Online Appendix for "Endogenous Information and Simplifying Insurance Choice"

Zach Y. Brown
University of Michigan & NBER

Jihye Jeon
Boston University

October 2023

A Model Details

A-1 Model Derivation

Utility is given by

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

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

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

$$\max_{P_{i1}^0,..,P_{iJ}^0} \int_{\boldsymbol{\xi_i}} \lambda \log \Sigma_{j=1}^J P_{ij}^0 \exp\left[(\xi_{ij} + \alpha p_j + \theta X_j^k + \epsilon_{ij})/\lambda\right] G(d\boldsymbol{\xi_i})$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{J}} P_{ij}^0 = 1,\, P_{ij}^0 \geq 0\ \forall j \tag{A-2}$$

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

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

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

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

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

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

Note that the joint error is $\xi'_{ij}/\lambda + e_j$. Given that e_j is distributed EV1, $Var[e_j] =$

$\frac{\pi^2}{6}$ so

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

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

$$Var_j[\ell(\sigma_i,\lambda)e^{'}_{ij}] = \frac{\sigma_i^2}{\lambda^2} + \frac{\pi^2}{6}$$
$$\ell(\sigma_i,\lambda)^2 = \frac{6\sigma_i^2}{\pi^2\lambda^2} + 1$$

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

$$\lambda\mathbb{E}_{e'}\left[\max_j((\alpha p_{ij} + \xi^0_{ij} + \theta X^k_j + \epsilon_{ij})/\lambda + \log(P^0_{ij}) + \xi^0_{ij}/\lambda + \ell(\sigma_i,\lambda)e^{'}_{ij})\right] - \lambda\gamma^e. \quad \text{(A-7)}$$

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

$$\lambda\mathbb{E}_{e'}\left[\max_j((\alpha p_{ij} + \xi^0_{ij} + \theta X^k_j + \epsilon_{ij})/\lambda + \log(P^0_{ij}) + \xi^0_{ij}/\lambda + \ell(\sigma_i,\lambda)e^{''}_{ij})\right] - \lambda\ell(\sigma_i,\lambda)\gamma^e. \quad \text{(A-8)}$$

Now we can again apply the formula from Small and Rosen (1981), this time in reverse. In particular, note that $\mathbb{E}_e\left[\max_j(v_j + \ell e_j)\right] = \ell\log\sum_j e^{v_j/\ell} + \ell\gamma^e$ where e_j is EV1. This implies that equation (A-8) can be expressed as

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

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

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

In the maximization problem we have ignored terms that do not affect the solution.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
e^p &= \frac{\partial P_{ij}}{\partial p_j}\frac{p_j + v_{ij}}{P_{ij}} \\
&= \frac{\partial V_{ij}}{\partial p_j}P_{ij}(1-P_{ij})\frac{p_j+v_{ij}}{P_{ij}} \\
&= \alpha_i\frac{\ell_{it}}{k_{it}\lambda(\ell_{it}-1)}(1-P_{ij})(p_j+v_{ij}),
\end{aligned}
\tag{A-12}
$$

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

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

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

A-2 Basic Model for Simulation

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

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

where

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

It is as if the agent maximizes expected utility

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

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

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

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

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

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

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

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

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

A-3 Empirical Model Likelihood function

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

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

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

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

where

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

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

A-4 Derivation of Welfare

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

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

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

$$CS^{RI} = \frac{1}{-\alpha_i}\mathbb{E}[\tilde{u}_{ij^*t} + d_{ij^*t}]$$

$$= \frac{1}{-\alpha_i}\mathbb{E}[\max_j \tilde{u}_{ijt}] + \frac{1}{-\alpha_i}\sum_j P_{ijt}d_{ijt}$$
$$= \frac{1}{-\alpha_i}\log\sum_j \exp[\tilde{\nu}_{ijt}] + \frac{1}{-\alpha_i}\sum_j P_{ijt}[\nu_{ijt} - \tilde{\nu}_{ijt}]$$

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

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

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

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

B Details on Data Construction

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

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

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

C Details of Alternative Models without Endogenous Information

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

Standard logit model

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

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

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

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

Coverage characteristics model

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

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

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

Differential weight model

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

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

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

Results from Alternative Models

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

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

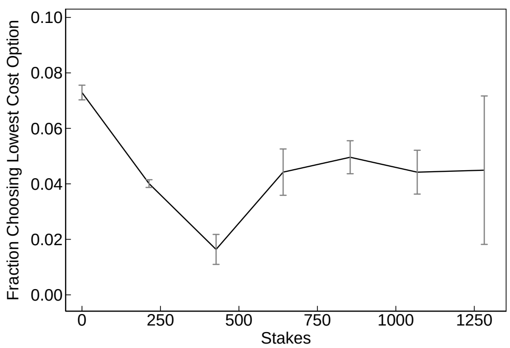
Table A-2 shows results from counterfactual experiments under the alternative models.

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

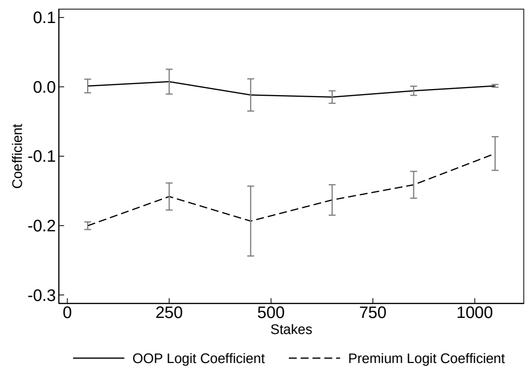
	Standard Logit		Coverage Characteristics		Differential Weight	
Total cost	-0.0773^{***}	(0.0034)				
Total cost × Income	0.0003^{***}	(0.0000)				
Risk	0.0027^{**}	(0.0011)	-0.0009	(0.0011)	0.0007	(0.0011)
Premium			-0.1705^{***}	(0.0052)	-0.1130^{***}	(0.0042)
Premium × Income			0.0004^{***}	(0.0000)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

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

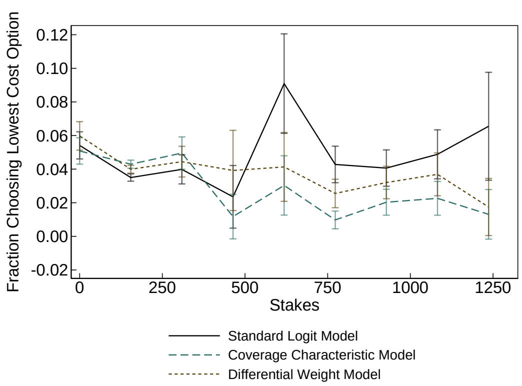
Figure A-1
Fit of Endogenous Information Model and Alternative Models



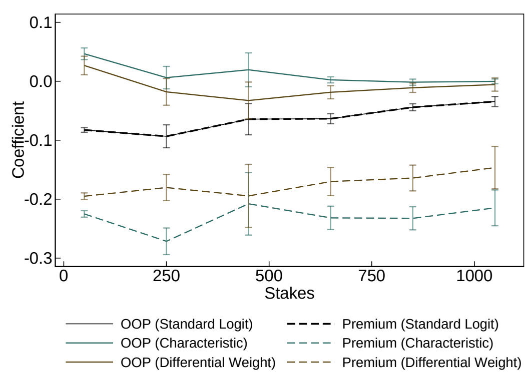
a. Fraction Choosing Lowest Cost Plan
Endogenous Information Model



b. Logit Coefficient on Premium & OOP
Endogenous Information Model



c. Fraction Choosing Lowest Cost Plan
Alternative Models



d. Logit Coefficient on Premium & OOP
Alternative Models

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

Table A-2

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

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

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

D Robustness Results for Motivating Evidence

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

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

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

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

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

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

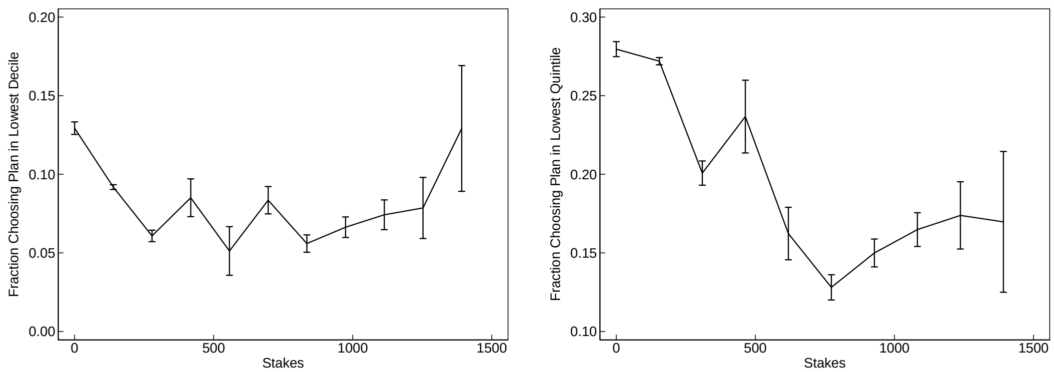
Notes: Estimates from linear probability model where dependent variable is the indicator variable for whether the individual chooses the lowest cost plan. Standard errors clustered at the market level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

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

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

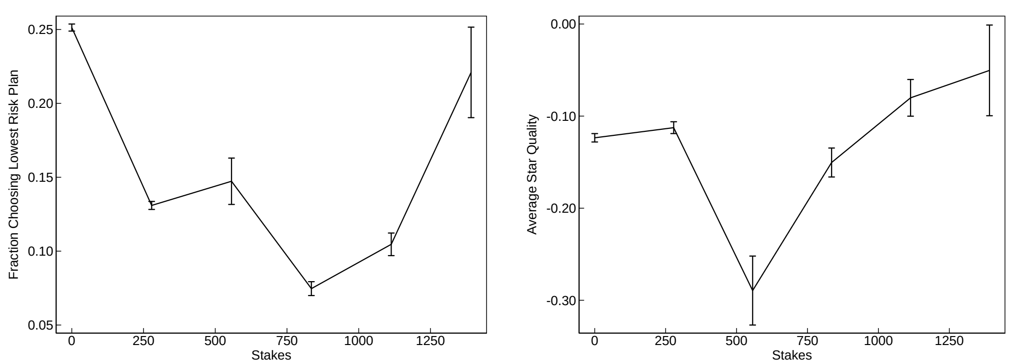
Figure A-2
Alternative Measures of Probability of Choosing Low Cost Plan by Stakes



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

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

Figure A-3
Alternative Measures of Choice Quality by Stakes



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

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

E Results from Alternative Specifications and Additional Counterfactuals

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

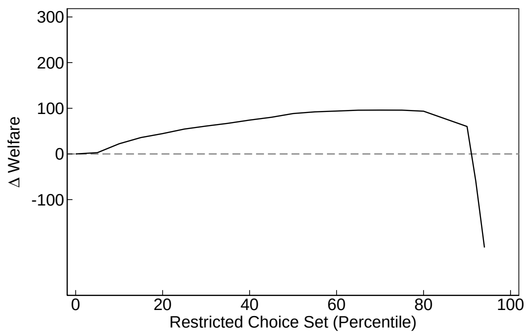
Table A-6
Estimates for Demand Model with Endogenous Information Acquisition
Alternative Specifications

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

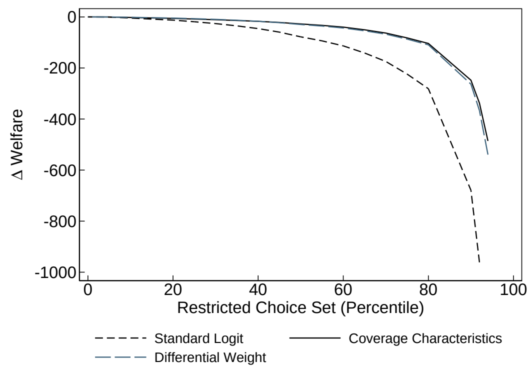
Notes: Specification 1 does not include insurer fixed effects. Specification 2 includes individuals with outlier stakes, which are not included in the baseline specification. Premium and out-of-pocket cost are in hundreds of dollars. Continuous individual characteristics (income, education, age, and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

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

Figure A-4
Counterfactual Welfare Effects of Restricted Choice Set



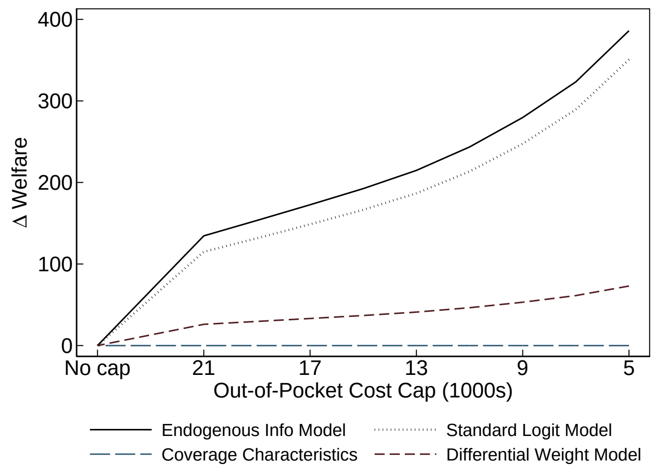
a. Endogenous Information Model



b. Alternative Models

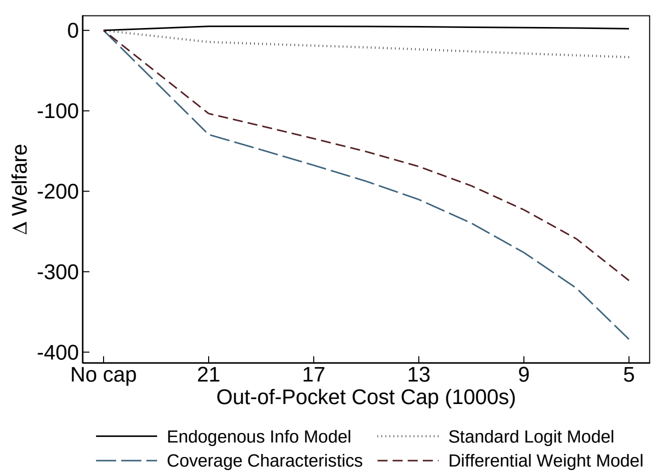
Notes: Chart shows counterfactual average change in welfare per enrollee from removing plans with mean utility below a given percentile where average utlity is computed for each individual. Counterfactual estimates from model with endogenous information acquisition are contrasted with counterfactual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

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



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

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



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

F Identification

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

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

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

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

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

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

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

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

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

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

using

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

G Monte Carlo Analysis to Assess Sensitivity to Distributional Assumptions

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

Table A-7
Parameter Values for a Monte Carlo Simulation

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

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

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

Table A-8
Monte Carlo Results

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

Notes: Standard errors are in parentheses.

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

Appendix References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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