternatives, they may first choose how much time and effort to put into learning about their options. If individuals face low stakes then they may choose to acquire little information and resulting choices will largely be based on easy to observe characteristics or initial beliefs. Conversely, individuals facing high stakes may incur significant cost conducting research in order to make an informed choice.

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 how much time and effort to put into learning about out-of-pocket costs for each plan, as well as other opaque characteristics. Depending on the cost of information acquisition and who has incentives to acquire it, some individuals may choose dominated insurance plans, with important implications for regulation of insurance markets.

We develop a micro-founded theoretical framework for examining demand in the presence of an information acquisition cost for a subset of product characteristics. We focus on insurance choice, although the model can be applied more broadly. The model generates predictions that are distinct from those of standard discrete-choice demand models. We find reduced-form evidence consistent with the model's predictions using data from Medicare prescription drug insurance, also known as Medicare Part D. We then use an empirical model directly based on the theoretical framework to examine the welfare effects of information frictions through counterfactual analysis.

The model builds on theoretical work incorporating rational inattention in discrete choice models (Matějka and McKay 2015). In the model, individuals first decide how much to research their options given easy to observe information such as plan premiums and their prior beliefs. The more research that 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 those expecting few claims, acquire less information than individuals with large consequences. When individuals choose not to acquire much information, they mainly base their decision on easily observed characteristics such as the premium. Due to the fact that some characteristics are always observed, this generates a non-monotonic relationship between the stakes and the quality of choices that individuals make. In addition, the model implies that the relative weight that individuals place on premiums versus out-of-pocket cost depends on the stakes.

Reduced-form results are 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 address the concern that preferences may be 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 better choices, which suggests they are acquiring more information. Therefore, the individual's demand is more elastic with respect to out-of-pocket cost in these years. These results are not consistent with standard models of insurance demand that have been previously used in the literature.

We develop an empirical model directly based on the theoretical framework. A key challenge of the rational inattention framework is that the complexity of the model generally makes estimation infeasible. Importantly, we derive a novel analytical solution for choice probabilities incorporating preference heterogeneity, allowing for a feasible estimation strategy. This allows us to recover the marginal cost of information, a key policy-invariant parameter. By incorporating heterogeneous preferences, including a taste shock, the model allows for the fact that individuals may not always choose the lowest cost plan even if they have full information. We also allow for heterogeneous marginal cost of information. For instance, researching plans may be easier when individuals have previous experience with Medicare Part D 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 $587 to $645 but annual out-of-pocket costs would decline from $668 to $577. Assuming that these results can be applied to all enrollees, this implies savings of about $1.1 billion for the overall Medicare Part D market. 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 $459 per individual excluding information acquisition costs. The average annual information acquisition costs are $273 per enrollee for those making active choices. However, given heterogeneity in the marginal cost of information and the incentives to acquire information, there is large variation in individuals' cost of information acquisition.

A key policy question we ask is how restricting the choice set affects welfare. Policy makers often set minimum standards for insurance plans, implicitly restricting the choice set.[1] In standard demand models, restricting the choice set strictly decreases welfare, which seems at odds with individuals' strong desire for a reduced and simplified choice set documented in existing papers and surveys.[2] By contrast, the model presented in this paper implies that policy makers can increase welfare by strategically removing plans. This simplifies the choice set, lowering information acquisition costs and reducing the likelihood of poorly informed individuals choosing dominated plans. Removing plans with average utility in the lowest quartile increases annual welfare by $364 per enrollee, almost half of which is due to a reduction in information acquisition costs. However, if the choice set is restricted too much, individuals with heterogeneous preferences cannot find a plan that is a good match, reducing welfare.

Next, we use the model to simulate changes to cost sharing. Cost sharing is increasing for Medicare prescription drug plans even though premiums have remained fairly constant.[3] On one hand, this potentially raises the stakes for individuals, increasing incentives for re-

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

[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] See Figure 3.

search. On the other hand, it increases part of the plan characteristics that is hard to observe, possibly leading to larger mistakes. Therefore, it is important to understand how information acquisition and plan choice change when cost sharing changes. In counterfactual simulations, we find that imposing an out-of-pocket cost cap substantially lowers information acquisition costs and reduces the probability that individuals "accidentally" choose low premium but high out-of-pocket cost plans. A $15,000 out-of-pocket cap increases welfare by $184, substantially more than what would be estimated using a standard demand model. The results imply that policy makers should take into account how cost sharing affects incentives for information acquisition.

The counterfactual analysis focuses on the demand-side effects of endogenous information acquisition, however we also argue there are potentially important implications for insurer competition. Using the model, we find demand elasticity with respect to premiums is -1.2 while demand elasticity with respect to out-of-pocket costs is only -0.4, implying that insurers have more incentive to compete on premiums.[4]

We also examine the implications of endogenous information acquisition for adverse selection. While an influential theoretical literature on adverse selection implies that high risk types always choose high coverage plans, there is evidence that this is not always the case (Cutler et al. 2008). One explanation for this puzzle is that information frictions make it difficult to choose high coverage plans, especially for those with a high marginal cost of information. We use the empirical model to examine this issue and find that costly information acquisition attenuates adverse selection relative to the full information case.

Finally, we briefly discuss how the model can be applied to other settings with endogenous information acquisition, especially complex choices in which some characteristics are easier to observe than others.

1.1 Related Literature

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

[4]There are implications for other dimensions of insurer competition, such as the number and complexity of plan offerings.

insurance plans and the fact that individuals are not using all available information. For instance, Handel and Kolstad (2015) survey individuals choosing health insurance and find that they do not fully understand the insurance plans, making it difficult to choose correctly. There is also evidence that consumers respond to easy-to-use information (Kling et al. 2012; Bundorf et al. 2019). In related work, Handel et al. (2015) examine the implications for regulation of insurance markets using a model with exogenous information frictions.

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. (2019). Other theoretical work incorporating rational inattention in a discrete choice framework includes Caplin et al. (2016a) and Caplin et al. (2016b). In this framework, decision makers choose how much and what type of information to acquire. Given a cost of acquiring information, which is proportional to the change in entropy, individuals optimally learn about the payoff structure of various options. There is very limited work incorporating the rational inattention framework into structural models. One exception is recent work by Joo (2019) who uses the rational inattention framework to examine quantity-surcharges for laundry detergent.

A related literature also examines consumer inertia in insurance plan choice (e.g. Handel 2013; Polyakova 2016; Ho et al. 2017; Abaluck and Adams 2018). Individuals may remain enrolled in a high cost plan even when there are lower cost plans that become available. In contrast to this literature, we focus on how individuals tradeoff premium and out-of-pocket payments. In our empirical analysis, we examine individuals that cannot choose to remain enrolled in a previous plan and have to make an active choice, allowing us to abstract from switching costs. However, information acquisition costs can provide a micro-foundation for endogenous switching costs. Individuals may choose to remain enrolled in an insurance plan to avoid repeatedly incurring information acquisition costs.[5]

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 (e.g. Hortaçsu and Syverson 2004; Hong and Shum 2006;

[5]Nudging individuals to switch plans would then impose an additional cost on individuals. Also see related discussion in Ho et al. (2017).

De Los Santos et al. 2012; Seiler 2013; Honka 2014). In search models, individuals start with full information about one option "for free" and then pay a cost to become fully informed about other options in their choice set, searching either sequentially or non-sequentially. One implication is that consumers will, at a minimum, have full information about the option they choose. In contrast to standard search models, 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 addition, 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 insurance plans, including their chosen option.[6] 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.

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 information. We leverage theoretical results linking the rational inattention framework with discrete choice models (Matějka and McKay 2015; Fosgerau et al. 2019). 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. The results from the simple theoretical framework help fix ideas and motivate our reduced-form analysis in the following section. In Section 5, we present a

[6] See, for instance, Handel and Kolstad (2015).

richer empirical framework that accounts for individual risk aversion, preferences over non-cost characteristics, and 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 solves

$$\max_{j \in J} \left[-p_j - \mathbb{E}_{B_i}[v_{ij}]\right] \tag{2}$$

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 $F(\mathbf{v}_i|\mathbf{s}_i)$.

Given constant marginal cost of information λ, total cost of information takes the form

$$c(F) = \lambda \left(-\int_{v_i} g(\mathbf{v}_i) \log g(\mathbf{v}_i) d\mathbf{v}_i + \mathbb{E}_{\mathbf{s}_i}[\int_{\mathbf{v}_i} f(\mathbf{v}_i|\mathbf{s}_i) \log f(\mathbf{v}_i|\mathbf{s}_i) d\mathbf{v}_i] \right) \tag{3}$$

where $g(\mathbf{u}_i)$ is the pdf of the prior and $f(\mathbf{u}_i|\mathbf{s}_i)$ is the pdf of the joint distribution of signals and payoffs.

As is standard in the rational inattention literature, the cost of information is proportional to the change in entropy between the prior and signal, often referred to as the mutual information. This can be thought of as a measure of the reduction in uncertainty after signals are received. This cost function has attractive properties and is meant to reflect the time and cognitive load necessary to acquire and process information. In particular, the cost function is consistent with an individual asking a series of yes-no questions with a fixed cost per question.[7]

The individual chooses an information acquisition strategy that solves

$$\max_{F(\mathbf{s}_i,\mathbf{v}_i)\in\Delta(\mathbb{R}^{2N})} \int_{\mathbf{v}_i} \int_{\mathbf{s}_i} \max_{j\in J} \left[-p_j - \mathbb{E}_{B_i}[F(\cdot\,|\mathbf{s}_i)] F(d\mathbf{s}_i|\mathbf{v}_i) G(d\mathbf{v}_i)\right] - c(F) \tag{4}$$

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

After information acquisition, individuals maximize expected utility, $\mathbb{E}_{B_i}[u_{ij}] = -p_j - \mathbb{E}_{B_i}[v_{ij}]$. Posterior beliefs, B_i, are determined by the signals received by the individual. 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.

$$P_{ij} = \frac{P^0_{ij} e^{(-p_j - v_{ij})/\lambda}}{\sum_{j=1}^{N} P^0_{ij} e^{(-p_j - v_{ij})\lambda}} \tag{5}$$

[7]See discussion in Cabrales et al. (2013) and Matějka and McKay (2015).

The choice probabilities in equation 5 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.

$P^0_{i1}, .., P^0_{iN}$ can be interpreted as the expected choice probabilities for each option in the choice set based the prior but before the realization of signals. These are obtained from solving

$$\max_{P^0_{i1},..,P^0_{iN}} \int_{\mathbf{v}} \lambda \log \Sigma_j P^0_{ij} e^{(-p_j - v_{ij})/\lambda} G(d\mathbf{v}) \text{ s.t. } \sum_j P^0_{ij} = 1, P^0_{ij} \geq 0 \; \forall j \tag{7}$$

We show that there exists a closed-form solution for P^0_{ij}, and therefore P_{ij}, under relatively innocuous assumptions. Developing a tractable model of demand with endogenous information is important for a few reasons. First, equation 7 makes it difficult to interpret demand 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 assume that individuals have a common prior 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. Furthermore, we assume that the distribution of the prior, $G(\mathbf{v})$, follows the conjugate of a scaled EV1 distribution.[8]

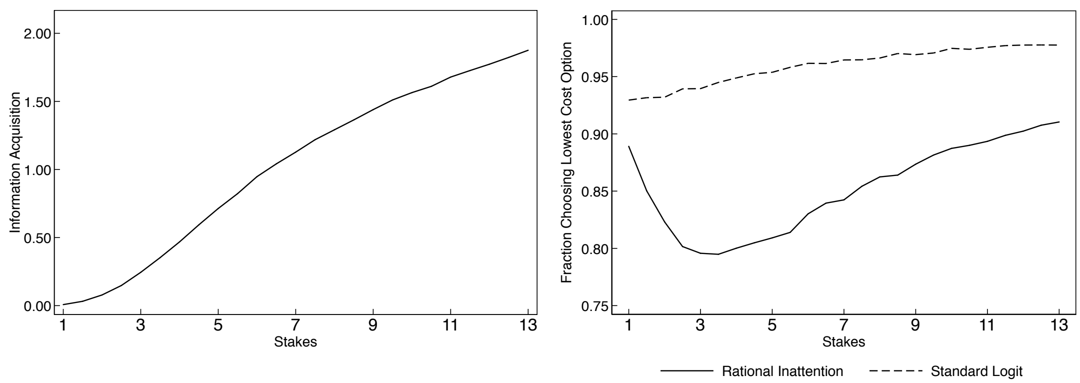
We show that choice probabilities now take a relatively simple form given by

$$P_{ij} = \frac{e^{(-p_i l/(l-1) - v_{ij})/\lambda}}{\sum_k e^{(-p_k l/(l-1) - v_{ik})/\lambda}} \tag{8}$$

where $l^2 = \frac{6\sigma^2}{\pi^2 \lambda^2} + 1$. We present the derivation of equation 8 and discuss the key distributional assumption in Appendix A. In Appendix G we conduct a Monte Carlo exercise to assess the importance of the distributional assumption regarding the prior and argue the the model is an accurate approximation even if the distribution of the prior is misspecified and

[8] This implies that when v is added to a random variable with a type 1 extreme value distribution, the resulting distribution is also type 1 extreme value. 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 mean fraction of individuals choosing lowest cost option from simulations based on endogenous information model with 3 options, $\lambda = 2$, $\sigma = 10$, and premium standard deviation of 4.

is actually normally distributed.

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

$$\tilde{u}_{ij} = -p_j - \underbrace{\frac{(l-1)}{l}\left(v_{ij} - \lambda e_{ij}\right)}_{\substack{\text{Expected} \\ \text{OOP}}}. \tag{9}$$

The expected out-of-pocket cost 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

$$\tilde{u}'_{ij} = -\underbrace{\frac{l}{\lambda(l-1)}}_{\substack{\text{Premium} \\ \text{Weight}}} p_j - \underbrace{\frac{1}{\lambda}}_{\substack{\text{OOP} \\ \text{Weight}}} v_{ij} + \underbrace{e_{ij}}_{\substack{\text{Normalized} \\ \text{Belief Error}}} \tag{10}$$

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.

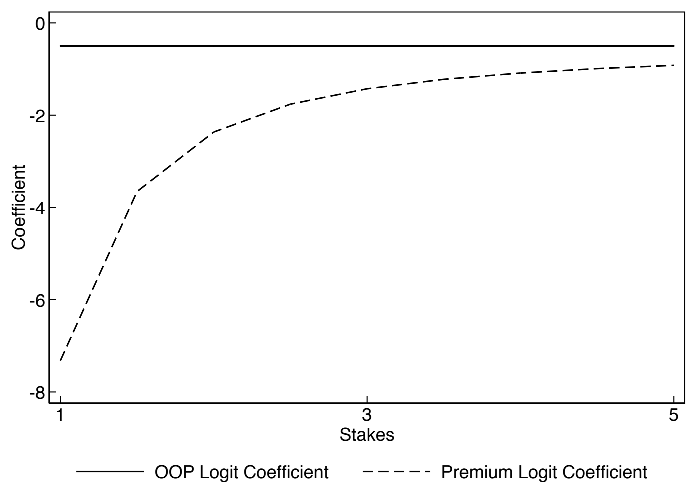
The model implies that choices depend on the stakes. To see this, note that σ, the standard deviation of the prior, can be interpreted as a measure of the stakes. When individuals have a less precise prior, i.e. when σ is large, individuals are more worries about making an incorrect choice when uninformed 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 3.

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 acquire little information, they often choose correctly just by choosing a plan with low premiums. As the stakes grow and comparisons become more complex, it becomes more difficult for individuals to choose the lowest cost plan despite the fact that they are acquiring more information. This implies a positive relationship between stakes and overspending. However, once the stakes are large enough, individuals become highly informed given the strong incentive to acquire information. In this range, there is a negative relationship between stakes and overspending.

Our model of endogenous information acquisition can be contrasted with standard demand models assuming full information. If utility is only a function of the cost, 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}, e.g. 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 10, the coefficient on v_{ij} is solely a function of the cost of information, however the coefficient on p_j depends

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, $\lambda = 2$, $\sigma = 10$, and premium standard deviation of 4.

on both the 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.[9] These elasticities converge as the stakes increase or the 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, premiums may simply be more "salient" than out-of-pocket costs when individuals are making a plan choice. In this case, 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

[9]The elasticities are derived in Appendix A.

2.

In Section 4 we test the predictions of the model using data on 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.

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 with high cost to research their Medicare plans.[10]

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.[11] In surveys, individuals often report that the plans are still too complicated and difficult to compare.[12] 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

[10]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.

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

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

Table 1
Summary of Insurance Choice for Forced Switchers

	Mean	SD
Demographics:		
Age	76.8	7.3
Female	0.606	0.489
Zip income (1,000s)	76.1	34.5
Zip education (pct BA)	29.2	16.9
Rural	0.072	0.259
Years enrolled in Part D	5.67	2.20
Alzheimers	0.093	0.290
Lung disease	0.109	0.312
Kidney disease	0.171	0.377
Heart failure	0.145	0.352
Depression	0.120	0.325
Diabetes	0.281	0.449
Other chronic condition	0.323	0.468
Chosen option:		
Annual premium	504.9	165.0
Out-of-pocket cost (RE)	681.5	978.1
Out-of-pocket cost (PF)	673.3	1379.7
Total spending	1178.2	1396.3
Relative to least expensive option:		
Difference (RE)	425.5	531.6
Percent difference (RE)	0.32	0.16
Difference (PF)	459.4	1217.2
Percent difference (PF)	0.34	0.20
Plans in Choice Set	26.5	6.9
Number of individuals	64,071	
Choice situations	84,193	

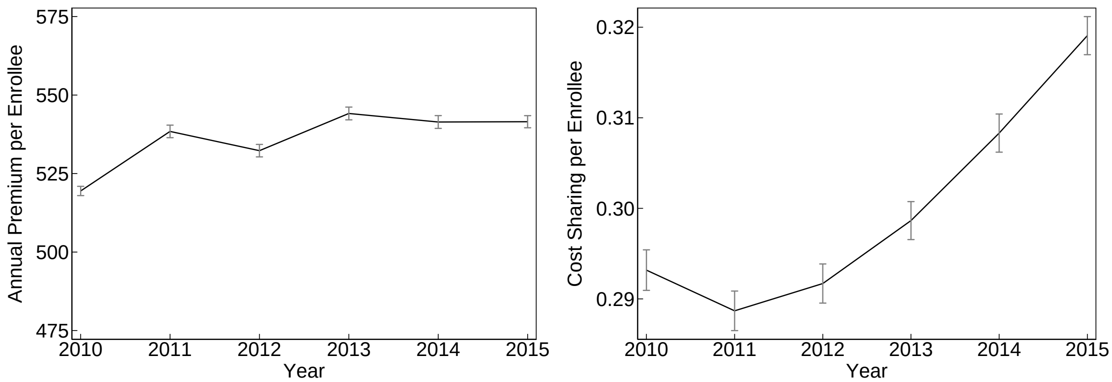
the cost of information to be lower in the market for Medicare Part D plans relative to other insurance markets.

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 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 when detailed drug formulary data becomes available. This allows us to more accurately construct out-of-pocket costs.[13]

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 cost of information goes to zero (or infor-

[13] Abaluck and Gruber (2016) use an earlier sample requiring formularies to be inferred from observed out-of-pocket payments.

Figure 3
Premiums and Cost Sharing Trends in Medicare Part D



a. Annual Premium

b. Cost Sharing

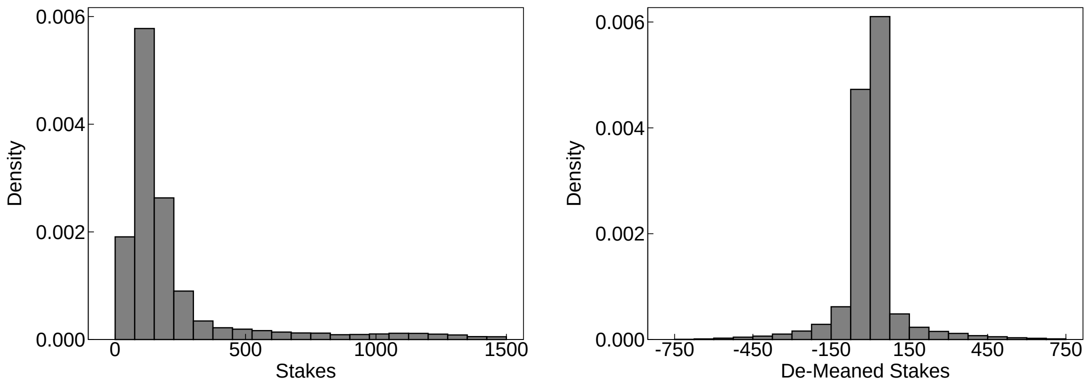
Notes: Cost sharing defined as out-of-pocket cost as a fraction of gross drug cost. Standard error bars show 95% confidence interval for the mean.

mation 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 expected out-of-pocket costs for chosen plans are very 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. In this approach, we assume 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.

Table 1 shows the calculated out-of-pocket costs for the two measures, rational expecta-

Figure 4
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.

tions (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.

The previous literature has documented the importance of consumer inertia in plan choice (e.g. Handel 2013; Polyakova 2016; Ho et al. 2017). 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. Importantly, these individuals do not have previous experience with any of the plans in their choice set, implying that they likely start with a common prior across options. Due to a change in 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. After these restrictions, we use a 1 percent sample for the primary analysis, 84,193 choice situations.

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 the individuals in the sample are presented in Table 1. The demographics of individuals that are forced to make an active choice are very similar to the demographics of the overall Medicare Part D population.[14]

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.[15] Therefore, the stakes are low when when expected out-of-pocket costs are similar across 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.[16] However, it is important to note that the stakes are not always higher when individuals face higher out-of-pocket costs. For instance, if individuals face very high out-of-pocket costs, they may hit the catastrophic coverage portion of Medicare Part D plans, leading to low variance in cost across plans. In this case, the individual could face relatively low stakes.

4 Reduced-Form Evidence

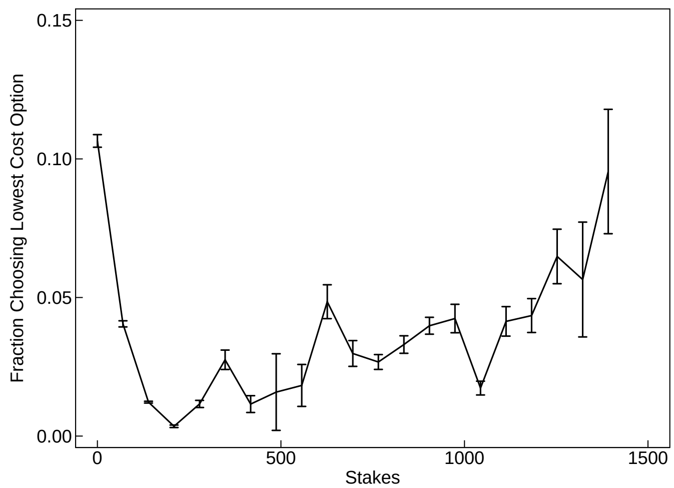
Motivated by the results of the model in Section 2, we now examine how insurance plan choice is affected by the stakes. We use individual-level data on Medicare prescription drug plan choice and exploit within-individual variation, i.e. the same individual who makes an active plan choice facing different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

[14] See Appendix A-2.

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

[16] See Appendix Table A-1.

Figure 5
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.

Stakes and Overspending

We start by examining the relationship between the fraction of individuals choosing the lowest cost plan and the stakes. Figure 5 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, there are concerns that individuals facing high stakes have different preferences than individuals facing low stakes.

In order to examine the causal effect of stakes on the fraction of individuals choosing the lowest cost plan, we exploit within-individual variation. For individual i in year t, we estimate the following linear probability model

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

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

	(1)	(2)	(3)	(4)	(5)
Stakes (100s)	-2.443^{***}	-1.969^{***}	-0.448^{***}	-0.521^{***}	-0.392^{***}
	(0.055)	(0.055)	(0.078)	(0.078)	(0.079)
Stakes Squared	0.222^{***}	0.184^{***}	0.064^{***}	0.071^{***}	0.059^{***}
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
Plan Characteristic Controls	No	Yes	No	Yes	Yes
Individual FEs	No	No	Yes	Yes	Yes
Year FEs	No	No	No	No	Yes
Implied minimum	551.1	534.0	351.2	369.1	333.0
Adjusted R2	0.010	0.034	0.396	0.399	0.400
Observations	84,193	84,193	84,193	84,193	84,193

Notes: Estimates from linear probability model where dependent variable is percent choosing lowest cost plan. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

where γ_i are individual fixed effects, θ_t are year fixed effects, and X_{it} are average plan characteristics.[17] 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, 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 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 5, 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 \$333, individuals are more likely to choose the lowest cost plan as the stakes increase.

One concern is that there is measurement error stemming from the fact that each indi-

[17] We include controls for star quality, deductible, generic coverage, coverage in the donut hole, and cost sharing. In addition, we control for within-plan out-of-pocket cost variance to account for risk-aversion.

vidual's out-of-pocket costs are predicted based on the average of similar individuals.[18] As a robustness check, we use individual's actual utilization to predict out-of-pocket costs, e.g. a perfect foresight assumption. The regression results are presented in Appendix Table A-3. All of the specification also imply a U-shaped relationship, although the standard errors are slightly larger.

The fraction of individuals choosing the lowest cost plan is only one measure of choice quality. In Appendix Figure A-1 we examine the fraction of individuals choosing a plan in the lowest decile of the plans in their choice set and 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-2. Across all of these alternative outcomes, we find evidence of a U-shaped relationship between the stakes and choice quality.

Stakes and Logit Coefficients

In order to examine the relative weight that individuals place on out-of-pocket cost and premiums, and how this varies according to the stakes, we estimate a model based on the standard logit framework. We use the model to further test the theoretical predictions in Section 2. 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

$$\nu_{ijt} = \alpha_1 p_{jt} + \alpha_2 p_{jt} Stakes_{it} + \gamma_1 v_{ijt} + \gamma_2 v_{jt} Stakes_{it} + \theta \tilde{\sigma}^2_{ijt} + \beta X_{ijt} \tag{12}$$

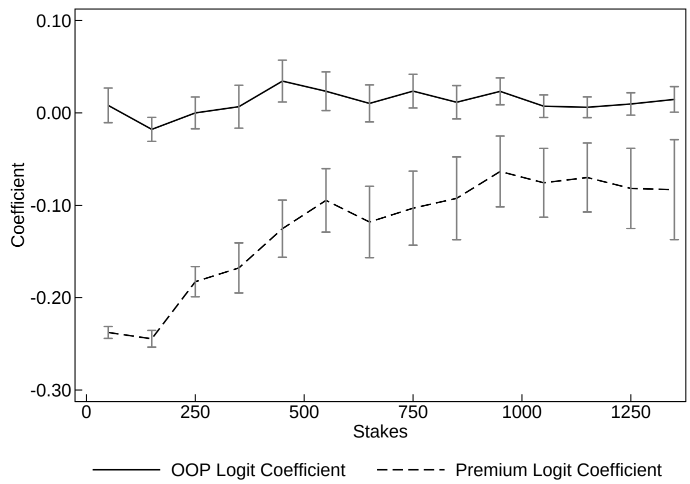
The specification controls for risk aversion by including $\tilde{\sigma}^2_{ijt}$, the variance of out-of-pocket costs for plan j.[19]. We also include other plan characteristics, X_{ijt}. Given additive i.i.d. EV1 error, choice probabilities are $P_{ijt} = \exp[\nu_{ijt}] / \left(\sum_k \exp[\nu_{ikt}]\right)$.

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

[18]The concern about measurement error in out-of-pocket costs is also alleviated by the fact that it is unlikely to generate the non-monotonic relationship seen in the data.

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

Figure 6
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.

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 6 presents the results in graphical form by interacting stake bins with coefficients on premium and out-of-pocket cost.[20] When the stakes are low, individuals appear to place a high value on reducing premiums relative to the value that they place on reducing out-of-pocket cost, i.e. the coefficient on premium is low relative to the coefficient on out-of-pocket cost. This is consistent with the idea that individuals do not have incentive to become informed about out-of-pocket costs. As the stakes rise, the relative weight that individuals appear to place on premiums declines, consistent with the model predictions depicted in Figure 2.

The results using the specification described in Equation 12 are presented in Table 3 Column 2. Consistent with the model, the interaction of premium and stakes is positive and

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

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

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

Notes: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 identified 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.

Finally, we examine whether the magnitude of the effect is different for an increase in the stakes compared to a decrease in the stakes. The results, which exploit the same within-individual variation, are presented in Table 3 Column 6. Focusing on how the stakes affect the weight that individuals put on premiums, the results imply a statistically significant and positive effect for both an increase and a decrease. However, the effect of a decrease in the stakes is larger in magnitude.

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-4. We also examine the results for both the baseline model and random coefficient model using the alternative definition of out-of-pocket cost, i.e. assuming perfect foresight. Results, presented in Appendix Table A-5 and Appendix Table A-6, also qualitatively similar.

Taken together, the reduced-form evidence implies that individuals respond to incentives to acquire information, consistent with the rational inattention model presented in Section 2. However, analyzing counterfactual welfare requires estimating a model directly based on the rational inattention framework.

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. 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 $\mathcal{J}_{it}$. Individuals have CARA utility, $\exp\left(-\gamma(W - C_{ijt})\right)$, where cost, C_{ijt}, is normally distributed. In particular, let $C_{ijt} \sim N(p_{jt} + v_{ijt}, \widetilde{\sigma^2_{ijt}})$ where p_{jt} is the premium, v_{ijt} is expected out-of-pocket cost, and $\widetilde{\sigma^2_{ijt}}$ is the within-plan variance of out-of-pocket cost.

Indirect utility can then be expressed as $-\alpha\left(\exp(\gamma(p_{jt} + v_{ijt}) + \frac{1}{2}\gamma^2\widetilde{\sigma^2_{ijt}}\right)$. Following the previous literature, we consider a first-order Taylor expansion of indirect utility. Reparameterizing and adding preferences over non-cost characteristics, utility can be expressed as

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

As in Section 2, a key assumption is that p_{jt} is initially observed while v_{ijt} can be observed only if 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^n_{jt}. In the baseline specification, we assume that plan quality 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.

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 of the scaled 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

[21] This implies that when v is added to a random variable with a type 1 extreme value distribution, the resulting distribution is also type 1 extreme value. 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 G.

[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.

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. Given that individuals do not have any previous experience with the plans, their prior variance is given by

$$\sigma^2_{it} = Var_j\left[\xi_{ijt}\right] \tag{14}$$

which is common to all options in an individual's choice set.[23] The prior distribution for each option are assumed to be independent. Let this multivariate distribution have CDF given by $G(\xi)$, following the same distribution as in Section 2.

We use the fact that the distribution of the prior and the taste shock emit a closed form solution for initial choice probabilities before information acquisition. This in turn allows as to derive an expression for choice probabilities after information acquisition:

$$P_{ijt} = \frac{\exp\left[\frac{\alpha_i v_{ijt}+\beta_1 X^u_{jt}+\beta_2\widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i l_{it} p_{jt}+l_{it}\beta_3 X^k_{jt}+\xi^0_{ijt}}{k_{it}\lambda_{it}(l_{it}-1)}\right]}{\sum_{k\in\mathcal{J}_{it}}\exp\left[\frac{\alpha_i v_{ikt}+\beta_1 X^u_{kt}+\beta_2\widetilde{\sigma^2_{ikt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i l_{it} p_{kt}+l_{it}\beta_3 X^k_{kt}+\xi^0_{ikt}}{k_{it}\lambda_{it}(l_{it}-1)}\right]}. \tag{15}$$

where

$$l^2_{it} \equiv \frac{6\sigma^2_{it}}{\pi^2\lambda^2_{it}} + 1,\;\; k^2_{it} \equiv \frac{l^2_{it} + \lambda^2_{it}(l_{it}-1)^2}{\lambda^2_{it}(l_{it}-1)^2} \tag{16}$$

The derivation for Equation 15 is found in Appendix A-2.

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

$$\tilde{u}_{ijt} = \alpha_i\frac{l_{it}}{k_{it}\lambda_{it}(l_{it}-1)}p_j + \alpha_i\frac{1}{k_{it}\lambda_{it}}v_{ij} + \frac{1}{k_{it}\lambda_{it}(l_{it}-1)}\xi^0_{ijt} + \beta_1\frac{1}{k_{it}\lambda_{it}}X^u_{jt} + \beta_3\frac{l_{it}}{k_{it}\lambda_{it}(l_{it}-1)}X^k_{jt} + e_{ij} \tag{17}$$

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

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

to zero. This is given by

$$\lim_{\lambda_{it}\to 0} P_{ijt} = \frac{\exp\left[\alpha_i(v_{ijt} + p_{jt}) + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \beta_3 X^k_{jt}\right]}{\sum_{k\in\mathcal{J}_{it}} \exp\left[\alpha_i(v_{ikt} + p_{kt}) + \beta_1 X^u_{kt} + \beta_2 \widetilde{\sigma^2_{ikt}} + \beta_3 X^k_{kt}\right]} \tag{18}$$

which are choice probabilities under full information.

We now describe the specific assumptions we make regarding heterogeneity in the price coefficient, α_i, the cost of information, λ_{it}, and the prior mean ξ^0_{ijt}. We allow for observable heterogeneity in price sensitivity by assuming

$$\alpha_i = -\exp(\beta^\alpha Z_i) \tag{19}$$

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

$$\lambda_{it} = \exp(\beta^{\lambda 1} Z_i + \beta^{\lambda 2} W_{it}) \tag{20}$$

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 individuals in our sample lack previous experience with any of the plans in their choice set. Given a prior mean that is common across options, ξ^0_{ijt} can be normalized to zero for every option. Since choice probabilities only depend on differences in expected utility, the normalization of the prior is inconsequential.

We also consider specifications in which individuals start with additional information about plans, i.e. allow for heterogeneous prior means. 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 specification that allow for individuals to have a heterogeneous prior that is based on observable characteristics, allow for unobserved plan quality, and use alternative measures of expected out-of-pocket cost. 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 models of demand used in the literature. As a benchmark, we estimate a *standard logit model* assuming that individuals have full information about both premiums and expected out-of-pocket cost. Next, we estimate a model in which demand is a function of premium and coverage characteristics, such as the deductible, rather than expected out-of-pocket cost. This approach is widely used in the empirical literature.[24] We call this model the *coverage characteristics model*. Finally, we estimate a *differential weight model* in which there is a different coefficient on premium and expected out-of-pocket cost. This approach has been previously applied in the context of Medicare Part D.[25]

The details of these alternative models are presented in Appendix D. Parameter estimates are in Table A-8.

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 individuals may have incorrect beliefs, leading to choices that are incorrect ex-post. In addition, total welfare should account for individual's information acquisition cost.

Under full information about product characteristics including out-of-pocket cost, but before the taste shock is known, consumer surplus for individual i in year t takes the usual

[24] This general approach has been used by Bundorf et al. (2012), Handel (2013), Decarolis et al. (2015), Polyakova (2016), Ericson and Starc (2016), Tebaldi (2017), and many others.

[25] See, for example, Abaluck and Gruber (2011), Abaluck and Gruber (2016), and Ho et al. (2017).

form[26]

$$CS_{it}^{FullInfo} = \mathbb{E}_e[\max(u_i)] = \frac{1}{|\alpha_i|}\log\left(\sum_{j\in\mathcal{J}_{it}}\exp(\nu_{ijt})\right) \tag{21}$$

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}.

Consumer surplus with endogenous information is given by

$$CS_{it}^{RI} = \frac{1}{|\alpha_i|}\log\sum_j e^{\tilde{\nu}_{ijt}} + \frac{1}{|\alpha_i|}\sum_j P_{ijt}[\nu_{ijt} - \tilde{\nu}_{ijt}] \tag{22}$$

where $\tilde{\nu}_{ijt} = \frac{\alpha_i v_{ijt}+\beta_1 X_{jt}^u+\beta_2\widetilde{\sigma_{ijt}^2}}{k_{it}\lambda_{it}} + \frac{\alpha_i l_{it} p_{jt}+\xi_{ijt}^0+\beta_3 l_{it} X_{jt}^k}{k_{it}\lambda_{it}(l_{it}-1)}$ is the belief utility excluding the i.i.d. shock e'_{ijt}. The first term is the expected welfare calculated as if beliefs were correct. Note that $1/|\alpha|$ 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 15.[27] Further detail is provided in Appendix A-3.

Following the assumptions of the rational inattention model in Section 2, the cost of information is determined by the mutual information

$$\hat{C}_{it} = \frac{\lambda_{it}}{|\alpha_i|}\mathbb{E}_\epsilon\left(H(G) - \mathbb{E}_s[H(F(\cdot|s))]\right) \tag{23}$$

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

$$\hat{C}_{it} = \frac{\lambda_{it}}{|\alpha_i|}\int_\epsilon\left(-\sum_{j\in\mathcal{J}_{it}} P_{ijt}^0(\epsilon)\log P_{ijt}^0(\epsilon) + \int_\xi\left(\sum_{j\in\mathcal{J}_{it}} P_{ijt}(\xi,\epsilon)\log P_{ijt}(\xi,\epsilon)\right)G(d\xi)\right)M(d\epsilon) \tag{24}$$

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.

[26]See Small and Rosen (1981). Note that this expression is up to a constant.

[27]See Train (2015) which considers the case in which beliefs are exogenously determined.

The welfare loss due to information frictions is then given by

$$\Delta CS_{it} = CS_{it}^{FullInfo} - CS_{it}^{RI} + \hat{C}_{it}. \tag{25}$$

5.3 Identification and Estimation

The key empirical challenge is separately identifying preferences and incorrect beliefs. Specifically, we wish to separately identify the price coefficient and the cost of information. In many applications, information frictions in which an individual receives a imprecise signal of product characteristics imply an error term that is essentially 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. If observed choices are equally sensitive to premiums and out-of-pocket costs, then we conclude that there are no information frictions and heterogenous preferences are largely a result of the taste shock (or preferences over not-price characteristics). This can be seen by noting that, under full information, a change in premiums and an equivalent change in out-of-pocket cost have the same effect on choice probabilities in Equation 18.

Conversely, if choices are more sensitive to premiums than out-of-pocket costs, we conclude that individuals must have an information acquisition cost making it costly to observe accurate out-of-pocket costs. This can be seen by noting that, for $\lambda_{it} > 0$, premiums and out-of-pocket cost enter differently in Equation 15.

Identifying heterogeneity in the price coefficient and cost of information follows a similar argument. If individuals with certain characteristics are more sensitive to both premiums and out-of-pocket cost, this group must have higher price sensitivity. However, if these individuals appear to be more sensitive to premiums relative to out-of-pocket costs holding the stakes fixed, it must be that they have a higher cost of information.

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

[28]One challenge is that the log-likelihood is prone to numerical rounding errors when λ_{it} is large relative to σ_{it}^2

2.

5.4 Empirical Model Estimates and Fit

The parameter estimates from the demand model are presented in Table 4. Focusing on Specification 1, which we take to be the baseline, average price sensitivity for individuals in the sample is estimated to be -0.12. The coefficient on age is highly significant indicating that older individuals are more price sensitive, although the quadratic term is positive and significant indicating that there is a nonlinear effect in age. Individuals in rural areas are less price sensitive.

The average marginal cost of information is 2.9, which can be interpreted as the cost of reducing entropy of beliefs by 1 unit.[29] The cost of information may reflect either the 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 cost of information would apply to the decision maker in question. We find that individuals in high income areas have a higher marginal cost of information, consistent with a higher opportunity cost of time. However, individuals in more educated areas have lower cost, consistent with the idea that it is easier for more educated individuals to research plans.

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. We find that the total cost of information is increasing in age. This can be seen in Figure A-3 which examines mean information cost by demographic characteristics.

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 cost of information acquisition. The distribution of the marginal cost of information and total cost of information is shown in Figure 7. A quarter of individuals in the sample spend less than $60 researching plans, however there are a small set of individuals who are estimated to

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.

[29]Marginal cost is in hundreds of dollars since premium and out-of-pocket costs are scaled for estimation. Note that for a normal distribution, entropy is given by $\frac{1}{2}\log(2\pi e\sigma^2)$. Therefore, for the special case of normally distributed beliefs, the marginal cost of information can be interpreted as the cost of increasing the precision of posterior beliefs by a factor of 100.

Table 4
Estimates for Demand Model with Endogenous Information Acquisition

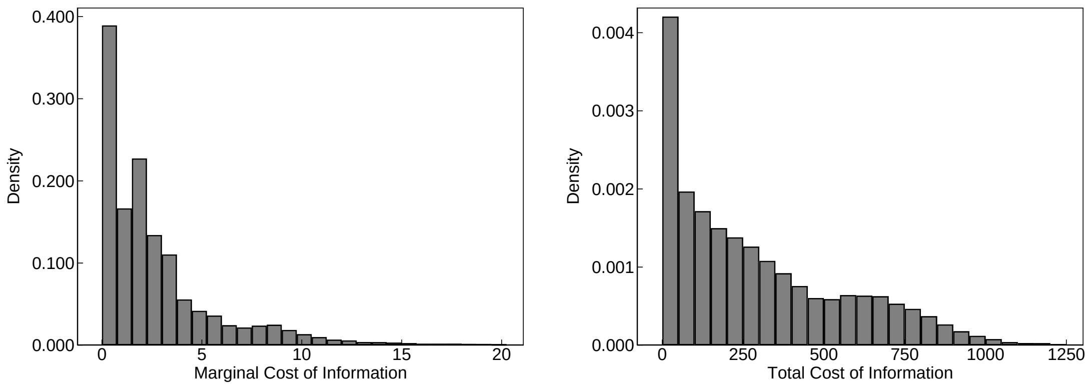
	Specification 1 Homogeneous Prior		Specification 2 Heterogeneous Prior Plan Average OOP	
	Estimate	SE	Estimate	SE
Price Sensitivity (β^{α})				
Constant	−2.1507	(0.0248)	−2.1752	(0.0279)
Income	−0.0004	(0.0008)	0.0009	(0.0008)
Education	−0.0028	(0.0016)	−0.0044	(0.0017)
Age	−0.1614	(0.0340)	−0.1805	(0.0359)
Age^2	0.0009	(0.0002)	0.0010	(0.0002)
Female	−0.0222	(0.0293)	−0.0104	(0.0315)
Rural	0.1259	(0.0520)	0.0789	(0.0572)
Other Plan Characteristics				
Plan Quality	1.5887	(0.0620)	1.6303	(0.0732)
Risk	−0.0343	(0.0017)	−0.0344	(0.0019)
Marginal cost of information (β^{λ})				
Constant	3.6751	(0.1109)	3.5234	(0.1033)
Income	0.0018	(0.0008)	0.0021	(0.0008)
Education	−0.0095	(0.0018)	−0.0095	(0.0018)
Age	0.3751	(0.0553)	0.3481	(0.0544)
Age^2	−0.0022	(0.0003)	−0.0020	(0.0003)
Female	0.0466	(0.0352)	0.0481	(0.0344)
Rural	0.0363	(0.0785)	−0.0178	(0.0758)
Part D Experience	−0.5712	(0.0247)	−0.5428	(0.0247)
Has alzheimers	0.0651	(0.0633)	0.0766	(0.0621)
Has lung disease	0.1135	(0.0585)	0.1397	(0.0580)
Has kidney disease	0.0122	(0.0474)	0.0192	(0.0463)
Has heart failure	0.1456	(0.0521)	0.1430	(0.0511)
Has depression	0.0084	(0.0532)	0.0121	(0.0522)
Has diabetes	0.1329	(0.0408)	0.1527	(0.0405)
Has other chronic condition	0.0197	(0.0402)	0.0254	(0.0394)
Mean price sensitivity	−0.1169		−0.1146	
Mean marginal cost of information	2.9134		2.8359	
LL	123,303.7		123,771.1	
Observations	1,058,745		1,058,745	

Notes: 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.

incur a cost of over $1,000 researching plans.

Specification 2 assumes that individuals have heterogenous prior means for options in their choice set and yields very similar parameter estimates. The mean price sensitivity and marginal cost of information are almost identical to Specification 1. In Appendix A-9 we also consider a specification with plan fixed effects. In this specification, price sensitivity and marginal cost of information are very similar to the baseline specification, indicating

Figure 7
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 24.

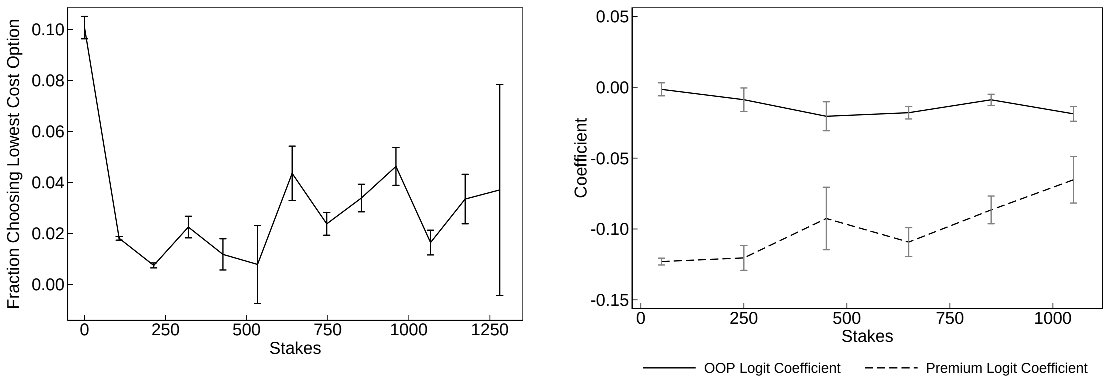
that dimensions of plan quality observed to individual but unobserved to the researcher are not driving the results. Finally, in Appendix A-9 we consider a specification in which individuals use easily observable characteristics—plan premium, deductible, generic coverage, coverage in the gap, and cost sharing—to predict out-of-pocket cost for each plan.[30] Given that individuals are assumed to have even more information before paying an information acquisition cost, the estimated marginal cost of information that rationalizes choices is higher than the baseline specification. Nevertheless, the qualitative conclusions of the model are similar.

As an additional robustness check, we also consider 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 implies quite similar parameter estimates. This can be seen in Table A-10.

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 8

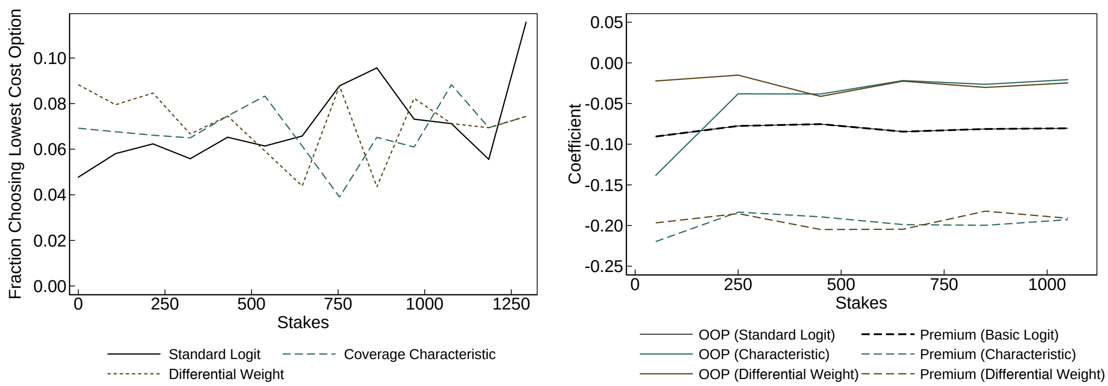
[30]The estimated relationship between these characteristics and out-of-pocket costs is given in Appendix A-7.

Figure 8
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 5 and Figure 6. For further description see Section 2.

show that the model can recover the patterns documented in Figure 5 and Figure 6 in the previous section. We also use the estimates from the alternative models that do not allow for endogenous information. As seen in Panel c and d, these alternative models cannot rationalize 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 $668 while the actual mean is $660. For premiums, it is $587 and $581 respectively. In addition, we examine the difference in cost between the chosen option and the plan with the lowest total cost. Again the model estimates are quite similar. 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.

6 Counterfactual Results

In this section, we explore the implications of the model using counterfactual analysis. We start by simulating insurance demand under full information in order to evaluate the welfare effects of endogenous information acquisition in Medicare Part D. 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 $91 lower, however these plans have premiums that are $58 higher. Given that individuals on average choose a plan that is $425.5 more expensive than the least expensive option, this suggests that individuals have strong preferences over non-cost characteristics such as quality and risk. 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, full information would result in total savings of $1.1 billion per year.[31] In addition, individuals choose plans that are high quality, as measured by Medicare star ratings, and lower risk. Overall, this implies that welfare, excluding information acquisition costs, increases by $459 per enrollee on average. Information acquisition costs are also substantial, averaging $273 per enrollee.

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 over the sample, including enrollees who do not make an active choice, is 33.9 million per year. We assume that savings would 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	660	604	668	577	840	670
Premium of chosen plan	581	637	587	645	618	673
Total cost of chosen plan	1241	1241	1256	1222	1459	1343
Cost difference compared to lowest cost plan	540	550	562	532	692	580
Adverse selection	0.020	0.099	0.024	0.095	0.117	0.173
Δ welfare ex. information acquisition cost				459		263
Δ information acquisition cost				273		646
Δ welfare ex. info acquisition cost (no taste shock)				1076		776
Out-of-pocket Elasticity			-0.37	-1.24	-0.67	-1.35
Premium Elasticity			-1.21	-1.24	-1.28	-1.35

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.

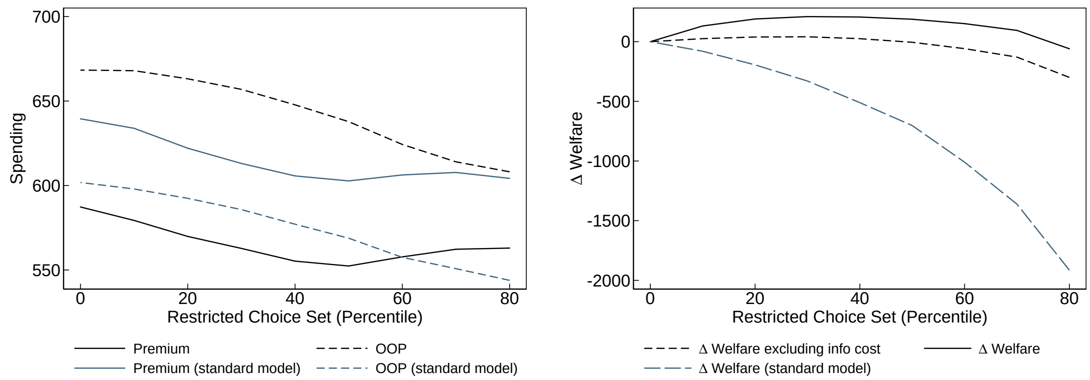
We also examine the implications of the model for adverse selection, which we define as the correlation between spending percentile and coverage choice percentile.[33] Predicted adverse selection from the baseline simulation is quite close the adverse selection in the data at 0.024 and 0.020 respectively. Adverse selection is relatively low in the baseline case, reflecting the fact that individuals cannot easily determine which are the high coverage plans due to information acquisition costs. This can be seen by noting that the degree of adverse selection increases markedly in the full information counterfactual. In the baseline case, the standard model over-predicts the degree of adverse selection. This is due to the fact that the standard model assumes that individuals are equally price sensitive when it comes to premiums and out-of-pocket costs, implying that individuals with high spending choose high coverage plans.

In the baseline case, the estimated elasticity of demand with respect to premiums is -1.2, however the elasticity with respect to out-of-pocket costs is only -0.4.[34] Elasticity with respect

[33]For the purposes of defining spending percentile, we calculate the average out-of-pocket spending across all options in individuals choice set and then consider the percentile rank across individuals in a given year. Coverage choice is defined as the percentile rank of out-of-pocket cost for the chosen option in an individual's choice set.

[34]Abaluck and Gruber (2011) reports a similar level of the average elasticity with respect to premiums ranging

Figure 9
Counterfactual Analysis of Restricted Choice Set



a. Spending

b. Change in Welfare

Notes: Chart shows counterfactual spending and change in welfare from removing plans with mean utility below a given percentile. Counterfactual estimates from model with endogenous information acquisition is contrasted with counterfactual estimates from standard logit demand model.

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 -1.3, the same for both premiums and out-of-pocket costs.

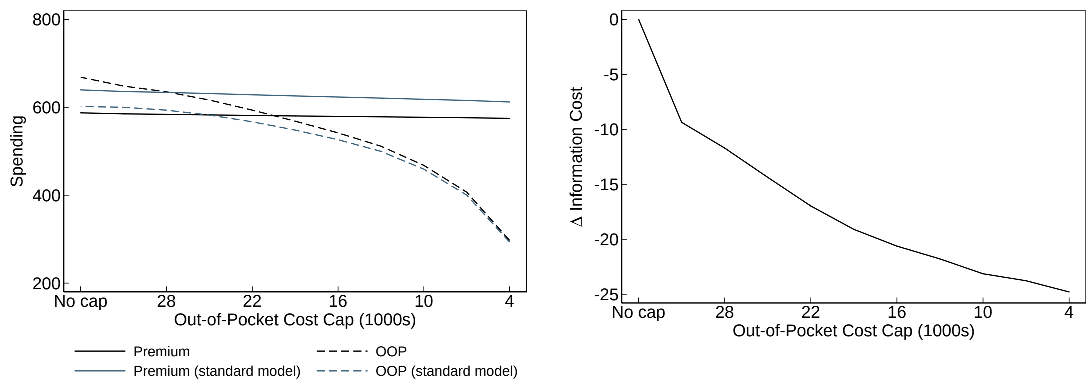
Table 5 also shows the results for individuals with information cost, $\hat{C}_{it}$, in the top quartile. These individuals may face higher stakes and therefore have more incentive to acquire information, or have higher marginal costs of acquiring information. For these individuals, the total cost saving is $116 in the full information case. Although the welfare effects excluding information acquisition costs are lower than the population as a whole, the information acquisition costs are more than double. Under full information, their demand is quite elastic, about -1.4.

Next, we use the model to examine the effect of 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.[35] We ask whether policy

from -0.75 to -1.17.

[35]For instance, in an experimental setting, Iyengar and Kamenica (2010) document that individuals may be better off with a smaller choice set.

Figure 10
Counterfactual Analysis of Out-of-Pocket Cost Cap



a. Spending b. Change in Information Acquisition Cost

Notes: Chart shows counterfactual spending and change in information acquisition cost from capping out-of-pocket cost at different levels. Counterfactual estimates from model with endogenous information acquisition is contrasted with counterfactual estimates from standard logit demand model.

makers can increase welfare by strategically eliminating plan offerings.[36] We simulate plan choices and welfare after eliminating plans with average utility in the lowest decile.[37] We then repeat this procedure after eliminating plans in subsequent deciles. The results are depicted in Figure 9. We assume that individuals are aware that "poor" plans are removed, thus affecting their incentive to research plans.[38]

Figure 9 panel a shows counterfactual spending as plans are removed. In the endogenous information model, individuals are willing to choose high premium plans in exchange for plans with low out-of-pocket costs once the choice set is simple enough. For this reason, premium and out-of-pocket spending starts to converge when the choice set is highly restricted.

When evaluating the welfare effects with endogenous information, there is a trade-off between simplifying the choice set in order to reduce information costs and allowing enough options that individuals can find a plan that is a good fit, i.e. has a high idiosyncratic

[36] 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.

[37] Average utility is defined as $v_{ijt} = \alpha_i(v_{ijt} + p_{jt}) + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}$.

[38] 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	
	25th Percentile Cutoff	50th Percentile Cutoff	\$5,000 Cap	\$15,000 Cap
Δ Premium	-20.7	-34.9	-12.0	-8.4
Δ Out-of-pocket cost	-8.0	-30.5	-325.9	-136.0
Δ Spending	-28.7	-65.5	-337.9	-144.4
Δ Welfare ex. info	202.6	188.5	356.2	163.1
Δ Information cost	-161.9	-193.8	-24.4	-21.0
Δ Welfare ex. info (no taste shock)	442.2	829.9	358.6	161.9

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 and simulates information acquisition and plan choice. Out-of-pocket cap counterfactual imposes limit on out-of-pocket cost of all plans and then simulates information acquisition and plan choice.

taste shock. As seen in Figure 9 panel b, removing plans is initially welfare increasing but removing too many plans decreases welfare. Welfare is maximized when plans in the 26th percentile and below are removed from individuals' choice sets. This can be contrasted with results from the standard demand model, also shown in Figure 9, which imply that any restriction of the choice set is strictly welfare decreasing.

The counterfactual results examining restricted plan choice are summarized in Table 6. Eliminating plans in the lowest quartile results in individual's choosing plans that have lower cost and better non-cost characteristics, resulting in welfare gains of \$203. In addition, individuals face lower stakes and therefore choose to acquire less information, resulting in information acquisition costs that are \$162 lower. Removing plans in the bottom 50th percentile leads to even lower information acquisition costs given the simplified choice set. However, welfare gains excluding information acquisition costs are not as large as the case in which the lowest quartile is removed since once the choice set becomes too small it becomes hard for individuals to find an option that is a good fit.

These results can be contrasted with alternative models that do not account for endogenous information. In addition to comparing the results to a standard logit model assuming full information, we compare results to what would be implied by alternative models of insurance demand commonly used in the literature. When restricting the choice set in the same way as in the main counterfactual, these alternative models yield quite different predictions for spending and welfare. The results, presented in Figure A-12, show that both the coverage characteristics model and the differential weight model predict a change in spending that is smaller than what is predicted by the endogenous information model. In all of these models, restricting the choice set implies a welfare reduction, the opposite of what is implied by the endogenous information model.

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-4, show that welfare generally decreases as options are randomly removed. However, if the taste shock is excluded from welfare calculations, welfare can in fact increase slightly as plans are randomly removed, largely due to lower information acquisition cost. These results are shown in panel b of Figure A-4.

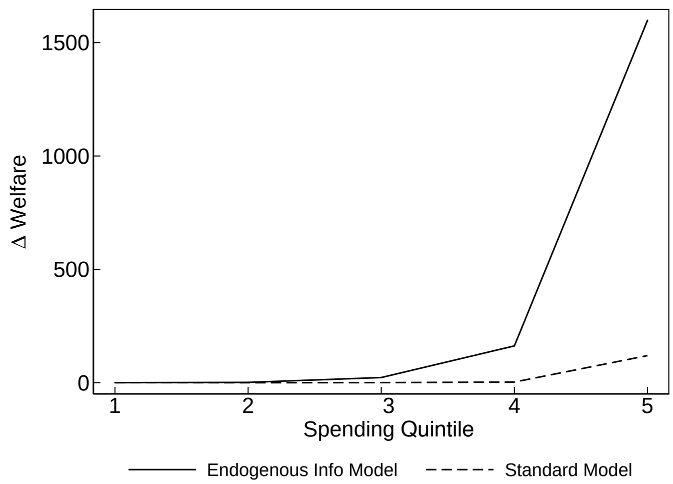
In order to examine how cost sharing interacts with endogenous information acquisition, we examine counterfactuals in which we impose an out-of-pocket cap. Currently, Medicare Part D enrollees who have out-of-pocket costs above the catastrophic threshold can still be liable for substantial costs.[39] Imposing an out-of-pocket cap effectively reduces the variance in out-of-pocket costs across plans, reducing the stakes as in the the previous counterfactual.

Figure 10 shows spending and change in information acquisition cost for different levels of an out-of-pocket cost cap. Unsurprisingly, the cap reduces out-of-pocket costs. However, the endogenous information model implies that, as the cap becomes more binding, out-of-pocket costs decrease faster than what would be implied by a standard model. There are two reasons why a cap on out-of-pocket costs generates additional welfare gains in the presence of endogenous information frictions. First, individuals are less likely to "accidentally" choose a plan with high out-of-pocket costs when the cap is binding. Second, Figure 10 panel b indicates that imposing the out-of-pocket cap also substantially reduces information acquisition costs. Since there is less risk of choosing a plan with very high out-of-pocket costs, individuals conduct less costly research.

The results, summarized in Table 6, indicate that imposing a \$15,000 cap implies a \$144 reduction in spending that generates a change in welfare of \$183 after accounting for the change in information cost. This can be compared to the standard demand model which implies a spending reduction of only \$102 and welfare effects of \$128. In order to understand

[39] 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.

Figure 11
Welfare Effect of Out-of-Pocket Cap by Spending Quintile



Notes: Chart shows welfare effect by spending quintile for counterfactual simulation with $5,000 out-of-pocket cap.

the mechanisms driving these results, we also simulate the endogenous information model holding individuals' prior fixed, therefore keeping information acquisition constant. The results, presented in Appendix Table A-11, indicate that the effect on spending is largely due to a reduction in mistakes.

Importantly, the endogenous information model implies that the welfare gains from an out-of-pocket cap accrue, in part, to individuals with spending below the cap. This is because these individuals spend less time and effort choosing a plan given that mistakes are less costly. This can be contrasted with the standard model which implies that only individuals with high spending benefit from the cap. This can be seen in Figure 11 which shows welfare effects by spending quintile for both the endogenous information model and standard model.[40]

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 cost of mistakes.

[40]In Figure A-12 we also present counterfactual results using the other alternative models. The coverage characteristics model does not imply any change in welfare since demand is not directly a function of out-of-pocket cost. The differential weight model estimates a much smaller welfare effect from the cap than the endogenous information model. This is due to the fact that the model estimates imply that individuals do not place much value on out-of-pocket costs.

7 Conclusion

We develop a micro-founded model of information frictions in insurance markets. Consistent with the model, we find evidence that individuals acquire more information as the stakes increase. The model can provide a unified framework that can also rationalize choice inconsistencies and choice overload. In addition, information acquisition costs may also help explain consumer inertia that has been documented in a variety of insurance markets.

We estimate an 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 concern about the complexity of insurance choice and how to regulate plan features. Standard demand models provide little insight into how information affects demand and how plans should be standardized. With this in mind, we use the model to examine how insurance regulation affects information acquisition. We find that accounting for endogenous information is important when considering policies that restrict plan choice or change cost sharing, both because of the change in information acquisition costs and the resulting effect on insurance choice.

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 information frictions. Future work should examine how endogenous information acquisition affects insurer competition over insurance premiums and out-of-pocket costs, as well as equilibrium responses on product positioning and plan complexity.

Furthermore, a contribution of the paper is to develop a tractable model of endogenous information frictions that can be used to empirically analyze other markets in which there are complex characteristics that are costly to research. Like insurance demand, other financial products, including mortgages and investment products, may require significant time and effort to understand. In addition, some attributes may be more difficult to observe than others. Related issues also arise in the context of sticker prices, which are easily observable, and shipping charges, taxes, or other surcharges that may be difficult to observe. Empirical methods incorporating endogenous information could also provide insight into how consumer protection laws should be designed in these markets by, for instance, regulating or standardizing product offerings.

References

Abaluck, Jason and Abi Adams, "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," Working Paper 2018.

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

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

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

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.

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," Working Paper 2016.

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.

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.

Cutler, David M, Amy Finkelstein, and Kathleen McGarry, "Preference heterogeneity and insurance markets: Explaining a puzzle of insurance," *American Economic Review*, 2008, *98* (2), 157–62.

Decarolis, Francesco, Maria Polyakova, and Stephen P Ryan, "The welfare effects of supply-side regulations in Medicare Part D," *NBER Working Paper*, 2015, *21298.*

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," Technical Report 2019.

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*, 2015, (0).

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.

Ho, Kate and Robin S Lee, "Insurer competition in health care markets," *Econometrica*, 2017, *85* (2), 379–417.

_ , 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.

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.

Iyengar, Sheena S and Emir Kamenica, "Choice proliferation, simplicity seeking, and asset allocation," *Journal of Public Economics*, 2010, *94* (7-8), 530–539.

Joo, Joonhwi, "Quantity-Surcharged Larger Package Sales as Rationally Inattentive Consumers' Choice," Working Paper 2019.

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.

Matějka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105* (1), 272–98.

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.

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:

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

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 $e_j \stackrel{iid}{\sim} EV1$ and C is a constant (Small and Rosen 1981). Applying this we have

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

where $le_j' \equiv -v_i/\lambda + e_j$ is the joint error and $l^2 \equiv (\frac{6\sigma^2}{\pi^2\lambda^2} + 1)$ 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 of the scaled EV1 distribution. Details about this distribution can be found in Cardell (1997).

Therefore,

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

Now the maximization problem in Equation A-1 becomes

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

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

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

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:

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

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

$$\begin{aligned}\sum_j P_j^0=\sum_j\left(l\eta e^{\frac{-p_j}{\lambda l}}\right)^{\frac{l}{1-l}}&=1\\ \eta=1/\sum_j\left(\frac{lp_j}{e^{\lambda l}}\right)&\end{aligned} \tag{A-5}$$

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

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

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

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

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

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

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

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

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_{ijt}^0, 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:

$$\max_{P_{i1t}^0,..,P_{iNt}^0} \int_{\boldsymbol{\xi}} \lambda_{it} \log \Sigma_{j=1}^N P_{ij}^0 \exp\left[(\alpha_i p_{jt} + \beta_3 X_{jt}^k + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it}\right] G(d\boldsymbol{\xi}) \text{ s.t. } \sum_{j=1}^N P_{ijt}^0 = 1,\ P_{ijt}^0 \geq 0\ \forall j \tag{A-9}$$

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

$$\begin{aligned}
&\int_{\boldsymbol{\xi}} \log \Sigma_j e^{\alpha_i p_{jt} + \beta_3 X^k_{jt} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it} + \log(P^0_{ijt})} G(d\boldsymbol{\xi}) \\
&\quad = \mathbb{E}_{\xi,e}\left[\max_j((\alpha_i p_{ijt} + \beta_3 X^k_{jt} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it} + \log(P^0_{ijt}) + e_{ijt})\right] + C \\
&\quad = \mathbb{E}_{\xi,e}\left[\max_j((\alpha_i p_{ijt} + \beta_3 X^k_{jt} + \epsilon_{ijt})/\lambda_{it} + \log(P^0_{ijt}) + \xi_{ijt}/\lambda_{it} + e_{ijt})\right] + C \\
&\quad = \mathbb{E}_{\xi',e}\left[\max_j((\alpha_i p_{ijt} + \beta_3 X^k_{jt} + \epsilon_{ijt})/\lambda_{it} + \log(P^0_{ijt}) + \xi^0_{ijt}/\lambda_{it} + \xi'_{ijt}/\lambda_{it} + e_{ijt})\right] + C \quad \text{(A-10)}
\end{aligned}$$

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

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

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

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

$$\begin{aligned}
Var[l_{it}e'_{ijt}] &= \frac{\sigma^2_{it}}{\lambda^2_{it}} + \frac{\pi^2}{6} \\
l^2_{it}\frac{\pi^2}{6} &= \frac{\sigma^2_{it}}{\lambda^2_{it}} + \frac{\pi^2}{6} \\
l^2_{it} &= \frac{6\sigma^2_{it}}{\pi^2\lambda^2_{it}} + 1
\end{aligned}$$

Then, equation A-10 can be rewritten as

$$\begin{aligned}
&\int_{\boldsymbol{\xi}} \log \Sigma_j e^{\alpha_i p_{jt} + \beta_3 X^k_{jt} + \epsilon_{ijt} + \xi_{ijt})/\lambda_{it} + \log(P^0_{ijt})} G(d\boldsymbol{\xi}) \\
&\qquad = \mathbb{E}_{e'}\left[\max_j((\alpha_i p_{ijt} + \xi^0_{ijt} + \beta_3 X^k_{jt} + \epsilon_{ijt})/\lambda_{it} + \log(P^0_{ijt}) + l_{it}e'_{ijt})\right] + C
\end{aligned}$$

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 of the scaled EV1 distribution.

Note that

$$\mathbb{E}_{e'}\left[\max_j((\alpha_i p_{ijt}+\xi^0_{ijt}+\beta_3 X^k_{jt}+\epsilon_{ijt})/\lambda_{it}+\log(P^0_{ijt})+l_{it}e'_{ijt})\right]+C$$
$$=\log\Sigma^N_{j=1}\exp[(\alpha_i p_{ijt}+\xi^0_{ijt}+\beta_3 X^k_{jt}+\epsilon_{ijt})/l_{it}\lambda_{it}+\log(P^0_{ijt})/l_{it}]$$

Now the maximization problem in Equation A-9 can be rewritten as

$$\max_{P^0_{i1t},..,P^0_{iNt}} \Sigma^N_{j=1}\exp[(\alpha_i p_{ijt}+\xi^0_{ijt}+\beta_3 X^k_{jt}+\epsilon_{ijt})/l_{it}\lambda_{it}+\log(P^0_{ijt})/l_{it}] \text{ s.t. } \sum_{j=1}^{N} P^0_{ijt}=1,\, P^0_{ijt}\geq 0\ \forall j$$

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

$$P^0_{ijt}=\frac{\exp\left[(\alpha_i p_{jt}+\xi^0_{ijt}+\beta_3 X^k_{jt}+\epsilon_{ijt})/(\lambda_{it}l_{it}-\lambda_{it})\right]}{\sum^N_{k=1}\exp\left[(\alpha_i(p_{kt}+v^0_{ikt})+\beta_3 X^k_{kt}+\epsilon_{ikt})/(\lambda_{it}l_{it}-\lambda_{it})\right]}.$$

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

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

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

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

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

$$\tilde{u}_{ijt}=(\alpha_i v_{ijt}+\beta_1 X^u_{jt}+\beta_2\widetilde{\sigma^2_{ijt}}+\alpha_i p_{jt}+\beta_3 X^k_{jt}+\epsilon_{ijt})/\lambda_{it}+(\alpha_i p_{jt}+\xi^0_{ijt}+\beta_3 X^k_{jt}+\epsilon_{ijt})/(\lambda_{it}l_{it}-\lambda_{it})+e_{ijt} \tag{A-11}$$

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

$$\begin{aligned}
\tilde{u}_{ijt} &= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt}}{\lambda_{it}} + \frac{\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt}}{\lambda_{it} l_{it} - \lambda_{it}} + \frac{\epsilon_{ijt}}{\lambda_{it} l_{it} - \lambda_{it}} + \frac{\epsilon_{ijt}}{\lambda_{it}} + e_{ijt} \\
&= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \alpha_i p_{jt} + \beta_3 X^k_{jt}}{\lambda_{it}} + \frac{\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt}}{\lambda_{it}(l_{it} - 1)} + \frac{l_{it}}{\lambda_{it}(l_{it} - 1)} \epsilon_{ijt} + e_{ijt} \\
&= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{\lambda_{it}} + \frac{(l_{it} - 1)\left(\alpha_i p_{jt} + \beta_3 X^k_{jt}\right)}{\lambda_{it}(l_{it} - 1)} + \frac{\alpha_i p_{jt} + \xi^0_{ijt} + \beta_3 X^k_{jt}}{\lambda_{it}(l_{it} - 1)} + \frac{l_{it}}{\lambda_{it}(l_{it} - 1)} \epsilon_{ijt} + e_{ijt} \\
&= \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{\lambda_{it}} + \frac{\alpha_i l_{it} p_{jt} + \xi^0_{ijt} + \beta_3 l_{it} X^k_{jt}}{\lambda_{it}(l_{it} - 1)} + \frac{l_{it}}{\lambda_{it}(l_{it} - 1)} \epsilon_{ijt} + e_{ijt} \qquad \text{(A-12)}
\end{aligned}$$

We define the joint error as $k_{it} e'_{ijt} \equiv \frac{l_{it}}{\lambda_{it}(l_{it}-1)} \epsilon_{ijt} + e_{ijt}$ 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,

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

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

$$\tilde{u}_{ijt} = \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{\lambda_{it}} + \frac{\alpha_i l_{it} p_{jt} + \xi^0_{ijt} + \beta_3 l_{it} X^k_{jt}}{\lambda_{it}(l_{it} - 1)} + k_{it} e'_{ijt}.$$

We renormalized the error and obtain

$$\frac{\tilde{u}_{ijt}}{k_{it}} = \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it} \lambda_{it}} + \frac{\alpha_i l_{it} p_{jt} + \xi^0_{ijt} + \beta_3 l_{it} X^k_{jt}}{k_{it} \lambda_{it}(l_{it} - 1)} + e'_{ijt}.$$

Therefore, the choice probabilities are

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

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

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

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

$$\begin{aligned} e^v &= \frac{\partial P_{ij}}{\partial v_{ij}} \frac{p_j + v_{ij}}{P_{ij}} \\ &= \frac{\partial V_{ij}}{\partial v_{ij}} P_{ij}(1 - P_{ij}) \frac{p_j + v_{ij}}{P_{ij}} \\ &= \alpha_i \frac{1}{k_{it}\lambda_i}(1 - P_{ij})(p_j + v_{ij}) \end{aligned} \tag{A-14}$$

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 $V_{ijt}(\alpha_i, \lambda_{it}, \beta_1, \beta_3) = \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i l_{it} p_{jt} + \xi^0_{ijt} + \beta_3 l_{it} X^k_{jt}}{k_{it}\lambda_{it}(l_{it} - 1)}$. 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 $\tilde{\nu}_{ijt} = \frac{\alpha_i v_{ijt} + \beta_1 X^u_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}}}{k_{it}\lambda_{it}} + \frac{\alpha_i l_{it} p_{jt} + \xi^0_{ijt} + \beta_3 l_{it} X^k_{jt}}{k_{it}\lambda_{it}(l_{it}-1)}$ 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, we define 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). 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 Robustness and Alternative Specifications of Demand with Endogenous Information

Alternatively, individuals may be able to initially observe some intermediate plan characteristics such as the plan deductible. These characteristics may partially inform them about expected out-of-pocket costs prior to doing research. We consider a third specification in which individuals are assumed to know the correlation between easily observable characteristics and the expected out-of-pocket costs. In particular, we let $\xi^0_{ijt} = \beta^{\xi} K_{jt}$ where K_{jt} includes premium, plan deductible, average cost sharing, generic coverage, coverage in the donut hole, and year fixed effects. We estimate β^{ξ} by regressing out-of-pocket costs on K_{jt} and then assume β^{ξ} is known by individuals.

Finally, we consider alternative measures of the variance of individual's prior $G\sigma^2$. 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.

In the baseline model, we directly include plan quality in the model, mitigating concerns about endogenous prices. Nevertheless, we also consider a specification with plan fixed effects which are assumed to be known by the individual.

The results for these alternative specifications can be found in Table A-9 and Table A-10.

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 the distribution of out-of-pocket cost.[41] 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

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

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, and X_{jt} is plan quality. 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. A related approach uses plan fixed effects to absorb differences in deductible, coinsurance, or other coverage characteristics. See, for instance,Ho and Lee (2017). In particular, we assume utility takes the form

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

$$u_{ijt} = \alpha_i p_{jt} + \beta_1 C_{jt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \beta_3 X_{jt} + \epsilon_{ijt} \tag{A-17}$$

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, and ϵ_{ijt} are the same as the previous 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), Abaluck and Gruber (2016), and Ho et al. (2017), assumes that the coefficients are fixed when considering counterfactual policies. For this model, we assume utility is given by

$$u_{ijt} = \alpha_i p_{jt} + \beta_1 v_{ijt} + \beta_2 \widetilde{\sigma^2_{ijt}} + \beta_3 X_{jt} + \epsilon_{ijt}. \tag{A-18}$$

We maintain assumptions regarding $\widetilde{\sigma^2_{ijt}}$, X_{jt}, α_i, 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-8.

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
Out-of-pocket cost (RE)	0.58	0.000	0.47	0.000
Out-of-pocket cost (PF)	0.73	0.000	0.69	0.000
Annual Premium	-0.00	0.646	0.00	0.957
Age	0.02	0.000	-0.00	0.911
Zip Income	0.04	0.000	0.03	0.006
Zip Education (%BA)	0.03	0.000	0.02	0.020
Chronic Condition	0.12	0.000	0.08	0.000

Notes: Shows correlation coefficient between relevant variable and stakes.

Table A-2
Summary of Demographics for Forced Switchers and All Enrollees

	All		Active Choice	
	Mean	SD	Mean	SD
Age	76.8	7.3	76.3	7.2
Female	0.606	0.489	0.603	0.489
Zip income (1,000s)	76.1	34.5	76.8	34.7
Zip education (pct BA)	29.2	16.9	29.6	17.1
Rural	0.072	0.259	0.070	0.255
Years enrolled in Part D	5.67	2.20	5.13	2.10
Alzheimers	0.093	0.290	0.087	0.282
Lung disease	0.109	0.312	0.104	0.305
Kidney disease	0.171	0.377	0.153	0.360
Heart failure	0.145	0.352	0.139	0.346
Depression	0.120	0.325	0.114	0.318
Diabetes	0.281	0.449	0.272	0.445
Other chronic condition	0.323	0.468	0.302	0.459

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

	(1)	(2)	(3)	(4)	(5)
Stakes (100s)	-2.268^{***}	-1.766^{***}	-0.143	-0.315^{***}	-0.210^{**}
	(0.061)	(0.062)	(0.090)	(0.090)	(0.091)
Stakes Squared	0.198^{***}	0.159^{***}	0.032^{***}	0.046^{***}	0.036^{***}
	(0.006)	(0.005)	(0.008)	(0.008)	(0.008)
Plan Characteristic Controls	No	Yes	No	Yes	Yes
Individual FEs	No	No	Yes	Yes	Yes
Year FEs	No	No	No	No	Yes
Implied minimum	571.4	554.1	223.0	339.1	288.9
Adjusted R2	0.007	0.028	0.347	0.351	0.353
Observations	200,701	200,701	200,701	200,701	200,701

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
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 $\times$ Indiv. avg stakes				0.023^{***}	0.021^{***}	0.021^{***}
				(0.001)	(0.002)	(0.002)
Premium $\times$ Stakes		0.023^{***}	0.022^{***}	0.009^{***}	0.010^{***}	
		(0.001)	(0.001)	(0.001)	(0.001)	
Premium $\times$ Stakes $\times$ $\mathbb{1}$(Stakes > 0)						0.008^{***}
						(0.002)
Premium $\times$ Stakes $\times$ $\mathbb{1}$(Stakes < 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 $\times$ Indiv. avg stakes				-0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)
OOP $\times$ Stakes		0.001	0.001	-0.001	-0.001	
		(0.001)	(0.001)	(0.001)	(0.001)	
OOP $\times$ Stakes $\times$ $\mathbb{1}$(Stakes > 0)						-0.000
						(0.001)
OOP $\times$ Stakes $\times$ $\mathbb{1}$(Stakes < 0)						-0.001
						(0.001)
Premium $\times$ Z_i	No	No	Yes	No	Yes	Yes
OOP $\times$ 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-5
Interaction of Stakes and Price Coefficient in Standard Logit Model
Robustness Check with Perfect Foresight Assumption

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

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

Table A-6
Interaction of Stakes and Price Coefficient in Standard Logit Model
Robustness Check with Random Coefficients and Perfect Foresight Assumption

	(1)	(2)	(3)	(4)	(5)	(6)
Premium (100s)	-0.310^{***}	-0.349^{***}	-0.625^{***}	-0.377^{***}	-0.641^{***}	-0.640^{***}
	(0.004)	(0.004)	(0.029)	(0.004)	(0.029)	(0.029)
sd	0.210^{***}	0.204^{***}	0.204^{***}	0.204^{***}	0.204^{***}	0.203^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Premium $\times$ Indiv. avg stakes				0.023^{***}	0.021^{***}	0.021^{***}
				(0.001)	(0.002)	(0.002)
Premium $\times$ Stakes		0.023^{***}	0.021^{***}	0.009^{***}	0.009^{***}	
		(0.001)	(0.001)	(0.001)	(0.001)	
Premium $\times$ Stakes $\times$ $\mathbb{1}$(Stakes $>$ 0)						0.008^{***}
						(0.002)
Premium $\times$ Stakes $\times$ $\mathbb{1}$(Stakes $<$ 0)						0.011^{***}
						(0.002)
Out-of-Pocket Cost (100s)	0.017^{***}	0.007	0.061^{**}	0.004	0.055^{*}	0.056^{*}
	(0.003)	(0.004)	(0.029)	(0.006)	(0.029)	(0.029)
sd	-0.086^{***}	-0.087^{***}	0.083^{***}	0.084^{***}	-0.083^{***}	-0.085^{***}
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
OOP $\times$ Indiv. avg stakes				0.004^{***}	0.004^{***}	0.004^{***}
				(0.001)	(0.001)	(0.001)
OOP $\times$ Stakes		0.004^{***}	0.004^{***}	0.001^{**}	0.001^{**}	
		(0.001)	(0.001)	(0.001)	(0.001)	
OOP $\times$ Stakes $\times$ $\mathbb{1}$(Stakes $>$ 0)						0.002^{***}
						(0.001)
OOP $\times$ Stakes $\times$ $\mathbb{1}$(Stakes $<$ 0)						0.001
						(0.001)
Premium $\times$ X_i	No	No	Yes	No	Yes	Yes
OOP $\times$ X_i	No	No	Yes	No	Yes	Yes
Log Likelihood	-112,179	-111,938	-111,509	-111,821	-111,409	-111,405
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 Regression Predicting
Out-of-Pocket Cost from Easily Observable
Characteristics

	Estimate	SE
Premium	-0.0188	(0.0037)
Drug deductible	0.0062	(0.0001)
Generic coverage	-2.8126	(0.0337)
Coverage in the coverage gap	0.9113	(0.0329)
Average cost sharing	5.7168	(0.1076)
Observations	1,058,745	

Notes: Includes year fixed effects. Standard errors in parentheses.

Table A-8
Estimates for Alternative Models of Insurance Demand
without Endogenous Information

	Standard Logit		Coverage Characteristics		Differential Weight	
Total cost	-0.504^{***}	(0.132)				
Total cost × Income	0.000	(0.000)				
Total cost × Education	0.000	(0.000)				
Total cost × Age	0.010^{***}	(0.003)				
Total cost × Age-squared	-0.000^{***}	(0.000)				
Total cost × Female	-0.003	(0.002)				
Total cost × Rural	-0.002	(0.004)				
Premium			-2.499^{***}	(0.332)	-1.419^{***}	(0.280)
Premium × Income			-0.000	(0.000)	-0.000	(0.000)
Premium × Education			0.000	(0.000)	0.000	(0.000)
Premium × Age			0.050^{***}	(0.008)	0.030^{***}	(0.007)
Premium × Age-squared			-0.000^{***}	(0.000)	-0.000^{***}	(0.000)
Premium × Female			0.002	(0.006)	0.002	(0.005)
Premium × Rural			0.002	(0.010)	-0.000	(0.009)
Deductible			-0.009^{***}	(0.000)		
Generic coverage			-0.473^{***}	(0.039)		
Coverage in gap			0.723^{***}	(0.040)		
Cost sharing			2.064^{***}	(0.093)		
OOP					-0.033^{***}	(0.002)
Other controls for plan characteristic	Yes		Yes		Yes	
Log Likelihood	-39,504		-31,990		-38,480	

Notes: 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-9
Estimates for Demand Model with Endogenous Information Acquisition
Alternative Specifications

	Specification 3 Plan Fixed Effects		Specification 4 Heterogenous Prior Predicted from Observables	
	Estimate	SE	Estimate	SE
Price Sensitivity (β^{α})				
Constant	−1.7991	(0.0214)	−2.1112	(0.0875)
Income	−0.0004	(0.0007)	0.0017	(0.0026)
Education	−0.0021	(0.0012)	−0.0044	(0.0056)
Age	−0.1915	(0.0273)	0.0238	(0.1395)
Age^2	0.0011	(0.0002)	−0.0003	(0.0009)
Female	−0.0252	(0.0234)	0.1085	(0.0969)
Rural	0.0750	(0.0441)	0.4103	(0.1432)
Other Plan Characteristics				
Plan Quality	1.4036	(0.0543)	3.3867	(0.8149)
Risk	−0.0432	(0.0018)	−0.0481	(0.0126)
Marginal cost of information (β^{λ})				
Constant	2.9379	(0.1194)	3.4565	(0.3415)
Income	0.0005	(0.0008)	0.0010	(0.0016)
Education	−0.0048	(0.0018)	−0.0067	(0.0037)
Age	0.2074	(0.0500)	0.4138	(0.1064)
Age^2	−0.0012	(0.0003)	−0.0024	(0.0006)
Female	0.0330	(0.0351)	−0.0075	(0.0764)
Rural	0.3632	(0.0825)	0.1288	(0.1375)
Part D Experience	−0.4030	(0.0189)	−0.3324	(0.0294)
Has alzheimers	0.0071	(0.0591)	−0.1358	(0.1357)
Has lung disease	0.0847	(0.0575)	0.0796	(0.1180)
Has kidney disease	0.0304	(0.0461)	−0.2010	(0.0940)
Has heart failure	0.1033	(0.0497)	0.1515	(0.1162)
Has depression	0.0372	(0.0521)	0.0443	(0.1118)
Has diabetes	0.1162	(0.0398)	0.0145	(0.0848)
Has other chronic condition	0.0141	(0.0391)	−0.0343	(0.0835)
Mean price sensitivity	−0.1656		−0.1348	
Mean marginal cost of information	2.8154		5.5449	
LL	107,842.09		118,556.31	
Observations	1,058,745		1,058,745	

Notes: 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
Estimates for Demand Model with Endogenous Information Acquisition
Robustness to Alternative Definition of Prior Variance Based on Group Average

	Specification 1 Homogenous Prior		Specification 2 Heterogenous Prior Plan Average OOP	
	Estimate	SE	Estimate	SE
Price Sensitivity (β^{α})				
Constant	−2.0477	(0.0260)	−2.0392	(0.0282)
Income	−0.0005	(0.0009)	0.0008	(0.0008)
Education	−0.0015	(0.0016)	−0.0032	(0.0017)
Age	−0.1518	(0.0354)	−0.1592	(0.0373)
Age^2	0.0009	(0.0002)	0.0009	(0.0002)
Female	−0.0251	(0.0302)	−0.0122	(0.0324)
Rural	0.1201	(0.0533)	0.0450	(0.0596)
Other Plan Characteristics				
Plan Quality	2.6502	(0.1664)	2.7878	(0.1642)
Risk	−0.0561	(0.0037)	−0.0573	(0.0036)
Marginal cost of information (β^{λ})				
Constant	3.2406	(0.0809)	3.2283	(0.0827)
Income	0.0009	(0.0005)	0.0009	(0.0005)
Education	−0.0068	(0.0012)	−0.0066	(0.0012)
Age	0.2733	(0.0354)	0.2546	(0.0337)
Age^2	−0.0016	(0.0002)	−0.0015	(0.0002)
Female	0.0305	(0.0235)	0.0293	(0.0231)
Rural	0.0528	(0.0467)	0.0345	(0.0458)
Part D Experience	−0.3680	(0.0118)	−0.3554	(0.0105)
Has alzheimers	0.0172	(0.0406)	0.0238	(0.0401)
Has lung disease	0.0438	(0.0376)	0.0577	(0.0375)
Has kidney disease	−0.0282	(0.0304)	−0.0220	(0.0301)
Has heart failure	0.0789	(0.0337)	0.0783	(0.0333)
Has depression	−0.0210	(0.0347)	−0.0168	(0.0343)
Has diabetes	0.0409	(0.0261)	0.0539	(0.0258)
Has other chronic condition	−0.0251	(0.0257)	−0.0193	(0.0253)
Mean price sensitivity	−0.1291		−0.1304	
Mean marginal cost of information	4.0605		4.2794	
LL	123,016.43		123,470.78	
Observations	1,058,745		1,058,745	

Notes: Premium and out-of-pocket cost are in hundreds of dollars. Continuous individual characteristics (income, education, age, and age squared) are demeaned. Prior variance defined as average choice set variance for similar individuals. Standard errors in parentheses.

Table A-11
Counterfactual Spending and Welfare for Out-of-Pocket Cap with Fixed Prior

	$5,000 Cap	$15,000 Cap
Endogenous Info Model		
Δ Premium	-11.0	-7.7
Δ Out-of-pocket cost	-325.6	-135.3
Δ Spending	-336.5	-142.9
Δ Welfare ex. info	0.0	0.0
Δ Information cost	-24.4	-20.8

Notes: Out-of-pocket cap counterfactual imposes limit on out-of-pocket cost of all plans and then simulates information acquisition and plan choice holding individuals prior fixed at the baseline.

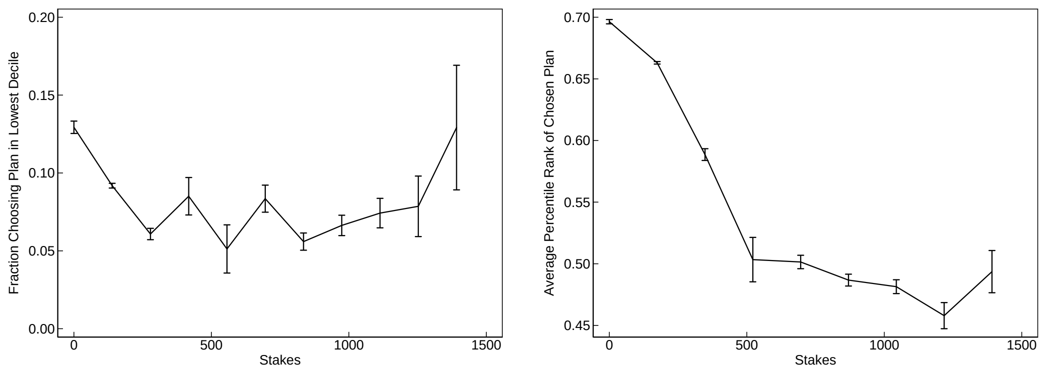
Table A-12
Counterfactual Spending and Welfare for Restricted Choice Set and Out-of-Pocket Cap from Alternative Demand Models

	Restricted Choice Set		Out-of-Pocket Cap	
	25th Percentile Cutoff	50th Percentile Cutoff	$5,000 Cap	$15,000 Cap
Standard logit model				
Δ Premium	-20.9	-35.9	-25.8	-16.6
Δ Out-of-pocket cost	-12.3	-32.4	-265.9	-85.3
Δ Spending	-33.2	-68.4	-291.7	-101.9
Δ Welfare	-271.6	-753.3	313.7	127.8
Coverage characteristics model				
Δ Premium	-8.7	-16.4	-	-
Δ Out-of-pocket cost	-6.5	-16.0	-358.4	-172.8
Δ Spending	-15.2	-32.4	-358.4	-172.8
Δ Welfare	-65.8	-191.5	-	-
Differential weight model				
Δ Premium	-16.9	-31.2	-14.5	-9.4
Δ Out-of-pocket cost	-6.9	-17.4	-316.3	-126.2
Δ Spending	-23.8	-48.5	-330.7	-135.6
Δ Welfare	-142.5	-407.4	145.1	64.2

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
Alternative Measures of Probability of Choosing Low Cost Plan by Stakes

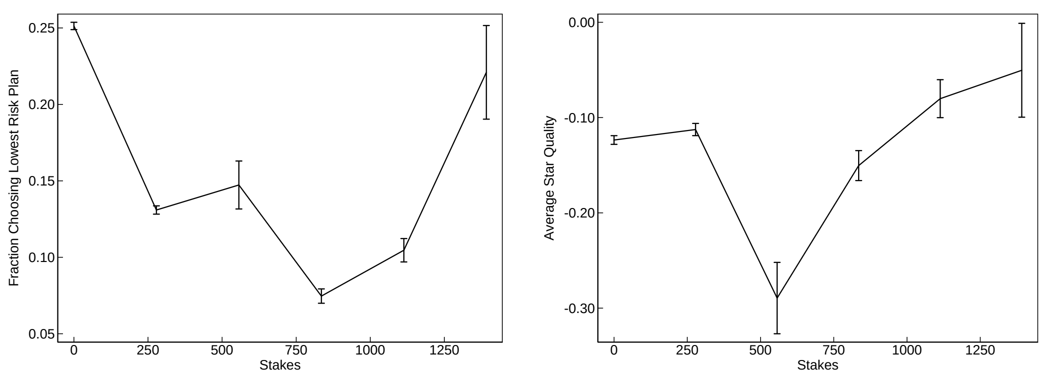


a. Plan in Lowest Decile

b. 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-2
Alternative Measures of Choice Quality by Stakes

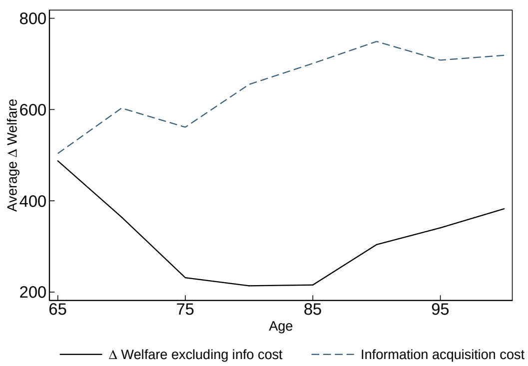


a. Fraction Choosing Lowest Risk Plan

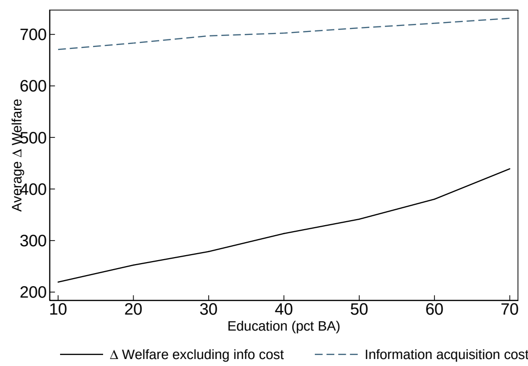
b. Average Plan Quality

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

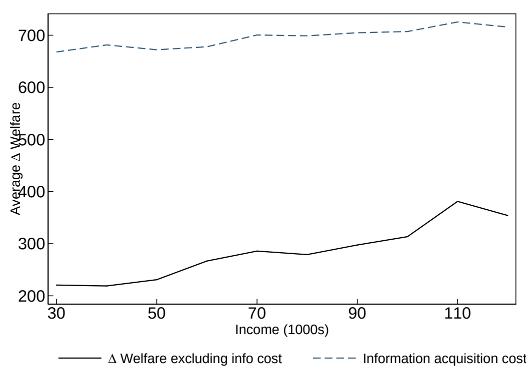
Figure A-3
Counterfactual Welfare Effects of Full Information
By Demographics



a. Age



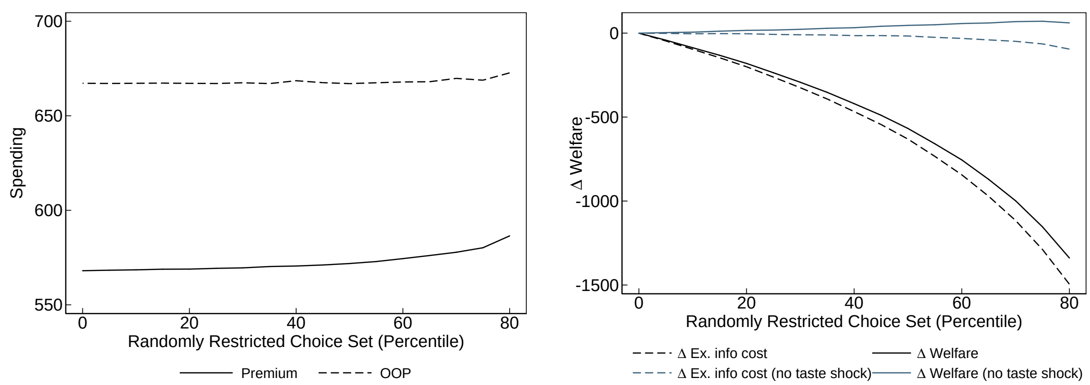
b. Education



c. Income

Notes: Shows difference in welfare between baseline simulation using Specification 1 and counterfactual simulation with full information for each demographic bin. Income and education are at the zip code level.

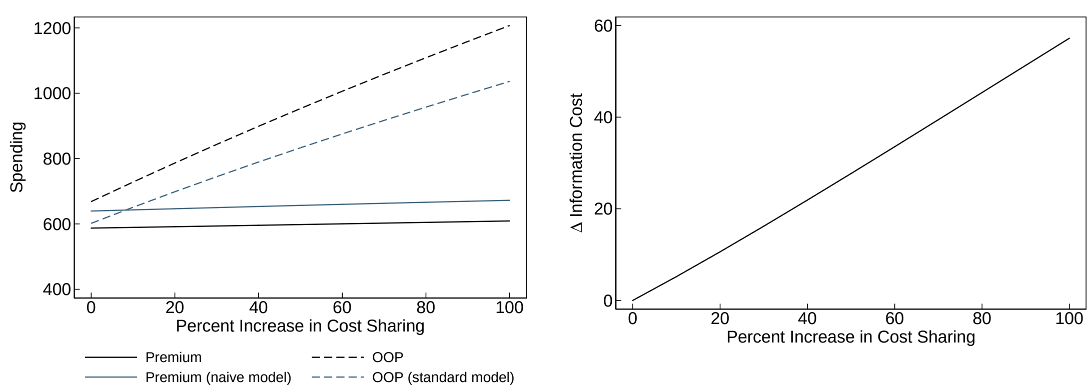
Figure A-4
Counterfactual Analysis of Randomly Removing Options from Choice Set



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-5
Counterfactual Analysis of Increased Cost Sharing



a. Spending b. Change in Information Acquisition Cost

Notes: Chart shows counterfactual spending and change in information acquisition cost from increasing cost sharing. Counterfactual estimates from model with endogenous information acquisition is contrasted with counterfactual estimates from standard logit demand model.

G Monte Carlo Analysis to Assess Sensitivity to Distributional Assumptions

We conduct a Monte Carlo exercise as part of our robustness analysis. In particular, we examine whether estimates are sensitive to the distributional assumption on the prior of out-of-pocket costs that is used in deriving the closed-form expression of choice probabilities (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-13 lists parameter values chosen for the simulation.

Table A-13
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

Notes:

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.

We simulate 1000 and 5000 choice situations under the two sets of assumptions and repeat each simulation 50 times. Table A-14 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

Table A-14
Monte Carlo Results

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

Notes: Standard errors are in parentheses.

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.