Appendix:
Information Frictions and Adverse Selection:
Policy Interventions in Health Insurance Markets

Benjamin R. Handel
UC Berkeley and NBER

Jonathan T. Kolstad
UC Berkeley and NBER

Johannes Spinnewijn
London School of Economics and CEPR

January 24, 2018

Abstract

This appendix provides supporting analysis for the primary manuscript 'Adverse Selection and Information Frictions: Policy Interventions in Health Insurance Markets. Appendix A presents proofs for the propositions and corollaries contained in the main text. Appendix B describes the cost model estimation in greater detail. Appendix C discusses choice model identification and estimation in greater detail. Appendix D discusses a series of analytical simulations designed to draw out results from our theoretical framework. Appendix E presents our counterfactual market analysis for the case where two types of competing health plans are priced in the market, rather than the case discussed in the paper where there is competition for offering one type of supplemental coverage. Appendix F presents additional analysis to supplement the analysis contained in the main paper.

A Appendix: Proofs

Proof of Proposition 1:

We consider a policy x that maintains the ordering of individuals' willingness-to-pay and thus the corresponding surplus from buying insurance conditional on the share of insured individuals $Q(x)$. We denote by $\tilde{w}(x)$ an individual's net willingness-to-pay and the corresponding density by $g^{\tilde{w}(x)}$. For the marginal individual, the net willingness-to-pay equals $\tilde{w}(x) = P(x) = D^{-1}(Q(x);x)$, while the original willingness-to-pay equals $\tilde{w}(0) = D^{-1}(Q(x);0)$. Equilibrium welfare for policy x (not accounting for the budgetary cost) equals

$$\begin{aligned}\mathcal{W}(x) &= \int_{P(x)} E_{\tilde{w}(x)=\tilde{w}'}(s)\, g^{\tilde{w}(x)}\left(\tilde{w}'\right) d\tilde{w}' \\ &= \int_{D^{-1}(Q(x);0)} E_{w=w'}(s)\, g\left(w'\right) dw'.\end{aligned}$$

The second equality follows by maintaining the ordering for any intensity of the policy x. Hence, the welfare effect of an increase in x only depends its impact on the marginal buyer. Using Leibniz' rule and $\frac{\partial}{\partial Q}D^{-1} = \frac{-1}{g^{\tilde{w}(x)}(P(x))}$, we find

$$\mathcal{W}'(x) = E_{P(x)}(s)\, Q'(x) = \left[P(x) - E_{P(x)}(c) - E_{P(x)}(f)\right] Q'(x).$$

Hence, a policy that increases Q, ceteris paribus, increases welfare if and only if $P(x) - E_{P(x)}(c) \geq E_{P(x)}(f)$.$\square$

Proof of Proposition 2:

The equilibrium price is characterized by $P^c(\alpha) = E_{\geq P^c(\alpha)}(c)$, where

$$E_{\geq P^c}(c) = \frac{1}{1 - G^{\tilde{w}(\alpha)}(P^c)} \int_{P^c}^{\infty} E\left(c|\tilde{w}(\alpha) = \tilde{w}'\right) g^{\tilde{w}(\alpha)}\left(\tilde{w}'\right) d\tilde{w}',$$

denoting an individual's net willingness-to-pay by $\tilde{w}(\alpha)$ and density by $g^{\tilde{w}(\alpha)}$. The corresponding equilibrium quantity equals $Q^c(\alpha) = D(P^c(\alpha);\alpha)$. We can solve for the impact of the policy on the equilibrium coverage by implicit differentiation of the equilibrium condition.

We consider first the impact of a change in the price on the equilibrium condition. Using Leibniz' rule, we find

$$\frac{\partial}{\partial P}\left[\int_{P^c} E\left(c|\tilde{w}(\alpha) = \tilde{w}'\right) g^{\tilde{w}(\alpha)}\left(\tilde{w}'\right) d\tilde{w}'\right] = -E_{P^c}(c)\, g^{\tilde{w}(\alpha)}(P^c).$$

Hence, using $\partial\left[1 - G^{\tilde{w}(\alpha)}(P)\right]/\partial P = -g^{\tilde{w}(\alpha)}(P)$, we find

$$\frac{\partial}{\partial P} E_{\geq P^c}(c) = \left[E_{\geq P^c}(c) - E_{P^c}(c)\right] \frac{g^{\tilde{w}(\alpha)}(P^c)}{1 - G^{\tilde{w}(\alpha)}(P^c)},$$

which depends on the difference between average and marginal cost.

We now consider the impact of a change in the policy on the equilibrium condition. We will show that

$$\frac{\partial}{\partial \alpha}\left[\int_{P^c} E\left(c|\tilde{w}(\alpha) = \tilde{w}'\right) g^{\tilde{w}(\alpha)}\left(\tilde{w}'\right) d\tilde{w}'\right] = -E_{P^c}(c \times f)\, g^{\tilde{w}(\alpha)}(P^c), \tag{5}$$

and, in an analogue way, we can find

$$\frac{\partial}{\partial\alpha}\left[\int_{P^c} g^{\tilde{w}(\alpha)}\left(w'\right)dw'\right] = -E_{P^c}\left(f\right)g^{\tilde{w}(\alpha)}\left(P^c\right).$$

Hence,

$$\begin{aligned}\frac{\partial}{\partial\alpha}E_{\geq P^c}\left(c\right) &= \left[E_{\geq P^c}\left(c\right)E_{P^c}\left(f\right) - E_{P^c}\left(c\times f\right)\right]\frac{g^{\tilde{w}(\alpha)}\left(P^c\right)}{1-G^{\tilde{w}(\alpha)}\left(P^c\right)} \\ &= \left[\left(E_{\geq P^c}\left(c\right) - E_{P^c}\left(c\right)\right)E_{P^c}\left(f\right) - cov_{P^c}\left(c,f\right)\right]\frac{g^{\tilde{w}(\alpha)}\left(P^c\right)}{1-G^{\tilde{w}(\alpha)}\left(P^c\right)},\end{aligned}$$

which again depends on the difference between average and marginal costs, but also on the covariance between costs and frictions among the marginal buyers, $cov_{P^c}\left(c,f\right)$*.*

The key step is to prove condition (5). Evaluating the impact of the information policy on $E_{\geq P^c}\left(c\right)$ *is more involved than evaluating the impact of a price change, since the policy changes the sorting into insurance. However, we can re-write*

$$\begin{aligned}&\int_{P^c} E\left(c|\tilde{w}\left(\alpha\right)=\tilde{w}'\right)g^{\tilde{w}(\alpha)}\left(\tilde{w}'\right)d\tilde{w}' \\ &= \int\int_{P^c} E\left(c|\tilde{w}\left(\alpha\right)=\tilde{w}', f=f'\right)g^{\tilde{w}(\alpha)|f}\left(\tilde{w}'|f'\right)d\tilde{w}'g^f\left(f'\right)df' \\ &= \int\int_{P^c} E\left(c|w=\tilde{w}'+\alpha f', f=f'\right)g^{w|f}\left(\tilde{w}'+\alpha f'|f'\right)d\tilde{w}'g^f\left(f'\right)df' \\ &= \int\int_{P^c+\alpha f'} E\left(c|w=w', f=f'\right)g^{w|f}\left(w'|f'\right)dw'g^f\left(f'\right)df'.\end{aligned}$$

The first equality follows from the law of iterated expectations. The second equality follows from the identity $\tilde{w}\left(\alpha\right)=w-\alpha f$*. Hence, conditional on* $f=f'$*, the density of the net willingness-to-pay* $\tilde{w}\left(\alpha\right)$ *equals the density of the willingness to pay* w *shifted by* $\alpha f'$*,* $g^{\tilde{w}(\alpha)|f}\left(\tilde{w}'|f'\right)=g^{w|f}\left(\tilde{w}'+\alpha f'|f'\right)$*. Moreover, conditioning on* $\tilde{w}\left(\alpha\right)=\tilde{w}'$ *and* $f=f'$ *is equivalent to conditioning on* $w=\tilde{w}'+\alpha f'$ *and* $f=f'$*. The last equality uses the substitution* $w'=\tilde{w}'+\alpha f'$*, for each* $f=f'$*, and thus* $dw'=d\tilde{w}'$*.*

After this manipulation, we can simply apply Leibniz' rule again

$$\begin{aligned}&\frac{\partial}{\partial\alpha}\left[\int_{P^c} E\left(c|\tilde{w}\left(\alpha\right)=\tilde{w}'\right)g^{\tilde{w}(\alpha)}\left(\tilde{w}'\right)d\tilde{w}'\right] \\ =&\ \int\frac{\partial}{\partial\alpha}\left[\int_{P^c+\alpha f} E\left(c|w=w', f=f'\right)g^{w|f}\left(w'|f'\right)dw'\right]g^f\left(f'\right)df' \\ =&\ -\int\left[E\left(c|w=P^c+\alpha f', f=f'\right)g^{w|f}\left(P^c+\alpha f'|f'\right)\times f'\right]g^f\left(f'\right)df' \\ =&\ -\int\left[E\left(c\times f|w=P^c+\alpha f', f=f'\right)g^{w|f}\left(P^c+\alpha f'|f'\right)\right]g^f\left(f'\right)df'.\end{aligned}$$

Using again the law of iterated expectations and the identity $\tilde{w}\left(\alpha\right)=w-\alpha f$*, we thus indeed find*

$$\frac{\partial}{\partial\alpha}\left[\int_{P^c} E\left(c|\tilde{w}\left(\alpha\right)=\tilde{w}'\right)g^{\tilde{w}(\alpha)}\left(\tilde{w}'\right)d\tilde{w}'\right] = -E_{P^c}\left(c\times f\right)g^{\tilde{w}(\alpha)}\left(P^c\right).$$

By implicit differentiation of the equilibrium condition $D^{-1}\left(Q^{c}\left(\alpha\right);\alpha\right) = E_{\geq D^{-1}\left(Q^{c}\left(\alpha\right);\alpha\right)}\left(c\right)$, *expressed in quantities, we find*

$$Q^{c\prime}\left(\alpha\right) = -\frac{\left[1-\frac{\partial}{\partial P^{c}}E_{\geq P^{c}}\left(c\right)\right]\frac{\partial D^{-1}\left(Q^{c};\alpha\right)}{\partial\alpha}-\frac{\partial}{\partial\alpha}E_{\geq P^{c}}\left(c\right)}{\left[1-\frac{\partial}{\partial P^{c}}E_{\geq P^{c}}\left(c\right)\right]\frac{\partial D^{-1}\left(Q^{c};\alpha\right)}{\partial Q}}.$$

Using $\frac{\partial D^{-1}\left(Q^{c};\alpha\right)}{\partial Q} = \frac{1}{-g^{\tilde{w}\left(\alpha\right)}\left(P^{c}\right)}$, $\frac{\partial D^{-1}\left(Q^{c};\alpha\right)}{\partial\alpha} = -E_{P^{c}}\left(f\right)$, $\frac{g^{\tilde{w}\left(\alpha\right)}\left(P^{c}\right)}{1-G^{\tilde{w}\left(\alpha\right)}\left(P^{c}\right)} = \frac{\left|\varepsilon_{D}\left(P^{c}\right)\right|}{P^{c}}$ *and the expressions derived before, we find*

$$Q^{c\prime}\left(\alpha\right) = -\frac{E_{P^{c}}\left(f\right)-cov_{P^{c}}\left(c,f\right)\frac{\left|\varepsilon_{D}\left(P^{c}\right)\right|}{P^{c}}}{1-\left[E_{\geq P^{c}}\left(c\right)-E_{P^{c}}\left(c\right)\right]\frac{\left|\varepsilon_{D}\left(P^{c}\right)\right|}{P^{c}}}g^{\tilde{w}\left(\alpha\right)}\left(P^{c}\right),$$

where the terms with $\left[E_{\geq P^{c}}\left(c\right)-E_{P^{c}}\left(c\right)\right]$ *dropped out of the numerator. Finally, for a uniform subsidy* S *such that* $P^{c} = E_{\geq P^{c}}\left(c\right)+S$, *we have*

$$Q^{c\prime}\left(S\right) = \frac{1}{1-\left[E_{\geq P^{c}}\left(c\right)-E_{P^{c}}\left(c\right)\right]\frac{\left|\varepsilon_{D}\left(P^{c}\right)\right|}{P^{c}}}g^{\tilde{w}\left(\alpha\right)}\left(P^{c}\right).$$

Defining this quantity effect as η^{c}, *the first expression in the Proposition follows.*

We now consider the impact on welfare. Welfare equals

$$\mathcal{W}\left(\alpha\right) = \int_{P\left(\alpha\right)}E\left(s|\tilde{w}\left(\alpha\right)=\tilde{w}'\right)g^{\tilde{w}\left(\alpha\right)}\left(\tilde{w}'\right)d\tilde{w}',$$

where $P\left(\alpha\right) = D^{-1}\left(Q\left(\alpha\right),\alpha\right)$. *The total impact of the policy on welfare depends on the policy's effect on the equilibrium quantity and its direct effect on welfare,*

$$\mathcal{W}'\left(\alpha\right) = \frac{\partial\mathcal{W}}{\partial P}\frac{\partial D^{-1}\left(Q\left(\alpha\right),\alpha\right)}{\partial Q}Q'\left(\alpha\right)+\frac{\partial\mathcal{W}}{\partial P}\frac{\partial D^{-1}\left(Q\left(\alpha\right),\alpha\right)}{\partial\alpha}+\frac{\partial\mathcal{W}}{\partial\alpha}.$$

By analogy to the above argument for the positive impact, we find

$$\begin{aligned}\frac{\partial\mathcal{W}}{\partial P} &= -E_{P\left(\alpha\right)}\left(s\right)g^{\tilde{w}\left(\alpha\right)}\left(P\left(\alpha\right)\right)\\ \frac{\partial\mathcal{W}}{\partial\alpha} &= -E_{P\left(\alpha\right)}\left(s\times f\right)g^{\tilde{w}\left(\alpha\right)}\left(P\left(\alpha\right)\right).\end{aligned}$$

Using $\frac{\partial D^{-1}\left(Q\left(\alpha\right);\alpha\right)}{\partial Q} = \frac{1}{-g^{\tilde{w}\left(\alpha\right)}\left(P\left(\alpha\right)\right)}$ *and* $\frac{\partial D^{-1}\left(Q\left(\alpha\right);\alpha\right)}{\partial\alpha} = -E_{P\left(\alpha\right)}\left(f\right)$, we find

$$\begin{aligned}\mathcal{W}'\left(\alpha\right) &= E_{P\left(\alpha\right)}\left(s\right)Q'\left(\alpha\right)+E_{P\left(\alpha\right)}\left(s\right)E_{P\left(\alpha\right)}\left(f\right)g^{\tilde{w}\left(\alpha\right)}\left(P\left(\alpha\right)\right)-E_{P\left(\alpha\right)}\left(s\times f\right)g^{\tilde{w}\left(\alpha\right)}\left(P\left(\alpha\right)\right)\\ &= E_{P\left(\alpha\right)}\left(s\right)Q'\left(\alpha\right)-cov_{P\left(\alpha\right)}\left(s,f\right)g^{\tilde{w}\left(\alpha\right)}\left(P\left(\alpha\right)\right)\end{aligned}$$

and the second expression of the Proposition immediately follows as well.□

Proof of Proposition 3 *For a risk-adjustment policy* β *the competitive equilibrium is determined by*

$$P^{c}\left(\beta\right) = E_{\geq P^{c}\left(\beta\right)}\left(\tilde{c}\left(\beta\right)\right)$$

and $Q^c(\beta) = D^{-1}(P^c(\beta))$. *By implicit differentiation, we find*

$$Q^{c\prime}(\beta) = -\frac{-\frac{\partial}{\partial\beta}E_{\geq P^c}(\tilde{c}(\beta))}{\left[1 - \frac{\partial}{\partial P^c}E_{\geq P^c}(\tilde{c}(\beta))\right]\frac{\partial D^{-1}(Q^c;\beta)}{\partial Q}}.$$

Like in the proof of Proposition 1, we find $\frac{\partial D^{-1}(Q^c;\beta)}{\partial Q} = \frac{1}{-g(P^c)}$ *and*

$$\frac{\partial}{\partial P^c}E_{\geq P^c}(\tilde{c}(\beta)) = \left[E_{\geq P^c}(\tilde{c}(\beta)) - E_{P^c}(\tilde{c}(\beta))\right]\frac{g(P^c)}{1 - G(P^c)}.$$

Moreover,

$$\begin{aligned}
\frac{\partial}{\partial\beta}E_{\geq P^c}(\tilde{c}(\beta)) &= \frac{\partial}{\partial\beta}\left[\frac{1}{1 - G(P^c)}\int_{P^c}^{\infty}E_{w'}(\tilde{c}(\beta))\,g(w')\,dw'\right] \\
&= \frac{1}{1 - G(P^c)}\int_{P^c}^{\infty}E_{w'}(-[c - Ec])\,g(w')\,dw' \\
&= E_{\geq P^c}(c) - Ec
\end{aligned}$$

Hence,

$$\begin{aligned}
Q^{c\prime}(\beta) &= \frac{E_{\geq P^c}(c) - Ec}{1 - \left[E_{\geq P^c}(c) - E_{P^c}(c)\right]\frac{|\varepsilon_D(P^c)|}{P^c}}g(P^c) \\
&= \eta^c \times \left[E_{\geq P^c}(c) - Ec\right],
\end{aligned}$$

where η^c *equals the equilibrium impact of a uniform subsidy.*

Welfare equals

$$\mathcal{W}(\beta) = \int_{D^{-1}(Q^c(\beta))}E_w(s)\,g(w')\,dw'.$$

Hence,

$$\mathcal{W}'(\beta) = -E_{P^c}(s)\frac{1}{-g(P^c)}g(P^c)\,Q^{c\prime}(\beta).$$

This proves the second expression of the Proposition.□

B Appendix: Cost Model Setup and Estimation

This appendix describes the details of the cost model, which is summarized at a high-level in section 4.[24] The output of this model, F_{kjt}, is a family-plan-time specific distribution of predicted out-of-pocket expenditures for the upcoming year. This distribution is an important input into the choice model, where it enters as a family's predictions of its out-of-pocket expenses at the time of plan choice, for each plan option.[25] We predict this distribution in a sophisticated manner that incorporates (i) past diagnostic information (ICD-9 codes) (ii) the Johns Hopkins ACG predictive medical software package (iii) a non-parametric model linking modeled health risk to total medical expenditures using observed cost data and (iv) a detailed division of medical claims and health plan characteristics to precisely map total medical expenditures to out-of-pocket expenses.The level of precision we gain from the cost model leads to more credible estimates of the choice parameters of primary interest (e.g., risk preferences and information friction impacts).

In order to most precisely predict expenses, we categorize the universe of total medical claims into four mutually exclusive and exhaustive subdivisions of claims using the claims data. These categories are (i) hospital and physician (ii) pharmacy (iii) mental health and (iv) physician office visit. We divide claims into these four specific categories so that we can accurately characterize the plan-specific mappings from total claims to out-of-pocket expenditures since each of these categories maps to out-of-pocket expenditures in a different manner. We denote this four dimensional vector of claims C_{it} and any given element of that vector $C_{d,it}$ where $d \in D$ represents one of the four categories and i denotes an individual (employee or dependent). After describing how we predict this vector of claims for a given individual, we return to the question of how we determine out-of-pocket expenditures in plan j given C_{it}.

Denote an individual's past year of medical diagnoses and payments by ξ_{it} and the demographics age and sex by ζ_{it}. We use the ACG software mapping, denoted A, to map these characteristics into a predicted mean level of health expenditures for the upcoming year, denoted θ:

$$A : \xi \times \zeta \to \theta$$

In addition to forecasting a mean level of total expenditures, the software has an application that predicts future mean pharmacy expenditures. This mapping is analogous to A and outputs a prediction λ for future pharmacy expenses.

We use the predictions θ and λ to categorize similar groups of individuals across each of four claims categories in vector in C_{it}. Then for each group of individuals in each claims category, we use the actual ex post realized claims for that group to estimate the ex ante distribution for each individual under the assumption that this distribution is identical for all individuals within the cell. Individuals are categorized into cells based on different metrics for each of the four elements of C:

$$\begin{aligned} \text{Pharmacy:} \quad & \lambda_{it} \\ \text{Hospital / Physician (Non-OV):} \quad & \theta_{it} \\ \text{Physician Office Visit:} \quad & \theta_{it} \\ \text{Mental Health:} \quad & C_{MH,i,t-1} \end{aligned}$$

For pharmacy claims, individuals are grouped into cells based on the predicted future mean phar-

[24]The model is similar to that used in Handel (2013).

[25]In the consumer choice model, this is mostly useful for estimating out-of-pocket expenditures in the HDHP, since the PPO plan has essentially zero expenditures.

macy claims measure output by the ACG software, λ_{it}. For the categories of hospital / physician (non office visit) and physician office visit claims individuals are grouped based on their mean predicted total future health expenses, θ_{it}. Finally, for mental health claims, individuals are grouped into categories based on their mental health claims from the previous year, $C_{MH,i,t-1}$ since (i) mental health claims are very persistent over time in the data and (ii) mental health claims are uncorrelated with other health expenditures in the data. For each category we group individuals into a number of cells between 8 and 12, taking into account the trade off between cell size and precision.

Denote an arbitrary cell within a given category d by z. Denote the population in a given category-cell combination (d, z) by I_{dz}. Denote the empirical distribution of ex-post claims in this category for this population $\hat{G_{I_{dz}}}(\cdot)$. Then we assume that each individual in this cell has a distribution equal to a continuous fit of $\hat{G_{I_{dz}}}(\cdot)$, which we denote G_{dz}:

$$\varpi : \hat{G_{I_{dz}}}(\cdot) \rightarrow G_{dz}$$

We model this distribution continuously in order to easily incorporate correlations across d. Otherwise, it would be appropriate to use $G_{I_{dz}}$ as the distribution for each cell.

The above process generates a distribution of claims for each d and z but does not model correlations over D. It is important to model correlation over claim categories because it is likely that someone with a bad expenditure shock in one category (e.g., hospital) will have high expenses in another area (e.g., pharmacy). We model correlation at the individual level by combining marginal distributions G_{idt} $\forall$ d with empirical data on the rank correlations between pairs (d, d').[26] Here, G_{idt} is the distribution G_{dz} where $i \in I_{dz}$ at time t. Since correlations are modeled across d we pick the metric θ to group people into cells for the basis of determining correlations (we use the same cells that we use to determine group people for hospital and physician office visit claims). Denote these cells based on θ by z_θ. Then for each cell z_θ denote the empirical rank correlation between claims of type d and type d' by $\rho_{z_\theta}(d, d')$. Then, for a given individual i we determine the joint distribution of claims across D for year t, denoted $H_{it}(\cdot)$, by combining i's marginal distributions for all d at t using $\rho_{z_\theta}(d, d')$:

$$\Psi : G_{iDt} \times \rho_{z_{\theta_{it}}}(D, D') \rightarrow H_{it}$$

Here, G_{iDt} refers to the set of marginal distributions $G_{idt} \forall d \in D$ and $\rho_{z_{\theta_{it}}}(D, D')$ is the set of all pairwise correlations $\rho_{z_{\theta_{it}}}(d, d') \forall (d, d') \in D^2$. In estimation we perform Ψ by using a Gaussian copula to combine the marginal distribution with the rank correlations, a process which we describe momentarily.

The final part of the cost model maps the joint distribution H_{it} of the vector of total claims C over the four categories into a distribution of out of pocket expenditures for each plan. For the HDHP we construct a mapping from the vector of claims C to out of pocket expenditures OOP_j:

$$\Omega_j : C \rightarrow OOP_j$$

This mapping takes a given draw of claims from H_{it} and converts it into the out of pocket expenditures an individual would have for those claims in plan j. This mapping accounts for plan-specific features such as the deductible, co-insurance, co-payments, and out of pocket maximums listed in table A-2. We test the mapping Ω_j on the actual realizations of the claims vector C to verify that our mapping comes close to reconstructing the true mapping. Our mapping is necessarily simpler

[26] It is important to use rank correlations here to properly combine these marginal distribution into a joint distribution. Linear correlation would not translate empirical correlations to this joint distribution appropriately.

and omits things like emergency room co-payments and out of network claims. We constructed our mapping with and without these omitted categories to ensure they did not lead to an incremental increase in precision. We find that our categorization of claims into the four categories in C passed through our mapping Ω_j closely approximates the true mapping from claims to out-of-pocket expenses. Further, we find that it is important to model all four categories described above: removing any of the four makes Ω_j less accurate.

Once we have a draw of OOP_{ijt} for each i (claim draw from H_{it} passed through Ω_j) we map individual out of pocket expenditures into family out of pocket expenditures. For families with less than two members this involves adding up all the within family OOP_{ijt}. For families with more than three members there are family level restrictions on deductible paid and out-of-pocket maximums that we adjust for. Define a family k as a collection of individuals i_k and the set of families as K. Then for a given family out-of-pocket expenditures are generated:

$$\Gamma_j : OOP_{i_k,jt} \rightarrow OOP_{kjt}$$

To create the final object of interest, the family-plan-time specific distribution of out of pocket expenditures $F_{kjt}(\cdot)$, we pass the total cost distributions H_{it} through Ω_j and combine families through Γ_j. $F_{kjt}(\cdot)$ is then used as an input into the choice model that represents each family's information set over future medical expenses at the time of plan choice. Figure B1 outlines the primary components of the cost model pictorially to provide a high-level overview and to ease exposition.

We note that the decision to do the cost model by grouping individuals into cells, rather then by specifying a more continuous form, has costs and benefits. The cost is that all individuals within a given cell for a given type of claims are treated identically. The benefit is that our method produces local cost estimates for each individual that are not impacted by the combination of functional form and the health risk of medically different individuals. Also, the method we use allows for flexible modeling across claims categories. Finally, we note that we map the empirical distribution of claims to a continuous representation because this is convenient for building in correlations in the next step. The continuous distributions we generate very closely fit the actual empirical distribution of claims across these four categories.

Cost Model Identification and Estimation. The cost model is identified based on the two assumptions of (i) no moral hazard / selection based on private information and (ii) that individuals within the same cells for claims d have the same ex ante distribution of total claims in that category. Once these assumptions are made, the model uses the detailed medical data, the Johns Hopkins predictive algorithm, and the plan-specific mappings for out of pocket expenditures to generate the the final output $F_{kjt}(\cdot)$. These assumptions, and corresponding robustness analyses, are discussed at more length in the main text.

Once we group individuals into cells for each of the four claims categories, there are two statistical components to estimation. First, we need to generate the continuous marginal distribution of claims for each cell z in claim category d, G_{dz}. To do this, we fit the empirical distribution of claims $G_{I_{dz}}$ to a Weibull distribution with a mass of values at 0. We use the Weibull distribution instead of the log-normal distribution, which is traditionally used to model medical expenditures, because we find that the log-normal distribution over-predicts large claims in the data while the Weibull does not. For each d and z the claims greater than zero are estimated with a maximum likelihood fit to the Weibull distribution:

$$\max_{(\alpha_{dz},\beta_{dz})} \Pi_{i\in I_{dz}} \frac{\beta_{dz}}{\alpha_{dz}}(\frac{c_{id}}{\alpha_{dz}})^{\beta_{dz}-1} e^{-(\frac{c_{id}}{\alpha_{dz}})^{\beta_{dz}}}$$

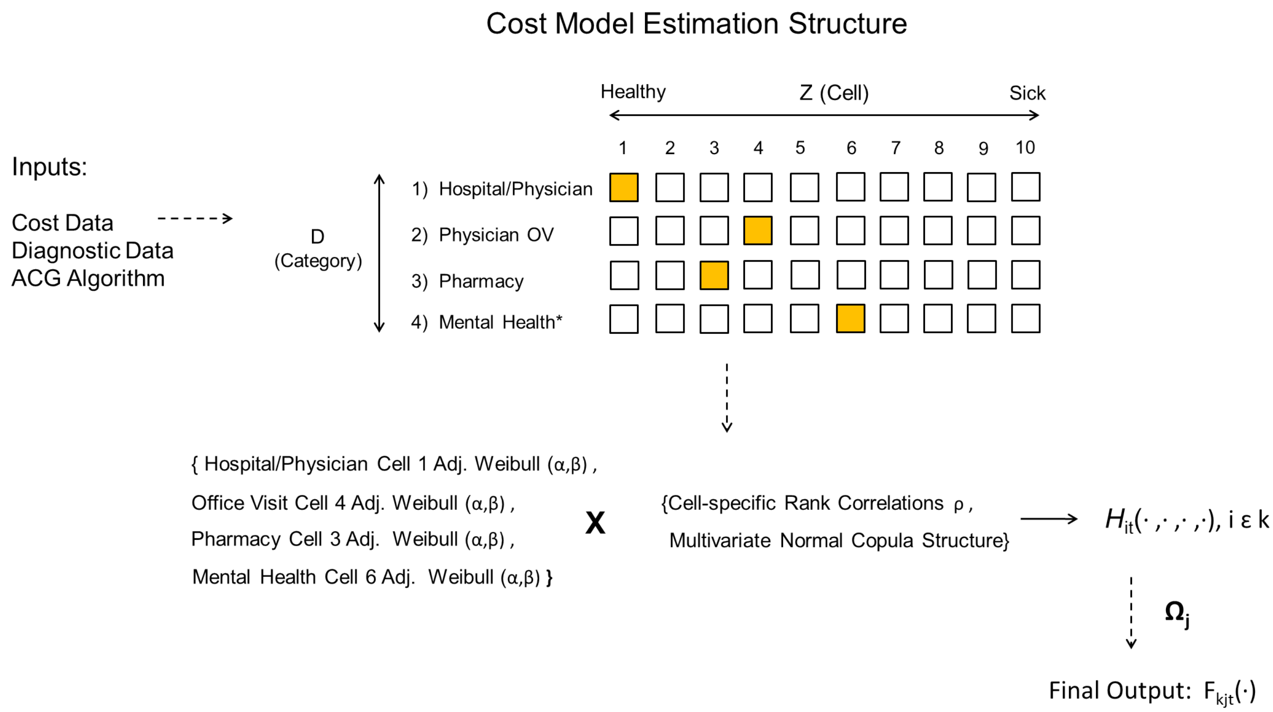

Figure B1: This figure outlines the primary steps of the cost model described in Appendix B. It moves from the initial inputs of cost data, diagnostic data, and the ACG algorithm to the final output F_{kjt} which is the family, plan, time specific distribution of out-of-pocket expenditures that enters the choice model for each family. The figure depicts an example individual in the top segment, corresponding to one cell in each category of medical expenditures. The last part of the model maps the expenditures for all individuals in one family into the final distribution F_{kjt}.

Here, $\hat{\alpha_{dz}}$ and $\hat{\beta_{dz}}$ are the shape and scale parameters that characterize the Weibull distribution. Denoting this distribution $W(\hat{\alpha_{dz}}, \hat{\beta_{dz}})$ the estimated distribution $\hat{G_{dz}}$ is formed by combining this with the estimated mass at zero claims, which is the empirical likelihood:

$$\hat{G_{dz}}(c) = \begin{cases} G_{I_{dz}}(0) & \text{if } c = 0 \\ G_{I_{dz}}(0) + \frac{W(\hat{\alpha_{dz}}, \hat{\beta_{dz}})(c)}{1 - G_{I_{dz}}(0)} & \text{if } c > 0 \end{cases}$$

Again, we use the notation $\hat{G_{iDt}}$ to represent the set of marginal distributions for i over the categories d: the distribution for each d depends on the cell z an individual i is in at t. We combine the distributions $\hat{G_{iDt}}$ for a given i and t into the joint distribution H_{it} using a Gaussian copula method for the mapping Ψ. Intuitively, this amounts to assuming a parametric form for correlation across $\hat{G_{iDt}}$ equivalent to that from a standard normal distribution with correlations equal to empirical rank correlations $\rho_{z_{\theta_{it}}}(D, D')$ described in the previous section. Let $\Phi^{i}_{1|2|3|4}$ denote the standard multivariate normal distribution with pairwise correlations $\rho_{z_{\theta_{it}}}(D, D')$ for all pairings of the four claims categories D. Then an individual's joint distribution of non-zero claims is:

$$\hat{H_{i,t}}(\cdot) = \Phi_{1|2|3|4}(\Phi_1^{-1}(\hat{G_{id_1t}}), \Phi_2^{-1}(\hat{G_{id_2t}}), \Phi_3^{-1}(\hat{G_{id_3t}}), \Phi_4^{-1}(\hat{G_{id_4t}})))$$

Above, Φ_d is the standard marginal normal distribution for each d. $\hat{H}_{i,t}$ is the joint distribution of claims across the four claims categories for each individual in each time period. After this is estimated, we determine our final object of interest $F_{kjt}(\cdot)$ by simulating K multivariate draws from $\hat{H}_{i,t}$ for each i and t, and passing these values through the plan-specific total claims to out of pocket mapping Ω_j and the individual to family out of pocket mapping Γ_j. The simulated $F_{kjt}(\cdot)$ for each k, j, and t is then used as an input into estimation of the choice model.

New Employees. For the first-stage full population model that compares new employees to existing employees to identify the extent of inertia, we need to estimate F_{kj} for new families. Unlike for existing families, we don't observe past medical diagnoses / claims for these families, we just observe these things after they join the firm and after they have made their first health plan choice with the firm. We deal with this issue with a simple process that creates an expected ex ante health status measure. We backdate health status in a Bayesian manner: if a consumer has health status x ex post we construct ex ante health status y as an empirical mixture distribution $f(y|x)$. $f(y|x)$ is estimated empirically and can be thought of as a reverse transition probability (if you are x in period 2, what is the probability you were y in period 1?). Then, for each possible ex ante y, we use the distributions of out-of-pocket expenditures F estimated from the cost model for that type. Thus, the actual distribution used for such employees is described by $\int_{x \in X} f(y|x)F(y)dy$. The actual cost model estimates $F(y)$ do not include new employees and leverages actual claims data for employees who have a past observed year of this data.

C Appendix: Choice Model Identification and Estimation

This appendix describes the algorithm by which we estimate the parameters of the choice model. The corresponding section in the text provided a high-level overview of this algorithm and outlined the estimation assumptions we make regarding choice model fundamentals and their links to observable data.

We estimate the choice model using a random coefficients probit simulated maximum likelihood approach similar to that summarized in Train (2009) and to that used in Handel (2013). The simulated maximum likelihood estimation approach has the minimum variance for a consistent and asymptotically normal estimator, while not being too computationally burdensome in our framework. We set up a likelihood function to predict the health choices of consumers in t_4. The maximum likelihood estimator selects the parameter values that maximize the similarity between actual choices and choices simulated with the parameters.

First, the estimator simulates Q draws for each family from the distribution of health expenditures output from the cost model, F_k for each family. The estimator also simulates D draws for each family-year from the distribution of the random coefficient γ_k, as well as from the distribution of idiosyncratic preference shocks ϵ_{kj}.

We define θ as the full set of model parameters of interest for the full / primary specification in Section 3:[27]

$$\theta \equiv (\mu_\gamma, \delta, \sigma_\gamma, \sigma_\epsilon, \eta_1, \eta_0, \beta).$$

We denote θ_{dk} as one draw derived from these parameters for each family, including the parameters that are constant across draws (e.g., for observable heterogeneity in γ or η) and those which change with each draw (unobservable heterogeneity in γ and ϵ):[28]

$$\theta_{dk} \equiv (\gamma_k, \epsilon_{kJ}, \eta_k, \beta)$$

Denote θ_{Dk} as the set of all D simulated parameter draws for family k. For each $\theta_{dk} \in \theta Dk$, the estimator uses all Q health draws to compute family-plan-specific expected utilities U_{dkj} following the choice model outlined earlier in section 3. Given these expected utilities for each θ_{dk}, we simulate the probability of choosing plan j^* in each period using a smoothed accept-reject function with the form:

$$Pr_{dk}(j = j^*) = \frac{\left(\frac{\frac{1}{-U_{dkj^*}}(\cdot)}{\Sigma_J \frac{1}{-U_{skj}}(\cdot)}\right)^\tau}{\Sigma_{\hat{j}}\left(\frac{\frac{1}{-U_{sk\hat{j}}}(\cdot)}{\Sigma_J \frac{1}{-U_{skj}}(\cdot)}\right)^\tau}$$

This smoothed accept-reject methodology follows that outlined in Train (2009) with some slight modifications to account for the expected utility specification. In theory, conditional on θ_{dk}, we would want to pick the j that maximizes U_{kj} for each family, and then average over D to get final choice probabilities. However, doing this leads to a likelihood function with flat regions, because for small changes in the estimated parameters θ, the discrete choice made does not change. The smoothing function above mimics this process for CARA utility functions: as the smoothing parameter τ becomes large the smoothed Accept-Reject simulator becomes almost identical to the

[27] While we discuss estimation for the full model, the logic extends easily to the other specifications estimated in this paper.

[28] Here, we collapse the parameters determining γ_k and η_k into those factors to keep the notation parsimonious.

true accept-reject simulator just described, where the actual utility-maximizing option is chosen with probability one. By choosing τ to be large, an individual will always choose j^* when $\frac{1}{-U_{kj^*}} > \frac{1}{-U_{kj}} \forall j \neq j^*$. The smoothing function is modified from the logit smoothing function in Train (2009) for two reasons: (i) CARA utilities are negative, so the choice should correspond to the utility with the lowest absolute value and (ii) the logit form requires exponentiating the expected utility, which in our case is already the sum of exponential functions (from CARA). This double exponentiating leads to computational issues that our specification overcomes, without any true content change since both models approach the true accept-reject function.

Denote any choice made $\mathbf{j}$ and the set of such choices as $\mathbf{J}$. In the limit as τ grows large the probability of a given $\mathbf{j}$ will either approach 1 or 0 for a given simulated draw d and family k. For all D simulation draws we compute the choice for k with the smoothed accept-reject simulator, denoted $\mathbf{j}_{dk}$. For any set of parameter values θ_{Sk} the probability that the model predicts $\mathbf{j}$ will be chosen by k is:

$$\hat{P}_k^{\mathbf{j}}(\theta, F_{kj}, X_{kt}^A, X_{kt}^B, \mathbf{Z'}) = \Sigma_{d \in D} \mathbf{1}[\mathbf{j} = \mathbf{j}_{dk}]$$

Let $\hat{P}_k^{\mathbf{j}}(\theta)$ be shorthand notation for $\hat{P}_k^{\mathbf{j}}(\theta, F_{kj}, X_{kt}^A, X_{kt}^B, \mathbf{Z'})$. Conditional on these probabilities for each k, the simulated log-likelihood value for parameters θ is:

$$SLL(\theta) = \Sigma_{k \in K} \Sigma_{\mathbf{j} \in \mathbf{J}} d_{k\mathbf{j}} \; ln \hat{P}_k^{\mathbf{j}}$$

Here $d_{k\mathbf{j}}$ is an indicator function equal to one if the actual choice made by family k was $\mathbf{j}$. Then the maximum simulated likelihood estimator (MSLE) is the value of θ in the parameter space Θ that maximizes $SLL(\theta)$. In the results presented in the text, we choose $Q = 50$, $S = 50$, and $\tau = 6$, all values large enough such that the estimated parameters vary little in response to changes.

A1 Model Implementation and Standard Errors

We implement the estimation algorithm above with the KNITRO constrained optimization package in Matlab. One challenge in non-linear optimization is to ensure that the algorithm finds a global maximum of the likelihood function rather than a local maximum. To this end, we run each model 12 times where, for each model run, the initial parameter values that the optimizer begins its search from are randomly selected from a wide range of reasonable potential values. This allows for robustness with respect to the event that the optimizer finds a local maximum far from the global maximum for a given vector of starting values. We then take the estimates from each of these 12 runs, and select the estimates that have the highest likelihood function value, implying that they are the best estimates (equal to or closest to a global maximum). We ran informal checks to ensure that, for each model, multiple starting values converged to very similar parameters similar to those with the highest likelihood function value, to ensure that we were obtaining robust results.

We compute the standard errors, provided in Appendix F, with a block bootstrap method. This methodology is simple though computationally intensive. First, we construct 50 separate samples, each the same size as our estimation sample, composed of consumers randomly drawn, with replacement, from our actual estimation sample. We then run each model, for 8 different starting values, for each of these 50 bootstrapped samples (implying 400 total estimation runs per model). The 8 starting values are drawn randomly from wide ranges centered at the actual parameter estimates. For each model, and each of the 50 bootstrapped samples, we choose the parameter estimates that have the highest likelihood function value across the 8 runs. This is the final estimate for each bootstrapped sample. Finally, we take these 50 final estimates, across the bootstrapped samples, and calculate the 2.5th and 97.5th percentiles for each parameter and statistic (we actually use the 4th and 96th percentiles given that 50 is a discrete number). Those percentiles are then, respectively, the upper and lower bounds of the 95% confidence intervals presented in Appendix F. See e.g., Bertrand et al. (2004) for an extended discussion of block bootstrap standard errors.

Finally, it is important to note that the 95% confidence intervals presented in Appendix F should really be interpreted as outer bounds on the true 95% intervals, due to computational issues with non-linear optimization. Due to time and computational constraints, we could only run each of the 50 bootstrap sample runs 8 times, instead of 12. In addition, we could not check each of these bootstrapped runs with the same amount of informal checks as for the primary estimates. This implies that, in certain cases, it is possible that one or several of the 50 estimates for each of the bootstrapped samples are not attaining a global maximum. In this case, e.g., it is possible that 45 of the 50 final estimates are attaining global maxima, while 5 are not. As a result, it is possible that the confidence intervals reported are quite wide due to computational uncertainty, even though the 45 runs that attain the global maximum have results that are quite close together. In essence, in cases where computational issues / uncertainty lead to a final estimate for a bootstrapped sample that is not a global maximum, the confidence intervals will look wide (because of these outlier / incorrect final estimates) when most estimates are quite similar. One solution to this issue would be to run each of the models more times (say 12 or 20) for each bootstrapped sample. This would lead to fewer computational concerns, but would take 1.5 to 2.5 times as long, which is substantial since the standard errors for one model take 7-10 days to run.

As a result, the confidence intervals presented should be thought of as outer bounds on the true 95% CIs. This means that for the models where these bounds are tight, the standard error results are conclusive / compelling since the true 95% CI lies in between these already tight bounds. In cases where the CI is very wide, this means that the true 95% CI lies in that wide range, and that we cannot draw meaningful conclusions due to computational uncertainty in all likelihood. Of course, it is possible the true CI is wide, but, in cases where 46 out of 50 bootstrapped parameter estimates are tight and four are outliers (without substantial variations in the underlying samples) this suggests that computational uncertainty is at fault for the wide bounds.

D Appendix: Simulations

To go beyond the local evaluations and provide further insights on how the different model components impact positive and normative outcomes under different policies, we present a series of simulations. We use these simulations to illustrate the role that the key micro-foundations described in this section play in determining market outcomes under (i) no policy interventions (ii) friction-reducing policies and (iii) risk-adjustment policies. Specifically, we distinguish between cases where friction-reducing interventions have positive vs. negative welfare impacts, and cases where effective risk-adjustment policies are essential prior to implementing friction-reducing policies.

Our focus is on a market setup in the mold of Einav et al. (2010), similar to our primary model, where insurers compete to offer supplemental insurance relative to a baseline publicly provided plan. See Appendix E for similar simulations on markets with two competitively priced plans, as studied in Handel et al. (2015).

The baseline plan for these simulations has a deductible of \$3,000, with 10% coinsurance after that point, up to an out-of-pocket maximum of $\$7,000$ (this plan has a 66% actuarial value for our baseline costs below). The supplemental coverage that insurers compete to offer covers all out-of-pocket spending in the baseline plan, and thus brings all consumers up to full insurance. These plans are similar to the minimum and maximum coverage levels regulated in the state-based exchanges set up in the ACA, and also mimic the plans we study in our empirical environment later in this paper. Importantly, in our environment with risk averse consumers and no moral hazard, all consumers purchase full insurance in the first-best. In each simulation we simulate the market for 10,000 consumers.

We study a range of scenarios that vary in terms of the underlying means and variances of (i) consumer surplus from risk protection (ii) consumer costs and (iii) consumer choice frictions. Table D1 describes the underlying distributions for the different cases we study. We simulate two scenarios for consumer yearly expected costs: both have the same mean of just above \$5,000. The first scenario has a high standard deviation of expected costs in the population of \$6,819 while the second has a low standard deviation of \$2,990. For each scenario, consumer expected costs are drawn from a lognormal distribution. The within-year standard deviation in costs for a given consumer is 3,000 plus 1.2 times their yearly expected costs in both scenarios, with each consumer's costs drawn from lognormal distributed as well. The impact of frictions on demand for generous insurance is generated from a normal distribution. The high (low) mean is a \$2,500 (\$0) shift in willingness-to-pay while the high (low) standard deviation we study is \$2,000 (\$500). We study all four combinations of these high/low means and variances. Finally, for consumer risk aversion, we also study four combinations from normal distributions with high/low means and variances. The high (low) CARA mean is $1 * 10^{-3}$ ($4 * 10^{-4}$) while the high (low) standard deviation is $4 * 10^{-4}$ ($1 * 10^{-4}$), with values truncated above 0.[29] The left panel in Figure 3 shows the two distributions of costs studied. The right panel in Figure 3 shows the distribution of surplus in the market when the variance in costs is high under the cases of (i) high mean and variance of risk aversion (ii) low mean and high variance of risk aversion and (iii) low mean and low variance of risk aversion.

We first present a specific simulation example to illustrate the very different impact frictions can have on equilibrium and welfare depending on the primitives of the model. We then systematically investigate positive and normative patterns across a wider range of simulations. The example we start with focuses on two markets that differ only in terms of mean surplus: one has low mean surplus and the other high mean surplus. Otherwise, both markets have a high variance of costs and surplus, and a low mean, but high variance of frictions.

[29]The variance in surplus is likely to increase further when allowing for horizontally differentiated plans.

Simulations **Key Micro-Foundations**		
Costs - μ_c*	Total costs - mean	5,373
Costs - High σ_c*	Total costs - sd	6,819
Costs - Low σ_c*	Total costs - sd	2,990
Surplus - High μ_s**	CARA - mean	$1 * 10^{-3}$
Surplus - Low μ_s**	CARA - mean	$3 * 10^{-4}$
Surplus - High σ_s**	CARA - sd	$4 * 10^{-4}$
Surplus - Low σ_s**	CARA - sd	$1 * 10^{-4}$
Frictions - High μ_f***	WTP shift - mean	2,500
Frictions - Low μ_f***	WTP shift - mean	0
Frictions - High σ_f***	WTP shift - sd	2,000
Frictions - Low σ_f***	WTP shift - sd	500

*Total costs simulated from lognormal distribution.

**CARA risk preferences simulated from normal distribution, truncated above 0.

***Shift in relative willingness-to-pay for high-coverage contract simulated from normal distribution.

Table D1: This table presents the parameter values of the distributions underlying the micro-foundations in our model for the different simulation scenarios we study.

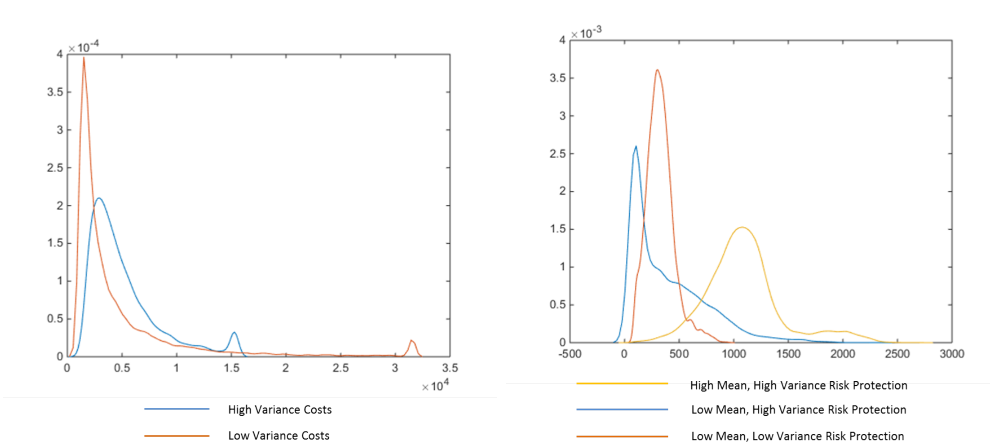

Figure D2: The left panel shows the two different distributions of total costs used in our simulations. The right panel shows the resulting surplus distributions under the different scenarios for the distribution of risk preferences, conditional on the cost distribution with high variance.

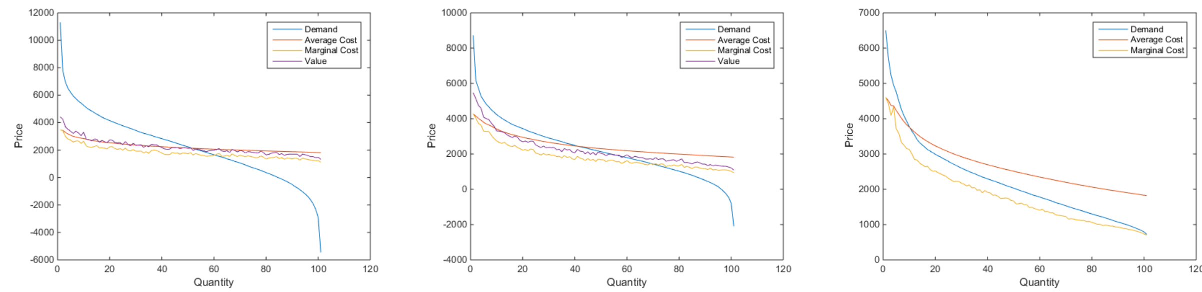

Figure D3: This figure shows the key market micro-foundations for the market with low mean surplus μ_s, in addition to high σ_s, low μ_f, high σ_f, and high σ_c. From left to right, the figure shows the three cases of (i) full frictions (ii) half frictions and (iii) no frictions.

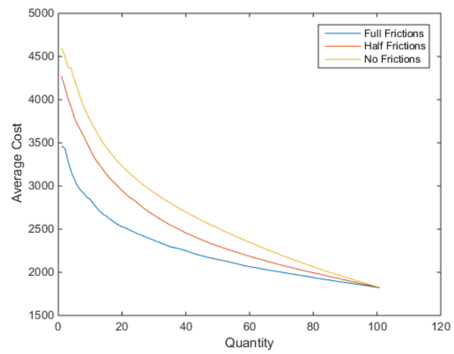

Figure D4: This figure shows the average cost curves, as a function of how much frictions are reduced for the specific simulation example with high σ_s, high σ_f, high σ_c. The mean surplus and friction do not affect this figure as they maintain the ordering of consumers.

Figure 4 shows the key micro-foundations of the market with low mean surplus for the three policy cases of full frictions, frictions reduced by 50%, and no frictions. The figure illustrates a number of properties of markets with low surplus relative to costs when the variance of frictions is meaningful. When full frictions are present, the demand curve is more heavily skewed due to impacts of very positive and negative friction draws. The variance in frictions swamps the variance in costs and surplus, and the market holds together, with quantity of incremental coverage purchased equal to 0.51. When frictions are reduced by 50% ($\alpha = 0.5$) the variation in willingness-to-pay becomes much closer to the variation in costs and value, but the presence of frictions still helps hold the market together, with quantity of incremental coverage equal to 0.41. When frictions are fully removed, the market almost completely unravels, with only 11% of consumers buying incremental coverage. As the figures reveal, as frictions are reduced in this environment, the demand curve becomes less skewed, making it harder to hold the market together at the top end. Consumers with the highest willingness to pay tend to overestimate the insurance value the most and the friction reducing policy reduces their demand for insurance. In addition, as Figure 5 shows, the average cost curves become steeper as frictions are reduced, reflecting increased sorting based on costs.

Figure 6 shows the key micro-foundations of the market with high mean surplus for the same policy interventions. In contrast to the market with low surplus, this case illustrates properties of markets where friction-reducing policies can be beneficial. In this case, when full frictions are present, 64% of consumers purchase coverage in equilibrium. Now, however, when frictions are reduced by 50%, the equilibrium percentage purchasing coverage increases to 79%, and when no

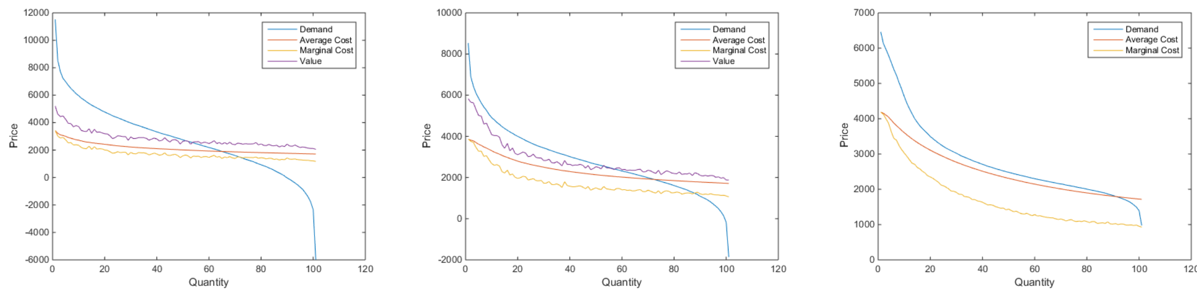

Figure D5: This figure shows the key market micro-foundations for the market with high mean surplus μ_s, in addition to high σ_s, low μ_f, high σ_f, and high σ_c. From left to right, the figure shows the three cases of (i) full frictions (ii) half frictions and (iii) no frictions.

frictions are present the percentage with coverage increases further to 91%. Here, friction-reducing policies have a positive impact on equilibrium coverage. As discussed before, when the share of consumers purchasing coverage is high (due to the high mean surplus), the marginal consumers are more likely to have a bias against purchasing more coverage (i.e., the marginal friction value is negative). As frictions are reduced, these consumers have that bias reduced so that the demand for insurance increases. The level effect of the policy is thus positive in this market. The incremental sorting based on costs when frictions are reduced is the same as in the market with low mean surplus, but this sorting effect is now more than offset by the reverse level effect so that equilibrium coverage increases. Note that the policy not only increases equilibrium coverage, but also increases the match quality and thus will improve welfare as well (as discussed shortly).

We now investigate a broad range of scenarios corresponding to different combinations of the underlying market micro-foundations. Table D2 shows the proportion of consumers buying supplemental insurance as a function of these different micro-foundations. We explore comparative statics for different cases with full frictions present and investigate what happens when those frictions are reduced.

There are several notable patterns. First, conditional on the population distributions of surplus from risk protection s and costs c, reducing the mean level of frictions (which favor purchasing generous coverage) reduces the overall demand for insurance and thus unambiguously reduces the equilibrium quantity purchased. More interestingly, following Proposition **??**, it is clear that the equilibrium implications of reducing the variance in frictions very much depends on the variance of costs. Comparing the first and second columns to the third and fourth columns shows that whether σ_c is high or low has an important impact on the degree of market unraveling as the mean and variance of frictions are reduced. For example, fixing μ_s as low and σ_s as high, when the frictions mean and variance is high, the market share of equilibrium coverage is 0.92 with low σ_c and 0.91 with higher σ_c. When the frictions changes to low mean, high variance, these quantities are 0.56 and 0.51 respectively. But, when the variance in frictions is also reduced to low (along with the mean), these quantities are 0.53 and 0.17. When frictions are fully removed, 39% of consumers purchase more generous coverage for this low σ_c case, but only 11% do in this high σ_c case. Thus, when potential surplus in the market is relatively low, high σ_c implies that reducing market frictions could be especially damaging for market function.

Table D2 also confirms that the mean and variance of surplus (relative to costs and frictions) have important implications for whether frictions are ‘good’ or ‘bad’ for market function. In the cases with low σ_c, low μ_s, and low μ_f, moving from high friction variance to low friction variance has little impact on the equilibrium quantity. When μ_s and μ_f are low, but σ_c is high, moving from high to low σ_f facilitates substantial unraveling (e.g., 0.51 to 0.17 purchasing under high σ_s).

Simulations **Equilibrium Quantities**					
	Low σ_c Low μ_s Low σ_s	Low σ_c Low μ_s High σ_s	High σ_c Low μ_s Low σ_s	High σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
High μ_f, High σ_f	0.93	0.92	0.90	0.91	0.95
High μ_f, Low σ_f	1	1	1	1	1
Low μ_f, High σ_f	0.59	0.56	0.49	0.51	0.64
Low μ_f, Low σ_f	0.65	0.53	0.07	0.17	0.79
No Frictions	0.45	0.39	0.06	0.11	0.91

Table D2: This table presents the proportion of the market purchasing incremental insurance in equilibrium, for a range of underlying population micro-foundations.

However, even with high σ_c, when μ_s and σ_s are high, with low μ_f reducing the variance of frictions *increases* the equilibrium quantity from 0.64 to 0.79. The role (negative) frictions play in pushing people away from generous coverage outweighs the role that they play in reducing adverse selection through sorting.

Taken all together, these results support the earlier analysis by illustrating that (i) reducing the mean impact of frictions on willingness-to-pay for insurance always reduces insurance coverage (ii) reducing the variance and impact of frictions can be good when the mean surplus is relatively high, but (iii) incremental adverse selection occurs and reduces coverage more when the variance in costs is relatively high. These results are further borne out in the bottom of Table D2, which studies the same scenarios, but under the policy where frictions are completely eliminated ($\alpha = 1$). In Appendix F, in Table F5, we also present results for simulations for $\alpha = 0.5$, or partially-reduced frictions, with the comparative statics intuitively following the patterns already described here.

Table D3 presents the proportion of the first-best surplus achieved in each scenario. Notably, welfare is increasing for friction-reducing policies when μ_s and σ_s are high, but decreasing when those values are lower. The sensitivity of the relationship to the level of σ_c is substantial: when σ_c is low the market does not unravel when frictions are reduced, but when σ_c is high it unravels rather quickly and so does the surplus achieved. The welfare implications tend to be in line with the implications for market function: equilibrium surplus increases when equilibrium coverage increases and vice-versa. The exception holds when the variance of surplus is high relative to the variance of costs. In particular, moving from high σ_f to low σ_f (keeping μ_f low), we find that equilibrium coverage decreases, while equilibrium surplus increases. The reason is that the positive matching effect of reduced frictions outweighs the negative equilibrium consequences of any incremental selection on costs, in line with the trade-off highlighted in Proposition **??**. Table F4 in Appendix F illustrates the improved matching by showing how the proportion of mistakes consumers make reduces, given the equilibrium price in each scenario. This table also highlights that while the market unraveling due to friction-reducing policies may decrease total welfare, some consumers will be better off as they now avoid making mistakes.[30]

Both the magnitude and direction of the welfare impact that friction-reducing policies have depend on how effective risk-adjustment transfers in the market are in mitigating adverse selection. Table D4 studies the interaction between friction-reducing policies and risk-adjustment policies.

[30] This table shows, among other things, that the proportion of mistakes made in equilibrium is only related to surplus achieved in the market when the mean and variance of surplus are large relative to frictions and costs.

Simulations **Equilibrium Surplus**					
	Low σ_c Low μ_s Low σ_s	Low σ_c Low μ_s High σ_s	High σ_c Low μ_s Low σ_s	High σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
High μ_f, High σ_f	0.94	0.94	0.90	0.94	0.95
High μ_f, Low σ_f	1	1	1	1	1
Low μ_f, High σ_f	0.62	0.61	0.51	0.61	0.67
Low μ_f, Low σ_f	0.72	0.66	0.14	0.33	0.84
No Frictions	0.56	0.55	0.10	0.23	0.95

Table D3: This table presents the proportion of first-best surplus achieved in the market for a range of underlying population micro-foundations.

We use the underlying distribution of frictions with low μ_f and high σ_f for all risk-adjustment scenarios.

Consider first the case with high σ_c, low μ_s, and high σ_s. When there is no risk-adjustment the market unravels and welfare decreases as frictions are reduced. With partially effective risk-adjustment ($\beta = 0.5$), reducing frictions still reduces equilibrium quantity and welfare, but by a much lesser degree. With full risk-adjustment ($\beta = 1$), there is almost no impact of reduced frictions on quantity, and welfare increases as frictions are reduced. Thus, in this scenario, friction-reducing policies become more tenable, and switch from 'bad' to 'good' as risk-adjustment is more effective. Figure 7 shows the outcomes in this market under full frictions and under no frictions for the three different risk-adjustment scenarios studied.

Compare this now to the case with low σ_c, keeping μ_s low, and σ_s high. With no risk-adjustment reducing frictions has a slight negative impact on equilibrium quantity and welfare. With partial risk-adjustment as frictions are reduced quantity is relatively unchanged but welfare increases substantially, reflecting the impact of better consumer-plan matches. Under full risk-adjustment, both quantity and welfare are strongly increasing as frictions are reduced. Finally, column 3 demonstrates that in the case of high mean surplus for which friction-reducing policies were good for equilibrium quantity and welfare even under no risk-adjustment, this gradient increases as risk-adjustment becomes more effective. Taken in sum, as the mean and variance of surplus increase relative to the mean and variance of costs in the population, the threshold of risk-adjustment necessary to make friction-reducing policies have a positive welfare impact is decreasing.

While our analysis focuses on the case where there is one type of supplemental insurance that is competitively provided, as in Einav et al. (2010), much of the intuition presented in this section extends to the type of market where two classes of plans with different actuarially levels are competitively offered (see e.g., Handel et al. (2015) or Weyl and Veiga (2017)). The key difference in practice between these two types of markets is that the market for supplemental coverage is less likely to unravel, because the supplemental insurer covers only incremental costs rather than the total costs of the sickest consumers. The comparative statics we study remain the same in spirit for this alternative market design: Appendix E presents simulation analysis similar to that presented in this section, but for the case of two priced classes of insurance offerings.[31]

[31] Several insights emerge. First, for a given set of micro-foundations, these multi-plan markets are much more likely to unravel. Consequently, the mean and variance of surplus relative to costs must be substantially higher for friction-reducing policies to have positive impacts in those markets, conditional on a given level of risk-adjustment. In markets with two priced plans, friction-reducing policies are always beneficial under full risk-adjustment, but risk-adjustment

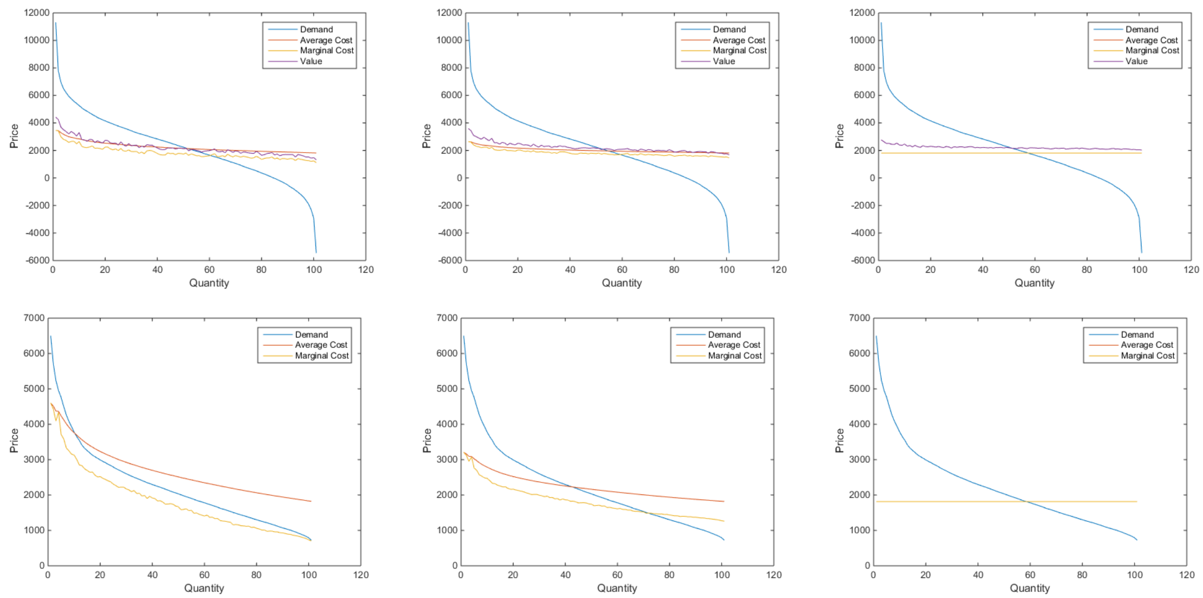

Figure D6: This figure shows market outcomes under different risk-adjustment transfer effectiveness levels. The top panel shows the impact of risk-adjustment with full frictions present, while the bottom shows the impact of risk-adjustment when no frictions are present. The market studied has high σ_c low μ_s, high σ_s, low μ_f, and high σ_f. From left to right, the figure shows the three cases of (i) no risk-adjustment (ii) partial risk-adjustment and (iii) full risk-adjustment.

must be much more effective than in the market for supplemental coverage to make friction-reducing policies welfare increasing. Thus, while the same basic intuition holds in markets with two plan types, policymakers should have a higher threshold for the effectiveness of risk-adjustment when considering the implementation of friction-reducing policies. See Appendix E for more detail on these markets, commonly referred to as exchanges.

Simulations Risk-Adjustment Quantity (% Surplus Achieved)			
	Low σ_c Low μ_s High σ_s	High σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
		No Risk-Adjustment ($\beta = 0$)	
Full Frictions	0.56 (61%)	0.51 (61%)	0.64 (67%)
Half Frictions	0.57 (66%)	0.41 (56%)	0.72 (76%)
No Frictions	0.39 (55%)	0.11 (23%)	0.91 (95%)
		Partial Risk-Adjustment ($\beta = .5$)	
Full Frictions	0.57 (62%)	0.54 (64%)	0.66 (70%)
Half Frictions	0.61 (69%)	0.52 (66%)	0.75 (79%)
No Frictions	0.62 (79%)	0.42 (60%)	0.93 (96%)
		Full Risk-Adjustment ($\beta = 1$)	
Full Frictions	0.58 (64%)	0.57 (66%)	0.68 (72%)
Half Frictions	0.64 (72%)	0.59 (72%)	0.77 (81%)
No Frictions	0.71 (86%)	0.58 (75%)	0.94 (97%)

Table D4: This table presents equilibrium quantity sold, and proportion of total surplus achieved, as a function of the underlying risk-adjustment (β) and friction-reducing policies (α). The entire Table considers the case of low μ_f and high σ_f.

E Appendix: Equilibrium with Two Types of Competing Plans

The primary empirical analysis discussed in the text is for an insurance market where there is a basic government option provided and insurers compete to provide supplemental insurance. As noted in the text, this setup is in the spirit of Einav et al. (2010). An alternative setup we describe in Section 2 and in Section **??** is that where insurers compete to offer two types of insurance plans simultaneously (so costs for both types of plans must break even with premiums in equilibrium). This latter setup is in the spirit of recent work by Handel et al. (2015) studying equilibrium in insurance exchanges. In an example, Veiga and Weyl (2016) illustrate how these two types of setups can lead to markedly different results, primarily because when costs are endogenized for basic coverage the costs incurred by each plan are similar to total expected costs, while when coverage is supplemental, costs are similar only to incremental spending in the supplemental coverage. Thus, the costs faced by the insurers providing generous coverage in the Handel et al. (2015) setup are substantially larger than those when the coverage insurers compete to offer is supplemental. This makes it more likely that equilibrium will unravel towards less generous coverage, because incremental premiums for generous coverage must reflect this larger cost difference.

The analysis in Section 2 considered the choice to buy incremental insurance from a competitive market or stick with a baseline option. Our comparative statics for how key micro-foundations interact with friction-reducing and risk-adjustment policies (and how those foundations determine equilibrium in the absence of such policies) remain the same in the case of more than one type of priced plan. The primary change is that both the high and low coverage plans must account for sorting based on costs in premium setting, whereas in the supplemental insurance case there is no premium for baseline coverage, so it does not adjust along with endogenous sorting. In practice, as shown in Weyl and Veiga (2017) and Handel et al. (2015), this internalization of costs by both plan types leads to unraveling in the market that is of an order of magnitude higher, conditional on the same population consumer micro-foundations.

We briefly illustrate this for a choice between two plans, a low-coverage plan L providing only and a high-coverage plan H. If both plans are priced in competitive markets (as in Weyl and Veiga (2017) and Handel et al. (2015)), each plan needs to internalize the full cost of its own consumers.

We relate our six potentially relevant dimensions of heterogeneity to our original setup as follows:

$$c = c^H - c^L,\ s = s^H - s^L,\ f = f^H - f^L \text{ and } P = P^H - P^L.$$

Since for each plan type the price equals the average cost of the individuals selecting the respective plan, the price differential equals

$$\begin{aligned} P &= E_{\geq P}\left(c^H\right) - E_{<P}\left(c^L\right) \\ &= E_{\geq P}\left(c\right) - \left[E_{<P}\left(c^L\right) - E_{\geq P}\left(c^L\right)\right]. \end{aligned}$$

The second term captures the difference in baseline coverage costs between those actually buying the low-coverage plan relative to those buying the high-coverage plan. If the baseline coverage costs are independent of the sorting of individuals, the previous equilibrium analysis entirely generalizes. If not, we need to take into account how the policy affects the selection based on the full cost into both plans.[32] Similarly, when evaluating policy interventions, the re-sorting based on the full cost in both plans determines the new equilibrium prices. The welfare analysis naturally generalizes

[32] If both plans insure the same underlying risk but differ in their overall coverage, we have "adverse selection" into the high-coverage contract, both for the baseline and supplemental coverage, i.e., $E_{\geq P}\left(c^L\right) \geq E_{<P}\left(c^L\right)$. If two plans insure different types of risk, we may well have "adverse selection" into both contracts, i.e., $E_{<P}\left(c^L\right) \geq E_{\geq P}\left(c^L\right)$.

since the change in total welfare only depends on the change in the differential surplus (as long as the purchase of baseline coverage is mandated). That is,

$$\begin{aligned}\mathcal{W} &= (1 - G(P)) E_{\geq P}\left(s^H\right) + G(P) E_{<P}\left(s^L\right) \\ &= (1 - G(P)) E_{\geq P}(s) + E\left(s^L\right).\end{aligned}$$

See Handel et al. (2015) and Weyl and Veiga (2017) for a much more complete discussion of equilibrium in markets with multiple tiers of competitively priced plans, and how they compare to the market with baseline coverage and privately-provided supplemental coverage. For this paper, it is only important to note that the comparative statics will be the same directionally, regardless, though of course the threshold for what makes a market unravel vs. not it much lower in the markets with two or more types of priced plans.

In Section D we presented simulations to illustrate the relationship between market micro-foundations and different policy recommendations, in the Einav et al. (2010) style market with one priced supplemental plan. Here, in Table E1 we present analogous results for the market with two priced plans. The underlying simulation micro-foundations for each scenario are the same as those described in the main text in Table D1.

In the supplemental market described in Section D, for the scenarios where the distribution of surplus was high relative to costs, the equilibrium held together and friction-reducing policies were welfare improving. In the market for two priced plans, this is not the case. With high mean and variance of frictions the case with high mean and variance of surplus has quantity equal to 0.55. When frictions are reduced, either by 50% or 100% the market completely unravels, in contrast to the supplemental market. This is for the case where the variance in costs is high. For the other two scenarios presented in Table E1, with low variance in costs, low mean surplus, and low or high surplus variance, the results have a similar flavor across the range of frictions present and friction policies. With high mean frictions, the market holds together and quantity provided is high. But, when the mean level of frictions is low, or policies are in place to reduce the high mean frictions, the market fully unravels and no generous insurance is purchased in equilibrium.

These results imply that the market with two priced plans is much more likely to unravel for a given set of micro-foundations. As a result, in this style market, policies to reduce frictions are more likely to be welfare decreasing than in the market with one competitively priced supplemental plan. While the mean and variance of surplus relative to the mean and variance of costs is still a crucial determinant of whether friction-reducing policies will be good or bad, now because of the nature of the market the distribution of insurance must be higher relative to the distribution of costs in order for the market to function and in order for friction-reducing policies to be welfare positive.

A corresponding implication is that the threshold for risk-adjustment that is necessary to make friction-reducing policies welfare positive is higher in the market with two priced plans. Table E2 presents market quantities and welfares for a range of interacted risk-adjustment and friction-reducing policies. As in the main text, results are presented for the case with low μ_f and high σ_f.[33] It is clear that in all cases studied, incremental risk-adjustment increases welfare and is absolutely crucial when implementing friction-reducing policies in the market. For any of the cases presented, when risk-adjustment is either partially effective ($\beta = 0.5$) or not present ($\beta = 0$) friction-reducing policies reduce equilibrium coverage and increase adverse selection. However, when full risk-adjustment is present, friction-reducing policies improve equilibrium quantity and welfare in

[33]Note that when the mean level of frictions are increased, the equilibrium is less likely to unravel, we present this case so it can be directly compared to the supplemental equilibrium in the text.

Simulations-Two Priced Plans **Equilibrium Quantities**			
	Low σ_c Low μ_s Low σ_s	Low σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
	Full Frictions		
High μ_f, High σ_f	0.83	0.79	0.55
High μ_f, Low σ_f	0.9	0	0
Low μ_f, High σ_f	0.33	0.28	0
Low μ_f, Low σ_f	0	0	0
No Frictions	0	0	0
	Half Frictions		
High μ_f, High σ_f	0	0	0
High μ_f, Low σ_f	0	0	0
Low μ_f, High σ_f	0	0	0
Low μ_f, Low σ_f	0	0	0
No Frictions	0	0	0

Table E1: This table presents the proportion of the market purchasing full insurance in equilibrium, for a range of underlying population micro-foundations. These results are for an insurance exchange where two types of plans are offered competitively.

the market.

Thus, the same underlying intuition holds for markets with two priced plans, but the threshold for what constitutes 'enough' risk-adjustment to implement friction-reducing policies is much higher because of the higher potential for adverse selection. This distinction is generally interesting, and reflects the underlying notion that, as the mean and variance of population costs becomes high relative to the mean and variance of surplus from risk protection, friction-reducing policies are more likely to be welfare -decreasing and more risk-adjustment is required for them to be welfare-increasing.

In addition to presenting these simulations, we also conduct the analog to our empirical analysis in the text for the case of two priced plans. Section **??** lays out the model for insurer competition in both the Einav et al. (2010) and Handel et al. (2015) cases. Since the mean and variance of costs are high relative to surplus in our empirical application, it is highly likely that the market will unravel except for cases with very high frictions or very effective risk-adjustment.

Figure E shows market equilibrium for the baseline case where $\alpha = 0$ and $\beta = 0$. The average cost line for generous coverage always lies above the demand curve, even in this case where substantial mean frictions push people towards that coverage. The high mean and variance of consumer costs, relative to their surplus from incremental coverage, leads to this scenario, which we also see in the simulation in Section 2. Not surprisingly, when frictions are partially and fully removed in figures E7 and E8 there is still no positive equilibrium market share of more generous coverage, which is expected given that the frictions we estimate push consumers toward that coverage.

There is some hope for maintaining generous coverage when there is insurer risk-adjustment.

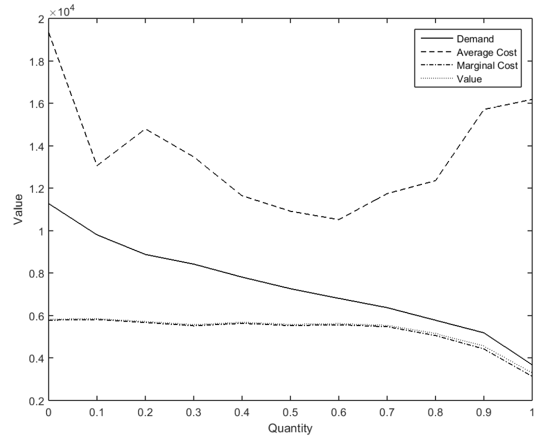

Figure E7: Market Equilibrium Including Information Frictions

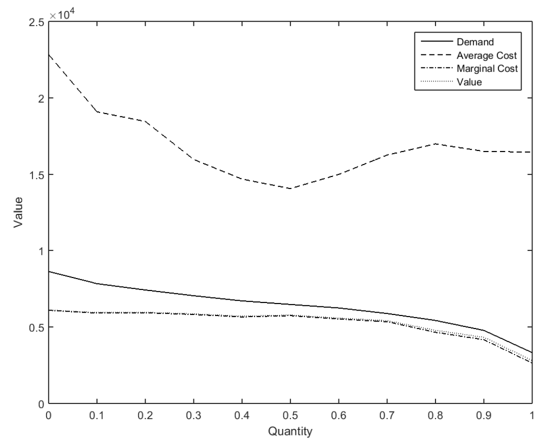

Figure E8: Market Equilibrium with Partial Information Frictions

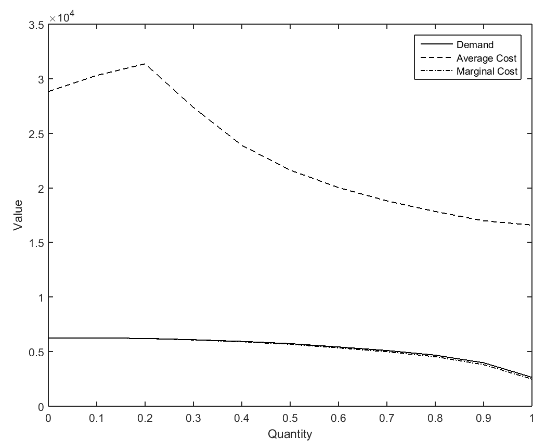

Figure E9: Market Equilibrium without Information Frictions

Simulations-Two Priced Plans Risk-Adjustment Quantity (% Surplus Achieved)			
	Low σ_c Low μ_s High σ_s	High σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
	No Risk-Adjustment ($\beta = 0$)		
Full Frictions	0.28 (35%)	0 (0%)	0 (0%)
Half Frictions	0 (0%)	0 (0%)	0 (0%)
No Frictions	0 (0%)	0 (0%)	0 (0%)
	Partial Risk-Adjustment ($\beta = .5$)		
Full Frictions	0.45 (50%)	0.01 (2%)	0.25 (30%)
Half Frictions	0.07 (11%)	0 (0%)	0 (0%)
No Frictions	0 (0%)	0 (0%)	0 (0%)
	Full Risk-Adjustment ($\beta = 1$)		
Full Frictions	0.58 (64%)	0.57 (66%)	0.68 (72%)
Half Frictions	0.64 (72%)	0.59 (72%)	0.77 (81%)
No Frictions	0.71 (86%)	0.58 (75%)	0.94 (97%)

Table E2: This table presents equilibrium quantity sold, and proportion of total surplus achieved, as a function of the underlying risk-adjustment (β) and friction-reducing policies (α). The entire table considers the case of low μ_f and high σ_f. Results presented are for the market with two competitively priced plans.

Figure E shows that some coverage is possible with full frictions and with either partial or full risk-adjustment. When frictions are removed, even with full risk-adjustment there is full unraveling of the market: this is because the cost of the average consumer for the family tier we study is higher than the top-end value of insurance coverage, given the way that the insurance contracts are set up relative to one another. Handel et al. (2015) shows that equilibrium in the market is harder to maintain the closer the two types of coverage are relative to one another, precisely for this reason.

Thus, with the limited surplus estimated in our environment from risk-protection, and the closeness of the two types of insurance contracts relative to average costs, the market outcome in our environment is almost always full unraveling. There are a few reasons why we might not see this in practice. First, consumers typically receive subsidies to purchase insurance coverage, either from the government in exchanges or from their employer in employer provided insurance. Though one typical principle of managed competition is that consumers receive a lump sum subsidy and pay the full marginal cost of generous coverage, in practice in many exchanges poorer consumers have caps on the premiums that they pay, limiting the relative premium spread between insurance contracts. The second reason is that consumers with frictions may follow decision models whereby they always choose more generous coverage no matter what. With the micro-foundations in our

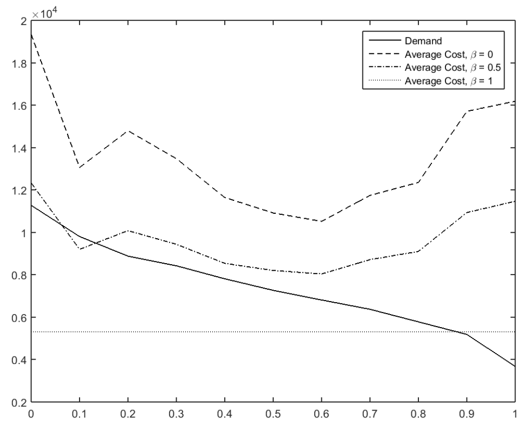

Figure E10: Market Equilibrium with full frictions and a range of risk-adjustment policies.

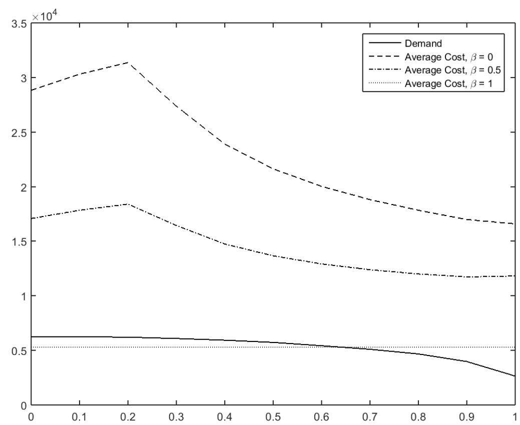

Figure E11: Market Equilibrium with no frictions and a range of risk-adjustment policies.

environment, even the presence of such consumers would not hold the equilibrium together, given the spread because average costs in the PPO and the relative generosity of that coverage, unless the consumers choosing generous coverage by mistake are the healthiest in the population.

F Appendix: Additional Analysis

Table E1 presents summary demographic statistics for the samples we study. The first column represents all employees who were present in our data and have complete records for at least eight months in the four years of data (t_1-t_4) that we observe.[34] The second column represents all employees who received our survey, regardless of whether or not they responded. The third column represents all employees who responded to our survey. Statistics from gender onwards represent only t_3, and use the re-weighted statistics for the second and third columns, as described in the text.

Table F2 presents the details of plan design for the two plans consumers choose between in our empirical environment. Table F5 presents the results for the simulations in Section 2.5 that are for the case of partially effective friction-reducing policies ($\alpha = 0.5$). Table F4 describes the proportion of consumers making choice mistakes in each of the simulation scenarios described in Section 2.5.

Figure E1 presents the smoothed distribution of expected costs from the cost model for families in the primary sample.

Figure F1 depicts the financial returns to selecting the HDHP option relative to the PPO option for an employee in the family tier, which has more than 50% of the employees in our sample.[35] The x-axis plots realized total health expenditures (insurer + insuree) and the y-axis plots the financial returns for the HDHP relative to the PPO as a function of those total expenditures. For a family, the range of potential ex-post value for the HDHP spans $[-\$2,500, +\$3,750]$, with the lower bound coming from cases with a lot of medical spending, the upper bound coming from the case of zero spending.[36] Based on ex post spending 60% of employees, across all tiers, are better off financially in the HDHP, though only 15% of employees actually choose that plan.

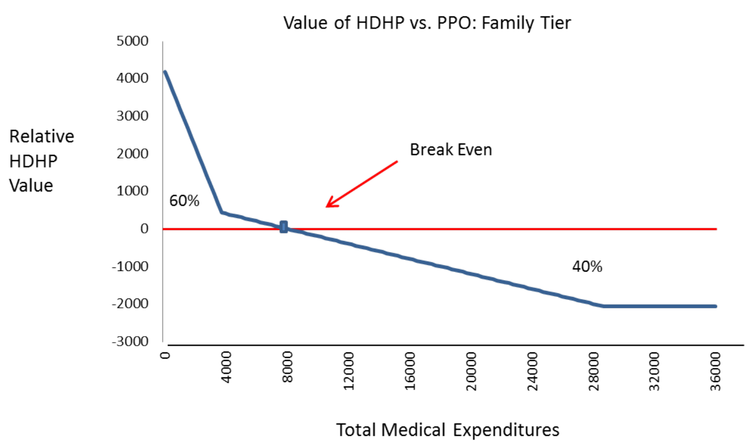

Figure F1: Description of HDHP financial value relative to the PPO in year t_4, for the family tier, as a function of total medical expenditures. This chart assumes that employees contribute 50% of the maximum possible incremental amount to their HSA, near the median in the population. 60% of *all* employees would be better of ex-post in the HDHP, given their respective coverage tiers.

Table F3 presents the results from the primary choice model we use in the main text, discussed

[34] We cannot provide the exact number of overall employees, to preserve the anonymity of the firm. As noted earlier, we cannot state the exact years of the data, though we note they are from a four-year period between 2008-2014.

[35] The same general structure holds for couples and families with shifts in the levels of the key plan terms.

[36] This range shifts upward by a constant amount if consumers derive value from incremental HSA contributions: the figure assumes consumers contribute 50% of the potential incremental contribution, up to the maximum allowed, consistent with what we find in the data. Note also that the relative value range for an individual / couple equals the family bounds multiplied by 0.4 (0.8).

Sample Demographics	Full Sample	Survey Recip. (Weighted)	Survey Resp. (Weighted)
N -Employees	[35,000 , 60,000]*	4500	1661
N_d - Emp.& Dep.	[105,000 , 200,000]*	11,690	4,584
t_3 PPO%	88.8	89.6	88.7
t_4 PPO%	82.7	83.0	81.6
t_3 HDHP %	11.2	10.4	11.3
t_4 HDHP %	17.3	17.0	18.4
Gender, Emp. and Dep. (% Male)	51.8	51.5	51.1
Age			
18-29	8.6%	14.9%	11.6%
30-39	41.1%	43.8%	42.7%
40-49	38.1%	32.7%	34.1%
50-59	10.9%	7.7%	10.5%
$\geq$60	1.3%	0.9%	1.2%
Income			
Tier 1 (< \$100K)	12.8%	15.3%	16.2%
Tier 2 (\$100K-\$150K)	65.8%	68.5%	69.2%
Tier 3 (\$150K-\$200K)	16.7%	14.3%	12.9%
Tier 4 (\$200K+)	3.5%	1.2%	0.9%
Family Size			
1	23.0%	29.0%	20.9%
2	19.0%	19.4%	21.9%
3+	58.0%	51.6%	57.2%
Family Spending			
Mean	\$10,191	\$8,820	\$11,247
Median	\$4,275	\$3,363	\$4,305
25th	\$1,214	\$878	\$1,176
75th	\$10,948	\$9,388	\$11,555
95th	\$35,139	\$32,171	\$41,864
99th	\$87,709	\$80,370	\$87,022

Table F1: This table gives summary statistics for the employees and dependents of the firm we use data from. When not stated, statistics are for year t_4. See Handel and Kolstad (2015) for more information on the population, their information about insurance options, and the link between costs, information, surplus, and insurance choices. Note that we cannot provide the exact sample size for all employees and dependents at the firm, to preserve the anonymity of the firm (though we can discuss sample size of the specific samples used in our analysis.

Health Plan Characteristics **Family Tier**	PPO	HDHP*
Premium	$0	$0
Health Savings Account (HSA)	No	Yes
HSA Subsidy	-	[$3,000-$4,000]**
Max. HSA Contribution	-	$6,250***
Deductible	$0****	[$3,000-$4,000]**
Coinsurance (IN)	0%	10%
Coinsurance (OUT)	20%	30%
Out-of-Pocket Max.	$0****	[$6,000-$7,000]**

* We don't provide exact HDHP characteristics to help preserve firm anonymity.

**Values for family coverage tier (2+ dependents). Single employees (or w/ one dependent) have .4× (.8×) the values given here.

***Single employee legal maximum contribution is $3,100. Employees over 55 can contribute an extra $1,000 in 'catch-up.'

****For out-of-network spending, PPO has a very low deductible and out-of-pocket max. both less than $400 per person.

Table F2: This table presents key characteristics of the two primary plans offered over time at the firm we study. The PPO option has more comprehensive risk coverage while the HDHP option gives a lump sum payment to employees up front but has a lower degree of risk protection. The numbers in the main table are presented for the family tier (the majority of employees) though we also note the levels for single employees and couples below the main table.

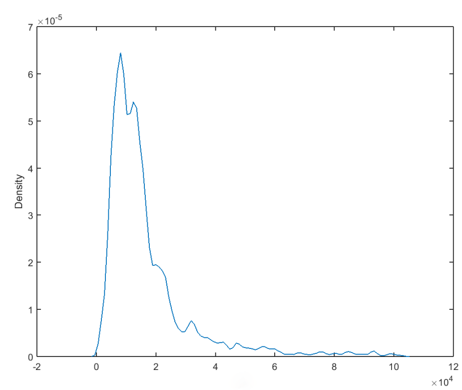

Figure F2: This figure presents the smoothed distribution of expected costs from the cost model for families in the primary sample. These costs are fully covered by the PPO and only partially by the HDHP.

in Section 3. See Handel and Kolstad (2015b) for a wider range of related specifications and estimates.

Primary Model Estimates	Model Estimate	95% CI
Average μ_γ	$8.6 \cdot 10^{-5}$	[$8.19 \cdot 10^{-5}$,$2.23 \cdot 10^{-4}$]
Std. Dev. μ_γ	$1.4 \cdot 10^{-5}$	[$9.41 \cdot 10^{-6}$,$4.41 \cdot 10^{-5}$]
Gamble Interp. of Average μ_γ	920.47	[822.51,924.23]
σ_γ	$2.2 \cdot 10^{-9}$	[$5.98 \cdot 10^{-6}$,$1.55 \cdot 10^{-4}$]
σ_ϵ, HDHP	0.11	[1.58,666.04]
Benefits knowledge:		
Any incorrect	98.04	[-614.70,377.52]
Any 'not sure'	-467.48	[-1670.66,127.94]
Time cost hrs. X prefs:		
Time cost hrs.	-9.72	[-90.07,118.86]
... X Accept, concerned	-118.15	[-282.81,-55.79]
... X Dislike	-128.98	[-293.99,-70.02]
Provider networks:		
HSP network bigger	-594.38	[-1842.45,562.52]
PPO network bigger	-2362.85	[-3957.68,-1286.62]
Not sure	-201.81	[-937.44,303.21]
TME guess:		
Overestimate	62.98	[-810.72,704.28]
Underestimate	-208.30	[-1154.63,837.19]
Not sure	-688.91	[-1987.28,320.99]
Average Friction Effect	-1787.40	[-2148.63,-906.96]
σ Friction Effect	1303.64	[1264.29,2329.12]
Likelihood Ratio Test Stat for Frictions	379.54	

Table F3: This table presents our primary estimates of our empirical choice framework. The first column presents the actual point estimates while the second column presents the 95% CI derived from the bootstrapped standard errors. Here, positive friction values indicate greater willingness-to-pay for high-deductible care.

Simulations **Mistakes in Equilibrium**	Low σ_c Low μ_s Low σ_s	Low σ_c Low μ_s High σ_s	High σ_c Low μ_s Low σ_s	High σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
			Full Frictions		
High μ_f, High σ_f	.23 (.18)	.31 (.27)	.47 (.45)	.36 (.33)	.11 (.07)
High μ_f, Low σ_f	.56 (.56)	.61 (.61)	.50 (.50)	.43 (.43)	.09 (.09)
Low μ_f, High σ_f	.39 (.13)	.39 (.15)	.35 (.24)	.35 (.20)	.35 (.06)
Low μ_f, Low σ_f	.26 (.13)	.21 (.11)	.02 (.01)	.06 (.04)	.17 (.04)
			Half Frictions		
High μ_f, High σ_f	.21 (.18)	.28 (.26)	.44 (.42)	.34 (.32)	.09 (.06)
High μ_f, Low σ_f	.35 (.35)	.30 (.30)	.50 (.50)	.42 (.42)	.09 (.09)
Low μ_f, High σ_f	.33 (.14)	.30 (.14)	.20 (.15)	.23 (.14)	.26 (.06)
Low μ_f, Low σ_f	.17 (.09)	.11 (.06)	.01 (0)	.02 (.01)	.09 (.03)

Table F4: This table presents the proportion of consumers making mistakes when purchasing coverage, given the equilibrium price, for a range of underlying population micro-foundations.

Simulations **Results for $\alpha = 0.5$**	Low σ_c Low μ_s Low σ_s	Low σ_c Low μ_s High σ_s	High σ_c Low μ_s Low σ_s	High σ_c Low μ_s High σ_s	High σ_c High μ_s High σ_s
		Half Frictions, % Purchase			
High μ_f, High σ_f	0.94	0.92	0.86	0.9	0.97
High μ_f, Low σ_f	1	1	1	1	1
Low μ_f, High σ_f	0.63	0.57	0.34	0.41	0.72
Low μ_f, Low σ_f	0.61	0.44	0.06	0.11	0.86
No Frictions	0.45	0.39	0.06	0.11	0.91
		Half Frictions % Surplus			
High μ_f, High σ_f	0.95	0.95	0.87	0.94	0.97
High μ_f, Low σ_f	1	1	1	1	1
Low μ_f, High σ_f	0.68	0.66	0.42	0.56	0.76
Low μ_f, Low σ_f	0.71	0.60	0.10	0.25	0.94
No Frictions	0.56	0.55	0.10	0.23	0.95

Table F5: This table presents the proportion of consumers purchasing generous coverage (top half) and the proportion of first-best surplus achieved in the market (bottom half) for a range of underlying population micro-foundations and a partially effective friction-reducing policy ($\alpha = 0.5$).