Mergers in the Presence of Adverse Selection: Online Appendix

A Supplemental Tables and Figures

Table A1: Control Function Estimation

	(1) All Products	(2) Products Purchased in Choice Data
Fraction of Consumers Under 35	-249	-272
	(94.5)	(120)
Firm	✓	✓
Metal Level	✓	✓
State	✓	✓

Note: This table displays the first stage of the control function estimation for the demand estimation procedure presented in Section 4. The dependent variable is the annual base premium. The first column shows estimates using all products in the markets included in the sample. The second column shows estimates using only products purchased in the choice data used in estimation. The control function is constructed from the residual of specification (1).

Table A2: Demand Estimation Results

	(1)	(2)	(3)
Premium	-2.07	-1.35	-1.32
	(0.02)	(0.02)	(0.01)
Age 31 - 40	0.30	0.28	0.29
	(0.02)	(0.02)	(0.02)
Age 41 - 50	0.62	0.44	0.43
	(0.02)	(0.02)	(0.02)
Age 51 - 64	1.20	0.71	0.70
	(0.01)	(0.01)	(0.01)
Family	0.01	0.06	0.04
	(0.01)	(0.01)	(0.01)
Subsidized	0.34	0.29	0.29
	(0.01)	(0.01)	(0.01)
Actuarial Value (AV)	7.03	11.98	11.70
	(0.05)	(0.08)	(0.08)
Risk Preference			
AV	0.59	0.55	0.54
	(0.00)	(0.00)	(0.00)
Firm - Risk Interaction	Y	Y	Y
Fixed Effects			
Age, Fam., Inc.	Y	Y	Y
Firm	Y		
Firm-Market		Y	
Firm-Category		Y	
Firm-Mkt-Cat.			Y

Note: This table displays the demand estimation results without using the control function as an additional source of identification. The specifications match those estimated in the three GMM columns of Table 1. The estimates are qualitatively similar.

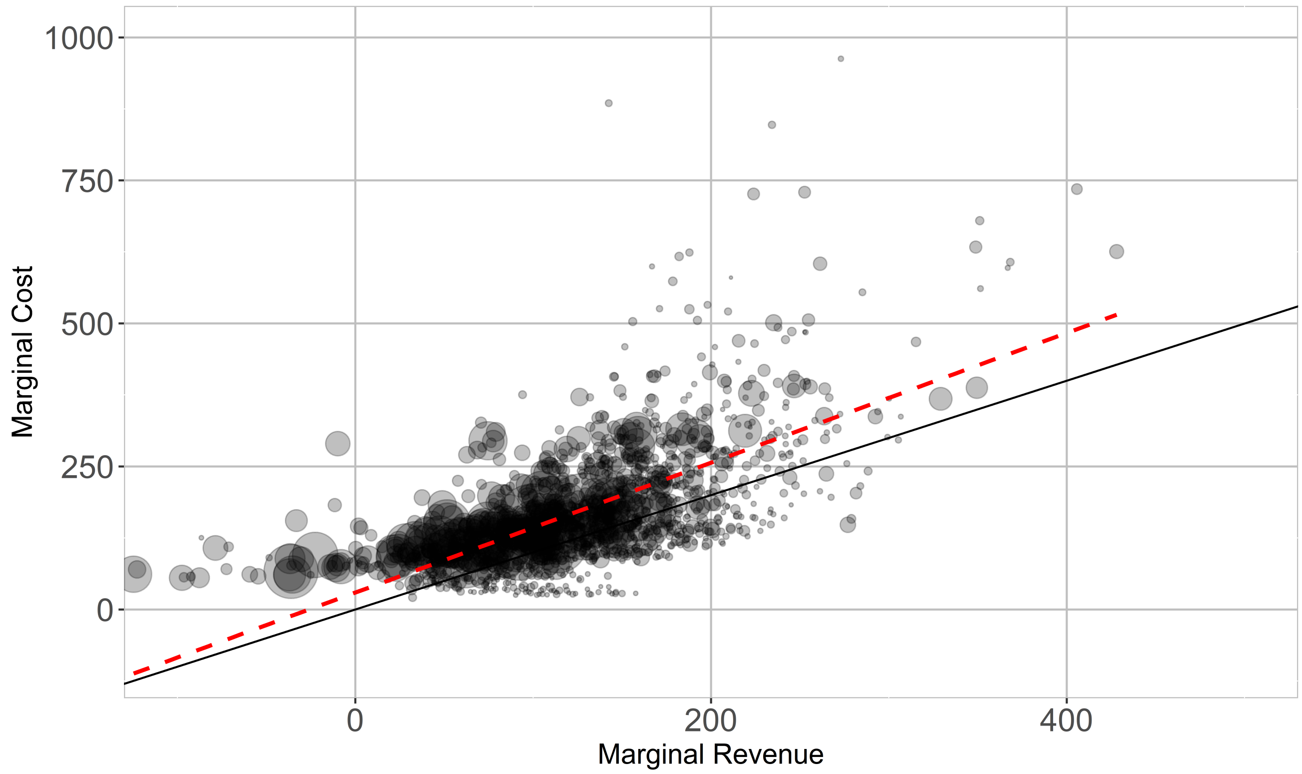

Figure A1: Marginal Revenue vs Marginal Cost in Baseline Model

Note: The product-level marginal cost and marginal revenue predicted by the estimated model are roughly equal on average. Each dot represents a product in a market. The size of the dots is proportional to the quantity sold. The model does relatively well with products that are close to the mean marginal revenue and costs but struggles to fit the outliers.

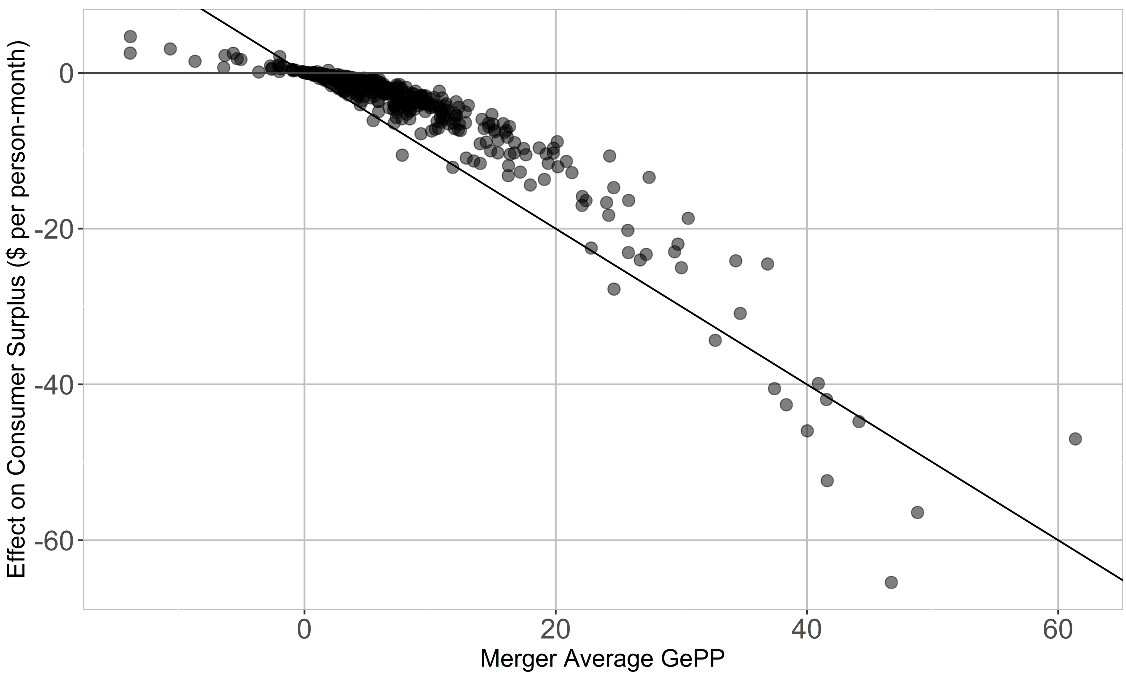

Figure A2: GePP Predicts Direction of Consumer Surplus Effect

Note: Average GePP forms a good prediction for the direction of the effect on consumer surplus. This figure compares the effect of a merger in a particular market relative to the average GePP across all products of the merging parties. Each dot represents a single merger-market. Sorting cost is displayed on a log scale. Both the welfare cost of sorting and the effect on consumer surplus are measured in dollars per person per month.The dark line represents the 45-degree line.

B Derivations for Section 2

B.1 The Average Cost Function

The cost to a particular product j of enrolling a household i is given by c_{ij}. The average cost of an insurance plan is the share-weighted average cost of all consumers that select that plan.

$$AC_j(\boldsymbol{p}) = \frac{1}{S_j(\boldsymbol{p})} \int_i S_{ij}(\boldsymbol{p}) c_{ij} di$$

The derivative of the average cost of product j with respect to the price of product k depends on the demand derivatives of the consumers of product j.

$$\frac{\partial AC_j}{\partial p_k} = \frac{1}{S_j}\int_i \frac{\partial S_{ij}}{\partial p_k} c_{ij} di - \frac{\frac{\partial S_j}{\partial p_k}}{S_j} \frac{\int_i S_{ij}(\boldsymbol{p}) c_{ij} di}{S_j}$$
$$\frac{\partial AC_j}{\partial p_k} = \frac{\frac{\partial S_j}{\partial p_k}}{S_j}\left(\frac{1}{\frac{\partial S_j}{\partial p_k}}\int_i \frac{\partial S_{ij}}{\partial p_k} c_{ij} di - AC_j\right)$$

B.2 Generalized Pricing Pressure

The optimal price for single-product firm, j, is derived below.

$$\begin{aligned}\Pi_j &= S_j(\boldsymbol{p})(p_j - AC_j(\boldsymbol{p}))\\ 0 &= \frac{\partial S_j}{\partial p_j}(p_j - AC_j) + S_j(1 - \frac{\partial AC_j}{\partial p_j})\\ p_j &= AC_j + \frac{S_j}{\frac{\partial S_j}{\partial p_j}}(1 - \frac{\partial AC_j}{\partial p_j}) \qquad (10)\end{aligned}$$

Consider now a multi-product firm with two products j and k that are offered to the same set of consumers, equivalent to the merged entity in the example given in Section 2. The optimal price for product j in this multi-product firm is derived below.

$$\begin{aligned}\Pi_j &= S_j(\boldsymbol{p})(p_j - AC_j(\boldsymbol{p})) + S_k(\boldsymbol{p})(p_k - AC_k(\boldsymbol{p}))\\ 0 &= \frac{\partial S_j}{\partial p_j}(p_j - AC_j) + S_j(1 - \frac{\partial AC_j}{\partial p_j}) + \frac{\partial S_k}{\partial p_j}(p_k - AC_k) - S_k\frac{\partial AC_k}{\partial p_j}\\ p_j &= AC_j + \frac{S_j}{\frac{\partial S_j}{\partial p_j}}(1 - \frac{\partial AC_j}{\partial p_j}) - \frac{\frac{\partial S_k}{\partial p_j}}{\frac{\partial S_j}{\partial p_j}}(p_k - AC_k) + \frac{S_k}{\frac{\partial S_j}{\partial p_j}}\frac{\partial AC_k}{\partial p_j} \qquad (11)\end{aligned}$$

GePP is defined as the difference between the pre-merger and post-merger first order conditions for a particular product's price, both normalized to be quasi-linear in marginal cost (Jaffe and Weyl (2013)). In the example given in Section 2, GePP given by the price defined in Equations 11 less the price defined in Equation (10).

$$GePP_{jk}(\boldsymbol{p}) = -\frac{\frac{\partial S_k(\boldsymbol{p})}{\partial p_j}}{\frac{\partial S_j(\boldsymbol{p})}{\partial p_j}}(p_k - AC_k(\boldsymbol{p})) + \frac{S_k(\boldsymbol{p})}{\frac{\partial S_j(\boldsymbol{p})}{\partial p_j}}\frac{\partial AC_k(\boldsymbol{p})}{\partial p_j}$$

Importantly, GePP is a function of prices. In practice, it is typically evaluated at pre-merger prices.

B.3 Socially Optimal and Constrained Optimal Prices

The social welfare function, $SW(\cdot)$, is given by the sum of consumer surplus and producer profits.

$$SW(\boldsymbol{p}) = \int_i CS_i(\boldsymbol{p})di + \sum_{k\in J} S_k(p_k - AC_k)$$

The socially optimal price for a particular product j is derived below, using the result that $\frac{\partial CS_i}{\partial p_j} = -S_j$.

$$\begin{aligned}
0 &= \int_i \frac{\partial CS_i}{\partial p_j}di + \frac{\partial S_j}{\partial p_j}(p_j - AC_j) + S_j(1 - \frac{\partial AC_j}{\partial p_j}) + \sum_{k\neq j}\frac{\partial S_k}{\partial p_j}(p_k - AC_k) - S_k\frac{\partial AC_j}{\partial p_j} \\
0 &= \frac{\partial S_j}{\partial p_j}(p_j - AC_j) - S_j\frac{\partial AC_j}{\partial p_j} + \sum_{k\neq j}\frac{\partial S_k}{\partial p_j}(p_k - AC_k) - S_k\frac{\partial AC_j}{\partial p_j} \\
p_j^W &= AC_j + \frac{S_j}{\frac{\partial S_j}{\partial p_j}}\frac{\partial AC_j}{\partial p_j} + \left(\sum_{k\neq j}\frac{S_k}{\frac{\partial S_j}{\partial p_j}}\frac{\partial AC_j}{\partial p_j} - \frac{\frac{\partial S_k}{\partial p_j}}{\frac{\partial S_j}{\partial p_j}}(p_k - AC_k)\right) = \int_i \sum_k \frac{\partial S_{ik}}{\partial p_j}c_{ij}di \Big/ \int_i \frac{\partial S_{ij}}{\partial p_j}di
\end{aligned}$$

The problem of a constrained social planner that chooses product-level prices subject to a promise of total profit $\bar{\Pi}$ to the insurance industry is given below.

$$\max_{\{p_j\}_{j\in J}} \int_i CS_i(\boldsymbol{p})di$$
$$\text{such that } \sum_{k\in J} S_k(p_k - AC_k) \geq \overline{\Pi}$$

The constrained optimal price for product j is derived below, where λ is the Lagrange multiplier on the profit constraint.

$$\mathscr{L} = \int_i CS_i(\boldsymbol{p})di + \lambda(\sum_{k\in J} S_k(p_k - AC_k) - \overline{\Pi})$$

$$0 = \int_i \frac{\partial CS_i}{\partial p_j}di + \lambda\left(\frac{\partial S_j}{\partial p_j}(p_j - AC_j) + S_j(1 - \frac{\partial AC_j}{\partial p_j}) + \sum_{k\neq j}\frac{\partial S_k}{\partial p_j}(p_k - AC_k) - S_k\frac{\partial AC_j}{\partial p_j}\right)$$

$$0 = S_j(1 - \frac{1}{\lambda}) + \frac{\partial S_j}{\partial p_j}(p_j - AC_j) - S_j\frac{\partial AC_j}{\partial p_j} + \sum_{k\neq j}\frac{\partial S_k}{\partial p_j}(p_k - AC_k) - S_k\frac{\partial AC_j}{\partial p_j}$$

$$p_j^{CE} = -\frac{\lambda - 1}{\lambda}\frac{S_j}{\frac{\partial S_j}{\partial p_j}} + AC_j + \frac{S_j}{\frac{\partial S_j}{\partial p_j}}\frac{\partial AC_j}{\partial p_j} + \left(\sum_{k\neq j}\frac{S_k}{\frac{\partial S_j}{\partial p_j}}\frac{\partial AC_j}{\partial p_j} - \frac{\frac{\partial S_k}{\partial p_j}}{\frac{\partial S_j}{\partial p_j}}(p_k - AC_k)\right)$$

B.4 Proof of Proposition 2.2

In Example 2.1, it is possible that preferences and costs are such that only one product is purchased in equilibrium, and it is no longer a meaningful example for a potential merger. In any equilibrium in which both products are purchased, it must be that $p_H^* > p_L^*$. Otherwise, the demand for product L is 0.

The marginal consumers between the two products are those that are indifferent between product H and product L. Specifically, the marginal consumers are those with $c_i = \hat{c}$, where

$$\hat{c} = \frac{p_H^* - p_L^*}{(\nu - 1)\beta}$$

Thus that marginal cost, $MC_{LH} = \hat{c}$, and the marginal cost $MC_{HL} = \nu\hat{c}$. Because $\nu > 1$, $MC_{HL} > MC_{LH}$.

All consumers with $c_i > \hat{c}$ prefer product H and all consumers with $c_i < \hat{c}$ prefer product L. Because we have assumed that the mass of consumers that purchase each product is non-zero, there exist other consumers that purchase both H and L with $c_i \neq \hat{c}$. Therefore, the average value of c_i for consumers that purchase H must be strictly greater than $\hat{c}$, and $AC_H > MC_{HL}$. Similarly, the average value of c_i for consumers that purchase L must be strictly less than $\hat{c}$, and $AC_L < MC_{LH}$.

Thus, it must be that $AC_L < MC_{LH} < MC_{HL} < AC_H$. Following Equation (8), this implies that

$\frac{\partial AC_L}{\partial p_H} > 0$ and $\frac{\partial AC_H}{\partial p_L} < 0$.

B.5 Proof of Proposition 2.4

In any symmetric equilibrium, $p_A^* = p_B^* = p^*$. The quantities, average costs, marginal costs, and other features of the equilibrium will be identical across the two products. As such, I will drop the product subscripts in the following proof.

Given the equilibrium prices, let $s(c)$ be the share of consumers with $c_i = c$ that purchase either product A or B. Let $m(c)$ be the share of consumers with $c_i = c$ that both purchase either A or B and are indifferent between the two options. The average cost of both products is given by

$$AC = \frac{1}{S}\int_i s(c_i)c_i di$$
$$S = \int_i s(c_i)di$$

The average cost of consumers on the margin between products A and B is given by

$$MC_{AB} = \frac{1}{M}\int_i m(c_i)c_i di$$
$$M = \int_i m(c_i)di$$

Let $\tilde{s}$ and $\tilde{m}$ be normalized weighting functions that determine the composition of consumers that make up the average cost and marginal cost.

$$\tilde{s}(c) = \frac{s(c)}{S}$$
$$\tilde{m}(c) = \frac{m(c)}{M}$$

Following Equation (8), if $MC_{AB} > AC$ then $\frac{\partial AC_A}{\partial p_B} > 0$ and $\frac{\partial AC_B}{\partial p_A} > 0$.

In order to show that $MC_{AB} > AC$, it is sufficient to demonstrate that $\tilde{s}(c)$ and $\tilde{m}(c)$ satisfy a single crossing property such that there exists some point in the support $\hat{c}$ where for all $c < \hat{c}$, $\tilde{m}(c) < \tilde{s}(c)$ and for all $c > \hat{c}$ $\tilde{m}(c) > \tilde{s}(c)$. In this case, the weighting function $\tilde{m}(c)$ first-order stochastic dominates the weights given by $\tilde{s}(c)$, and by the properties of first-order stochastic dominance, $MC_{AB} > AC$.

The proof relies on demonstrating that $s(c)$ is "more concave" than $m(c)$. To see why this might be true, consider the distribution of idiosyncratic preferences F. Because preferences are i.i.d., the quantity demanded is equal to the share of consumers for whom both idiosyncratic preference terms fall below a threshold value. The share of marginal consumers are those for whom this threshold condition is true and the two draws of ε_A and ε_B are equal. See the expressions below for an unbounded distribution of ε.

$$s(c_i) = 1 - F(p^* - \beta c_i)^2$$
$$s(c_i) = 1 - \left(\int_{-\infty}^{p^* - \beta c_i} f(\varepsilon) d\varepsilon \right)^2$$
$$m(c_i) = 1 - \frac{1}{\Gamma} \int_{-\infty}^{p^* - \beta c_i} (f(\varepsilon))^2 d\varepsilon$$

where $\Gamma = \int_{-\infty}^{\infty} (f(\varepsilon))^2 d\varepsilon$. Very roughly speaking, the share function $s(c)$ follows the concave function, $f(x) = -x^2$, and the marginal function $m(c_i)$ is the expected value of the concave function. The technical details of rigorously establishing the single crossing condition for an arbitrary distribution are non-trivial. Here, I will show that this condition holds for three commonly used distributions for idiosyncratic preferences: uniform, type I extreme value, and normal.

In each of these proofs, I will assume that c has infinite support, which allows me to exploit limiting properties of the distributions (or bounds in the case of uniform distribution.) However, because $\tilde{s}(c)$ and $\tilde{m}(c)$ are strictly increasing functions that always integrate to 1, the crucial single crossing property is maintained for any bounded range of c.

The example also assumes that the distribution of c is uniform. Establishing the single cross condition of $s(c)$ and $m(c)$ does not depend on the distribution of c and is maintained for any continuous distribution. However, some distributions of c may preclude the existence of an equilibrium.

Uniform Distribution

Suppose that the distribution of idiosyncratic preferences is uniform, i.e. $\varepsilon \sim U(l,u)$. In this case,

$$s(c) = 1 - \left(\frac{p^* - \beta c - l}{u - l}\right)^2$$
$$m(c) = F(p^* - \beta c) = \frac{p^* - \beta c - l}{u - l}$$

Because consumers are uniformly distributed, the probability that $\varepsilon_{iA} = \varepsilon_{iB}$ is independent of the value of ε_{iA}. Therefore, the distribution of $F(\varepsilon_{iA}|\varepsilon_{iA} = \varepsilon_{iB}) = F(\varepsilon_{iA})$. Thus the fraction of consumers that are marginal is simply given by the c.d.f. of the preference shocks.

Because both $\tilde{s}(c)$ and $\tilde{m}(c)$ must both integrate to one and are not everywhere equal, these two lines must intersect at least once. Because $\tilde{s}(c)$ is strictly concave for values $c \in (\frac{p^*-h}{\beta}, \frac{p^*-l}{\beta})$, and $\tilde{m}(c)$ is linear, they can intersect only one time.

For every c, $s(c) \geq m(c)$, with a strict inequality whenever $s(c) \in (0,1)$. Therefore, $S > M$. And there exists some large enough c value such that $s(c) = m(c) = 1$. Thus, at this large value of c, it must be that $\tilde{s}(c) < \tilde{m}(c)$. Therefore, $\tilde{m}(c)$ must be strictly greater than $\tilde{s}(c)$ after this crossing point and strictly lesser before it.

Type I Extreme Value

Suppose that the distribution of idiosyncratic preferences is such that $\varepsilon_{iA} = \upsilon_{iA} - \upsilon_{i0}$ and $\varepsilon_{iB} = \upsilon_{iB} - \upsilon_{i0}$, where υ_{iA}, υ_{iB}, and υ_{i0} are all distributed type I extreme value with variance σ^{ε}. In this case,

$$
\begin{aligned}
s(c) &= \frac{e^{\frac{\beta c - p^*}{\sigma^{\varepsilon}}}}{1 + 2e^{\frac{\beta c - p^*}{\sigma^{\varepsilon}}}} \\
\tilde{s}(c) &= \frac{s(c)}{\int s(c_i)di} \\
m(c) &= \frac{1}{\sigma^{\varepsilon}} s(c)^2 \\
\tilde{m}(c) &= \frac{\frac{1}{\sigma^{\varepsilon}} s(c)^2}{\frac{1}{\sigma^{\varepsilon}} \int s(c_i)^2 di} \\
\tilde{m}(c) &= \frac{s(c)^2}{\int s(c_i)^2 di}
\end{aligned}
$$

These functions satisfy three conditions that are sufficient to establish the single crossing property. First, $\lim_{c \to -\infty} s(c) = \lim_{c \to -\infty} m(c) = 0$ and $\lim_{c \to \infty} s(c) = \lim_{c \to \infty} m(c) = 1$. Second, $s(c) > m(c)$ for all c.

Third, the derivatives of $s(c)$ and $m(c)$ satisfy a single crossing property such that there exists a $\bar{c}$ where for all $c < \bar{c}$, $s'(c) > m'(c)$ and for all $c > \bar{c}$, $s'(c) < m'(c)$. Note that $m'(c) = 2s(c)s'(c)$. Because s is a strictly increasing function, this single crossing point is the value $\bar{c}$ where $s(\bar{c}) = \frac{1}{2}$.

Because both $\tilde{m}(c)$ and $\tilde{s}(c)$ must integrate to 1 and are not everywhere equal, these two functions must intersect at least once. Because both s and m are strictly increasing with single crossing derivatives, the intersections must be countable. In other words, the functions only intersect at points rather than overlapping for an interval.

Because $s(c) > m(c)$ for all c, $\int s(c_i)di > \int s(c_i)^2 di$. Because the denominator for $\tilde{s}(c)$ is strictly larger and $\lim_{c \to \infty} s(c) = \lim_{c \to \infty} m(c) = 1$, there is some final crossing point $\hat{c}$ such that for all $c > \hat{c}$, $\tilde{s}(c) < \tilde{m}(c)$.

Suppose for a contradiction that there is another crossing point at some value $c' < \hat{c}$. It must be that s is less than m for $c < c'$ greater than m for $c > c'$ for values of c in the neighborhood of c'. Therefore, $\frac{\partial \tilde{s}(c')}{\partial c} > \frac{\partial \tilde{m}(c')}{\partial c}$. Because $\int s(c_i)di > \int s(c_i)^2 di$, it must also be that $\frac{\partial s(c')}{\partial c} > \frac{\partial m(c')}{\partial c}$. Because of the single crossing condition for these derivatives, it must be that for all values of $c < c'$, $\frac{\partial \tilde{s}(c')}{\partial c} > \frac{\partial \tilde{m}(c')}{\partial c}$. Thus, for values of $c < c'$, $\tilde{s}(c') < \tilde{m}(c')$, and because of the derivative inequality, these two functions are diverging as c goes to negative infinity. However, $\lim_{c\to-\infty} s(c) = \lim_{c\to-\infty} m(c) = 0$ and thus $\lim_{c\to-\infty} \tilde{s}(c) = \lim_{c\to-\infty} \tilde{m}(c) = 0$. This is a contradiction. Therefore, we know that these functions must cross only once at $\hat{c}$. Moreover, $\tilde{m}(c)$ must be strictly greater than $\tilde{s}(c)$ after this crossing point and strictly lesser before it.

Normal Distribution

Suppose that the distribution of idiosyncratic preferences are normal, i.e. $\varepsilon \sim N(\mu, \sigma^{\varepsilon})$. In this case,

$$
\begin{aligned}
s(c) &= 1 - \Phi\left(\frac{p^* - \beta c}{\sigma^{\varepsilon}} - \mu\right) \\
m(c) &= 1 - \Phi\left(\sqrt{2}\left(\frac{p^* - \beta c}{\sigma^{\varepsilon}} - \mu\right)\right)
\end{aligned}
$$

That the marginal consumers follow a distribution scaled by the $\sqrt{2}$ can be seen from the square of the normal p.d.f.

$$
\begin{aligned}
\phi(x)^2 &= \left(\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}\right)^2 \\
&= \frac{e^{-x^2}}{2\pi} \\
&= \frac{e^{\frac{-(\sqrt{2}x)^2}{2}}}{2\pi} \\
&= \frac{1}{\sqrt{2\pi}}\phi\left(\sqrt{2}x\right)
\end{aligned}
$$

The extra $\frac{1}{\sqrt{2\pi}}$ will drop out when normalizing to the weighting function $\tilde{m}$, so we can drop it here.

The functions $s(c)$ and $m(c)$ satisfy the same three conditions as those generated by the type I extreme value preferences. First, the limiting values of both s and m are 0 as c approaches negative infinity and 1 as c approaches positive infinity. Second, $s(c) > m(c)$ for all c. This second condition is not generally true for an arbitrary scaling factor, but holds for $\sqrt{2}$.

Third, the derivatives of $s(c)$ and $m(c)$ satisfy a single crossing property such that there exists a $\bar{c}$ where for all $c < \hat{c},\ s'(c) > m'(c)$ and for all $c > \hat{c},\ s'(c) < m'(c)$. The derivatives are given by:

$$
\begin{aligned}
s'(c) &= \frac{2\beta}{\sigma^{\varepsilon}} 2\Phi\left(\frac{p^* - \beta c}{\sigma^{\varepsilon}} - \mu\right)\phi\left(\frac{p^* - \beta c}{\sigma^{\varepsilon}} - \mu\right) \\
m'(c) &= \frac{\sqrt{2}\beta}{\sigma^{\varepsilon}}\phi\left(\sqrt{2}\left(\frac{p^* - \beta c}{\sigma^{\varepsilon}} - \mu\right)\right)
\end{aligned}
$$

These are known parametric functions that meet this single crossing condition. For intuition, it is easy to verify that $2\Phi(x)\phi(x)$ and the standard normal p.d.f., $\phi(x)$, satisfy this condition by a similar argument to that used in the type I extreme value proof above. Consider all potential scaling factors d for $d\phi(dx)$. For small values of d, the p.d.f. is very disperse—large in the tails and small at the mean—and it intersects $2\Phi(x)\phi(x)$ twice. For large values of d, the p.d.f. is very concentrated—vanishing in the tails and large at the mean—and it also intersects $2\Phi(x)\phi(x)$

twice, but from the opposite directions. For values of d sufficiently close to 1—including $\sqrt{2}$—the two curves satisfy the single crossing property. Note that β, σ^{ε}, and μ, are all scaling parameters that apply equally and will not alter this relationship.

The proof then proceeds in the same manner as the previous proof for the type I extreme value distribution.

B.6 Simulation Parameters

The illustrative model outlined in Section 2 can be governed four parameters: $(\delta_0, \beta, \nu, \sigma^c)$. For the parameter δ_0, I calibrate it to target an initial market share $\bar{S} = S_{AL} + S_{AH} = S_{BL} + S_{BH}$. For the parameter, σ^c, I calibrate it to match a value of a skewness measure of the cost distribution, skew $= \frac{P_{90} - P_{50}}{P_{50} - P_{10}}$.

Table A3 displays the range of targeted values for these parameters. The final column of the table also shows the fixed value of the parameter in all panels of Figure 1 where that parameter is held constant.

Table A3: Data Description

Parameter	Minimum Value	Maximum Value	Fixed Value
$\bar{S}$	0.05	0.4	0.2
β	0	1.2	0.5
ν	0	0.6	0.2
skew	1	6	-

Note: This table displays the range of parameters used to simulate the mergers studied in Section 2

In order to generate Figure 1, I hold fixed two of the three parameters (δ_0, β, ν) fixed and simulate a merger for every value of the third parameter along a fine grid between the minimum and maximum value shown in Table A3. In each of the three panels, I repeat this for four values of skewness: 1, 1.5, 3, and 6.

In order to generate Figure 2, I simulate 1000 mergers by drawing parameters in a Monte Carlo simulation. I each simulation, I draw each of the four parameters from independent uniform distributions bounded between the minima and maxima listed in Table A3.

In each simulation, I solve for the symmetric equilibrium for both pre-merger and post-merger. Because of selection, some parameter values could potentially generate asymmetric equilibria, but I ignore this possibility. A symmetric equilibrium with positive market shares exists for all parameter values considered. I use 100 Gaussian quadrature points to integrate over the distribution of c.

In these simulations, I must also set a fifth parameter, σ^{ε}, which governs the variance of the idiosyncratic preferences. Because the distribution of c affects demand and is fixed, this normalization is non-trivial. I set $\sigma^{\varepsilon} = 0.125$. In the Monte Carlo simulations (explained below), this value of σ^{ε} produces reasonable average own-price elasticities that range from -3.9 to -9.2, with a mean value of -5.6. Smaller values of σ^{ε} lead to greater elasticities and make it more likely that mergers will benefit consumers. Greater elasticities lead to lower equilibrium prices and more potentially unprofitable consumers. Greater values of σ^{ε} lead to lower elasticities and have the opposite effect.

C Data Processing

C.1 Processing the Choice Data

The choice data contain only the ultimate choices made by the consumers, not the scope of available options. In order to construct choice sets, I use the HIX 2.0 data set compiled by the Robert Wood Johnson Foundation. This data set provides detailed cost-sharing and premium information on plans offered in the non-group market in 2015. The data set is nearly a complete depiction of the market for the entire United States, but there are some markets in which some cost-sharing information is missing, or insurance firms are absent altogether.

I restrict the analysis to markets in which I observe characteristics of the entire choice set and can be reasonably confident that the private marketplace presents nearly the complete choice set of health insurers. Using state-level market shares from the Medical Loss Ratio reporting data, I throw out any markets in which I do not observe any purchases from insurance firms that have more than 5% market share in the state. In this way, I hope to ensure that my sample of choices is not segmented to only a portion of the market.

In Table A4, I summarize the data sample used in estimation and compare it to other data on the non-group insurance market: the ACS and data reported by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) at the U.S. Department of Health and Human Services. The ACS survey design offers the broadest depiction of the market across all market segments. ASPE publishes detailed descriptive statistics on purchases made through the federally-facilitated HealthCare.gov. Relative to the ACS, enrollment through HealthCare.gov is weighted heavily towards low-income, subsidy-eligible consumers. As a result, the plan type market shares reported by ASPE are weighted heavily towards Silver plans that have extra cost-sharing benefits at low incomes. Although the private marketplace is tilted towards higher-income and younger households, the ACS weighting moves the demographic distributions and market shares closer to those in the other data sources. Ryan et al. (2022) investigate these relationships in more detail and show that the market shares, conditional on income and geography, are quite close to those reported by ASPE.

Choice Sets

A household's choice set depends on the age composition of its members and the household income. Because I observe only one age of the household, I use a simple rule to impute the age composition: any household with more than one individual contains two adults of the same age and additional persons are under the age of 21. The data contain information on the premium paid for a subset of the observations. In combination with the base premium of the purchased product, the premium paid can be used to impute household composition. Using the median base premium in the selected firm and metal-level, I construct an imputed household age-rating measure. The correlation between this imputation and the more simple age-rating rule applied to the rest of the sample is 0.90. The results are robust to alternative assumptions about age rating.

For subsidized consumers, income can be imputed from the observed subsidy value and the

Table A4: Data Description

	ACS	ASPE	Private Marketplace Un-weighted	Private Marketplace Weighted
		Age Distribution		
Under 18	0.0%	9.0%	0.0%	0.0%
18 to 25	7.6%	11.3%	11.1%	11.4%
26 to 34	17.2%	17.5%	30.8%	29.1%
35 to 44	22.2%	16.8%	21.4%	20.1%
45 to 54	25.3%	20.9%	19.9%	20.5%
55 to 64	27.7%	23.3%	16.8%	19.0%
		Income Distribution		
Under 250% FPL	32.1%	76.1%	30.8%	43.0%
250% to 400% FPL	24.5%	15.4%	9.1%	13.4%
Over 400% FPL	43.4%	8.5%	60.1%	43.6%
		Metal Level Market Shares		
Catastrophic		1.1%	5.0%	3.6%
Bronze		24.2%	39.2%	36.0%
Silver		66.4%	41.8%	48.8%
Gold		6.6%	11.1%	9.4%
Platinum		1.7%	2.9%	2.2%

Note: The table compares the weighted and unweighted distribution of consumers in the estimation data sample relative to other data sources on the non-group market. The age distributions reported are for the head-of-household with the exception of ASPE, which is the individual-level distribution.

household size. I use this imputed income for subsidized consumers with missing information. However, doing so is not possible for the consumers that do not receive a subsidy. I assume that those in the data without a reported subsidy amount have an income greater than the subsidy qualification threshold.

The choice set in each market is large. The typical market has about 150 plans to choose from, and these plans do not necessarily overlap with other markets. Because I observe only a sample of choices, there are many plans that I do not observe being chosen. The lack of observed choices does not necessarily imply that these plans have a zero market share and may be due to the fact that the number of options is large relative to the observed number of choices. The median number of choices per market is 300.

To simplify this problem, I aggregate to the level of firm-metal offerings in a particular market. For example, all Bronze plans offered by a single insurance firm are considered a single product. Firms typically offer more than one plan in a given metal level. The median number of plan offerings per metal level is three, and the 75^{th} percentile is five. Wherever there is more than one plan per category, I aggregate by using the median premium within the category. The only other product attributes I use in estimation are common to all plans in each category.

C.2 American Community Survey

Data on the size and demographic distributions of both the uninsured and insured populations in each market come from 2015 American Community Survey (ACS). The population of individuals who might consider purchasing non-group health insurance is any legal US resident that is not eligible for Medicaid, Medicare, and is not enrolled in health insurance through their employer. An individual that is not enrolled in employer sponsored insurance but has an offer that they chose not to accept is assumed to be in the non-group market. These consumers may be ineligible for subsidies but can often obtain waivers to get the same treatment as those without an employer offer. This population is small (Planalp et al. (2015)), and I treat them identically to the rest of the non-group market.

In order to address under-reporting of Medicaid enrollment, any parent that receives public assistance, any child of a parent that receives public assistance or is enrolled in Medicaid, any spouse of an adult that receives public assistance or is enrolled in Medicaid or any childless or unemployed adult that receives Supplemental Security Income payments are assumed to be enrolled in Medicaid. Besides Medicaid and CHIP enrollment, an individual is considered eligible for either program if his or her household income falls within state-specific eligibility levels. If an individual is determined to be eligible for Medicaid through these means but reports to be enrolled in private coverage, either non-group coverage or through an employer, they are assumed to be enrolled in Medicaid. This accounts for those that confuse Medicaid managed care programs with private coverage, and Medicaid employer insurance assistance.

This article follows the Government Accountability Office methods (GAO (2012)) to construct health insurance households. This method first divides households as identified in the survey data into tax filers and tax dependents, linking tax dependents to particular tax filers. A tax filing household, characterized by the single filer or joint filers and their dependents, is generally considered to be a health insurance purchasing unit. In some cases, certain members of a tax household will have insurance coverage through another source, e.g. an employer or federal program. In this case, the health insurance household is the subset of the tax household that must purchase insurance on the non-group market.

C.3 Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS) is a nationally representative household survey on demographics, insurance status, and health care utilization and expenditures. MEPS provides moments on the distribution of risk scores in the insured population and the relative costs of households by the age and risk score of the head of household and the risk. All moments are constructed using all surveyed households with a head of household under the age of 65.

The 2015 Medical Conditions File (MCF) of MEPS contains self-reported diagnosis codes. The publicly available data only list 3-digit diagnosis codes, rather than the full 5-digit codes. I follow McGuire et al. (2014) and assign the smallest 5-digit code for the purpose of constructing the condition categories. For example, I treat a 3-digit code of '571' as '571.00'. This implies that many conditions in the hierarchical risk prediction framework are censored. However McGuire et al. (2014) find that moving from 5-digit codes to 3-digit codes does not have a large effect on the predictive implications for risk scores.

I link the MCF to the Full Year Consolidated File to identify the age and sex of the individual, and then apply the 2015 HHS-HCC risk prediction methodology (Kautter et al. (2014a)). The risk coefficients are published by CMS and publicly available.

C.4 Medical Loss Ratio Data

CMS makes publicly available the state-level financial details of insurance firms in the non-group market for the purpose of regulating the MLR.[1] This information includes the number of member-months covered by the insurance firm in the state and total costs.

This article uses two pieces of information from the Medical Loss Ratio filings: average cost and average risk adjustment transfers.

Firms are defined by operating groups at the state level. Some firms submit several medical loss ratio filings under for different subsidiaries in a given state. I group these filings together.

Average cost is defined as total non-group insurance claims divided by total non-group member months, current as of the first quarter of 2016. This computation includes claims and member months that may not be a part of the non-group market as it is characterized in this analysis. For instance, grandfathered insurance plans that are no longer sold to new consumers are included. These are likely to be a small portion of the overall market.

To compute the average risk adjustment payment, some adjustment to the qualifying member months is required. Unlike medical claims, grandfathered plans (and other similar non-ACA compliant plans) are not included in the risk adjustment system. Dividing the total risk adjustment transfer by the total member months will bias the average transfer towards zero.

The interim risk adjustment report published by CMS includes the total member months for every state. And the MLR filings separately list the risk-corridor eligible member months, which are a subset of the risk adjustment eligible member months. I define "potentially non-compliant" member months as the difference between risk-corridor eligible member months and total member months. I scale the potentially non-compliant member months of all firms in each state proportionally so that total member months is equal to the value published by CMS, with two exceptions. First, firms that opted not to participate in the ACA exchange in that state have zero risk-corridor eligible member months. I do not reduce the member months of these firms, as I

[1] Insurance firms in this market are restricted in how much premium revenue they may collect, relative to an adjusted measure of medical costs. In 2015, this constraint is not often binding. Excess revenue is returned to consumers via a rebate.

cannot isolate the potentially non-compliant months. Second, if the risk-corridor eligible member months exceed the total member months published by CMS, I assume that the risk-corridor eligible member months are exactly equal to the risk adjustment eligible member months.

Computing Firm-level Risk

This article firm-level risk transfers to infer the equilibrium distribution of risk across firms. With a bit of simplification, the ACA risk transfer formula at the firm level can be written as

$$T_f = \left[\frac{\bar{R}_f}{\sum_{f'} S_{f'} \bar{R}_{f'}} - \frac{\bar{A}_f}{\sum_{f'} S_{f'} \bar{A}_{f'}} \right] \bar{P}_s$$

where $\bar{R}_f$ is the firm level of average risk and $\bar{A}_f$ is the firm level average age rating, where the average is computed across all the firms products and weighted by members, a geographic adjustment, and a metal-level adjustment. S_f is the firm's state-level inside market share, and $\bar{P}_s$ is the average total premium charged in the state.

Every element of this formula is data available in the Interim Risk Adjustment Report on the 2015 plan year, except for the plan-level market shares, the plan-level average age rating, and the plan-level average risk. As a simplification, I assume that the average age rating is constant across all firms, and that the weighting parameters in the risk component are negligible. In reality, variation in the average age rating is not very large, and incorporating this variation in the moment matching dramatically increases the computational burden.

I compute the implied firm-level average risk as

$$\bar{R}_f = \left(\frac{T_f}{\bar{P}_s} + 1 \right) \bar{R}$$

where the risk transfer T_f is the average firm-level risk adjustment transfer from MLR data, $\bar{P}_s$ is the average state level premium reported in the interim risk adjustment report, and $\bar{R}$ is the national average risk score reported in the interim risk adjustment report.[2] In the estimation, I target the difference between $\bar{R}_f$ and an adjusted average risk score for the state that accounts for grandfathered insurance products not sold through the marketplace.

C.5 Rate Filing Data

The Center for Medicare and Medicaid Services (CMS) tabulates the Premium Rate Filings that insurance firms must submit to state insurance regulators if they intend to increase the premiums for products they will continue to offer. In these filings, insurance firms include information on the cost and revenue experience of the insurance product in the prior year and projections for the following year.

The rate filing data are divided into two files—a firm-level worksheet and a plan-level worksheet—and contain information on the prior year experience of the plan and the projected experience of the plan in the coming year. I use projected firm-level average cost and the average ratio of experienced costs across metal levels for all firms. Using projected average costs for the firms leads to the best fit for the first order conditions, which are not imposed in estimation. This may be because it more accurately represents firms' expectations when setting their costs. Although the decision to use projected or experienced costs does affect the marginal cost estimation, it does not qualitatively impact the results.

To construct moments on the ratio of average cost across metal level categories, I use the prior year experience submitted in the 2016 rate filings data. To recover the average cost after reinsurance, I subtract the experienced total allowable claims that are not the issuer's obligation and the experienced risk adjustment payments from the total allowable claims.

The ratio of average cost across each metal level category is computed as the weighted average of every within firm ratio. I compute the average cost across all plans within each metal level category in each firm, and then compute the weighted average of the ratios across each firm.

[2] The formula implies that the state average risk score should go in place of the national average. However, I do not allow the risk distribution among consumers to vary by geography (other than through composition). I use the national risk score to abstract from these geographical differences.

Each step is weighted using the reported experienced member months. The model moments are constructed in the same manner.

To estimate firm average costs, this article takes advantage of the firm's projected costs for the 2015 plan year. I use the projected firm level average cost from the 2015 plan year firm-level rate filing data. I compute post-reinsurance projected costs by subtracting projected reinsurance payments from "projected incurred claims, before ACA Reinsurance and Risk Adjustment."

Some firms do not appear in the risk filing data. For these firms, I compute the projected average cost for those firms by adjusting the experienced average cost reported in the Medical Loss Ratio filings by the average ratio of projected to experienced claims. In 2015, the average ratio of project to experienced claims for firms in my sample is 71.5%.

D Demand Estimation

The demand model closely follows that of Tebaldi (2024), with the difference that cost heterogeneity is modeled through this calibrated random coefficient on risk score rather than directly associated to the willingness to pay for insurance. This specification allows me to incorporate more data on the risk distribution of consumers to match heterogeneity in risk preferences across firms—via the firm fixed effect—in addition to the amount of insurance. In this section, I detail the calibration of the risk score distribution and the demand estimation methodology.

D.1 Risk Score Distribution

The risk scores in the demand model correspond to the output of the Health and Human Services Hierarchical Condition Categories risk adjustment model (HHS-HCC) used in the non-group market for the purpose of administering risk adjustment transfers. The HHS-HCC risk adjustment model is designed to predict expected plan spending on an individual based on demographics and health condition diagnoses. It is the result of a linear regression of relative plan spending on a set of age-sex categories and a set of hierarchical condition categories derived from diagnosis codes.

$$\frac{\text{Plan Spending}_{it}}{\text{Avg. Plan Spending}_t} = \gamma_0 + \sum_g \gamma_{tg}^{age,sex} Age_{ig} Male_{ig} + \sum_{g'} \gamma_{tg'}^{HCC} HCC_{ig'} + \eta_{it}$$

The prediction regressions are performed separately for different types of plans t, where t represents the metal category of the plan. The resulting risk score for an individual is a normalized predicted relative-spending value. Because all regressors take a value of either 1 or 0, the risk score is equal to the sum of all coefficients that apply to a particular individual.

$$r_{it} = \underbrace{\sum_g \gamma_{tg}^{age,sex} Age_g Male_g}_{r_{it}^{dem}} + \underbrace{\sum_{g'} \gamma_{tg'}^{HCC} HCC_{g'}}_{r_{it}^{HCC}}$$

Unless specifically noted, r_i^{HCC} will refer to the Silver plan HCC risk-score component and represent standard a measure of health status across all product types.

Parametric Distribution

The distribution of risk scores, $\hat{G}$, is estimated from the 2015 Medical Conditions File (MCF) of the Medical Expenditure Panel Survey. The MCF contains self-reported diagnosis codes and can be linked to demographic information in the Population Characteristics file. The publicly available data only list three-digit diagnosis codes, rather than the full five-digit codes. I follow McGuire et al. (2014) and assign the smallest five-digit code for the purpose of constructing the condition categories and matching the HHS-HCC risk coefficient.[3] See Appendix Section C.3 for detail on processing the data.

In the data, a majority of individuals have no relevant diagnoses, i.e., $r_i^{HCC} = 0$.In order to match this feature of the data, the distribution combines a discrete probability that an individual has a non-zero risk score and a continuous distribution of positive risk scores. With some probability

[3]For example, I treat a three-digit code of '301' as '301.00'. McGuire et al. (2014) find that moving from five-digit codes to three-digit codes does not have a large effect on the predictive implications for risk score estimation. In this case, there is measurement error as the model used was originally estimated on 5-digit codes.

$\delta(Z_i)$, the household has a non-zero risk score drawn from a log-normal distribution, i.e., $r_i^{HCC} \sim$ Lognormal$(\mu(Z_i), \sigma^r)$. With probability $1-\delta(Z_i)$, $r_i^{HCC} = 0$. I allow the probability of having any relevant diagnoses and the mean of the log-normal distribution to vary by two age categories for the head of household and two income categories—above and below 45 years old, and above and below 400 percent of the federal poverty level.

Table A5 displays the moments of the risk score distributions for each metal level in the data. Figure A3 compares the risk distribution in the MCF with the simulated risk distribution in the estimation sample.

Table A5: Parametric Distribution of Risk Scores

Age	Income		Bronze		Silver		Gold		Platinum	
	(% of FPL)	$\delta(Z_i)$	$\mu(Z_i)$	σ^r	$\mu(Z_i)$	σ^r	$\mu(Z_i)$	σ^r	$\mu(Z_i)$	σ^r
$\leq$45	$\leq$400%	0.15	2.86	19.7	2.99	19.5	3.11	19.8	3.31	20.8
	>400%	0.13	3.02	19.7	3.22	19.5	3.22	19.8	3.40	20.8
>45	$\leq$400%	0.31	3.49	19.7	3.73	19.5	3.73	19.8	3.97	20.8
	>400%	0.24	3.25	19.7	3.46	19.5	3.46	19.8	4.67	20.8

Note: This table displays three aspects of the distribution of HHS-HCC risk scores in the 2015 Medical Conditions File of the MEPS. The first column displays the portion of risk scores that are positive for four categories divided by age and income. The next columns display the mean and variance for each metal-level specific risk score. The mean depends on these same demographic groups, and the variance is calculated across the whole population.

D.2 Estimation and Identification

The demand model two primary identification concerns. First, a plan premium's price may be correlated with the unobserved quality ξ_{jm}, leading to biased estimates of α_i. In this environment, the premium regulations provide a source of variation in price, which is exogenous to variation in unobserved quality. The age-adjustment on premium, a_i, increases monotonically and non-linearly with age, and strictly increases with every age after 25. Income-based subsidies are available to households that earn below 400 percent of the federal poverty level. These subsidies decline continuously within the subsidy-eligible range. Moreover, the choice-set changes discretely at 250 percent, 200 percent, and 150 percent of the federal poverty level. At each of these income thresholds, the Silver plan becomes significantly more generous with no discontinuous change in

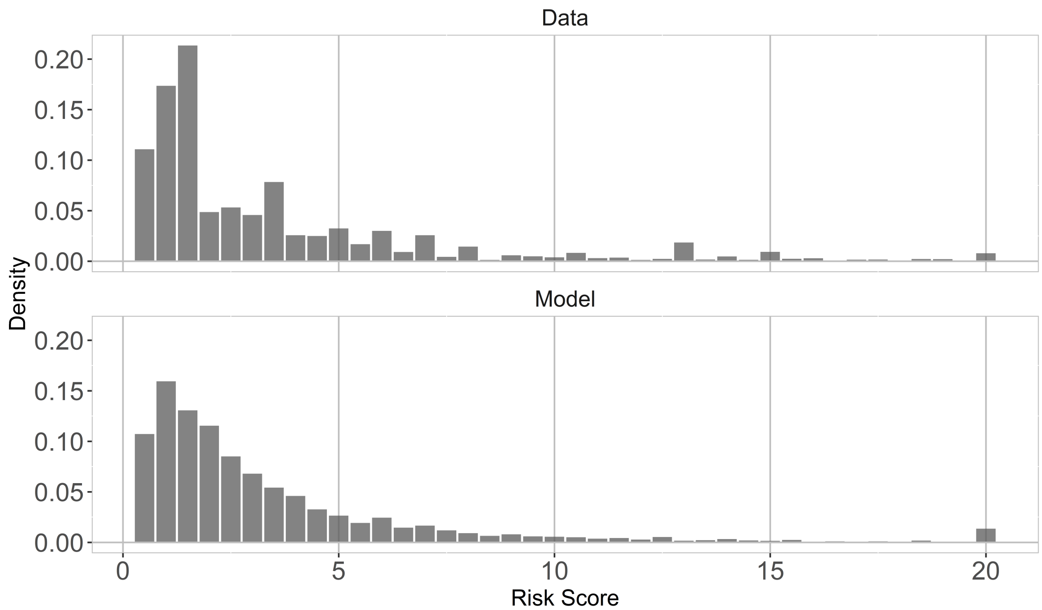

Figure A3: Risk Score Distribution Model Fit

Note: The data distribution comes from applying the HHS-HCC risk prediction methodology to the distribution of self-reported diagnoses in the 2015 Medical Conditions File of the MEPS. The model distribution comes from predicting the distribution of risk scores in the same MEPS sample. In both cases, the distribution of positive Silver metal-level risk scores are displayed.

the price (Lavetti et al. (2023)).

However, it is possible that preferences are also changing with age and income in a way that is hard to disentangle with the regulatory identification. I follow Tebaldi (2024) by augmenting this source of identification with variation in premiums due to local demographics (Waldfogel (2003)). The intuition is that areas with higher proportions of young consumers will have lower average costs and lower premiums. Conditional on the age of a consumer, the age of her neighbors does not affect her demand for insurance.

I implement the control function approach from Tebaldi (2024). In the first stage, I regress the base premium set for each product p_{jm} on the share of consumers in that market that are under the age of 35, and fixed effects for the firm, market, and metal-level of the product.

$$p_{jm} = \text{Share Under } 35_m + \gamma_{f(j)} + \gamma_m + \gamma_{metal} + \zeta_{jm}$$

The results of the first stage are presented in Appendix Table A1. I use the residual, $\widehat{\zeta_{jm}}$, in demand estimation as a control function. Specifically, I allow demand to depend on a 3rd order polynomial of the residual, with each term of the polynomial interacted with age group, as well as firm-market fixed effects. Appendix Table A2 includes demand estimates without the control function, which are qualitatively similar.

The second concern is the identification of the risk coefficients, $(\gamma_r, \{\beta_r^k\})$. These parameters are incorporated into the estimation equations in the same manner as variance parameters for distributions of unobserved consumer preferences (e.g. Berry et al. (1995)). However, because I have data on the distribution of risk scores in the market and moments on the average risk scores of individuals that choose certain products, I am able to incorporate these product-level moments to ensure that the model captures the appropriate risk-related substitution patterns and improve identification (Petrin (2002)).

The demand model targets eighty nine moments on the distribution of consumer risk scores: the average risk score of all insured consumers; the average risk score of enrollees in the Bronze, Silver, Gold, and Platinum plan categories; and the average risk score of each firm relative to the average risk score in the state for each firm-state combination in the data. Let l index the moments, let n index the $N = 500$ draws from the unobserved distribution of risk scores, and let $I(j)$ be the set of consumers that have product j in their choice set. For each group of products, J_l, I compute the moments as

$$M_l = \frac{\sum_{j \in J_l} \sum_{i \in I(j)} \sum_{n=1}^{N} w_i S_{inj} r_{inj}}{\sum_{j \in J_l} \sum_{i \in I(j)} \sum_{n=1}^{500} w_i S_{inj}} - R_m^{data}$$

where r_{inj} is a product-specific risk score draw to match the definition of the moments in the data and w_i is the weight consumer i (see Section 3 for more details on weighting).

To estimate the demand model, I follow Grieco et al. (2025) to combine a micro-data log-likelihood function with product-level GMM moments. The parameters maximize the sum of the log-likelihood of observed choices less the weighted moment objective value.

$$\hat{\theta} \in \text{argmax}_{\theta} \sum_{i} \sum_{j} Y_{ij} \log(\frac{1}{N} \sum_{n} S_{inj}) - M'WM \tag{12}$$

The estimation proceeds in two steps. First, I use the identity matrix as the weighting matrix, W. Second, I set the diagonal of the weighting matrix equal to the inverse of the moment variances evaluated at the parameters estimated in the first stage. Because the moments do not apply to all consumers in the data, I cannot directly compute the moment variances. Instead, I follow Petrin (2002) by computing the variance of a separate set of moments that can be used to construct the intended moments for estimation. In this case, the predicted choice probabilities, S_{ij}, and the average predicted risk score for each product, $\frac{1}{N}\sum_{n} S_{inj} r_{inj}$ are sufficient. The variance of the targeted moments can then be computed using the delta method.

This estimation procedure is analogous to a GMM estimation that uses the first order conditions of the likelihood function as moments (Grieco et al. (2025)). Using the likelihood function in place of an additional set of moments allows the estimation to maintain the desirable convergence and identification properties of maximum likelihood estimation. However, to compute standard errors, I exploit the analogous GMM framework and compute the typical GMM standard errors where the weighting matrix is a block diagonal matrix with the Hessian of the likelihood function in one block and the moment weighting matrix W in the other.

E Cost Estimation

The method of simulated moments estimation procedure targets four sets of moments which each identify four sets of parameters. The age and risk parameters are identified using moments from the Medical Expenditure Panel Survey (Appendix Section C.3). For clear identification of costs by age separate from risk score, the estimation targets age moments among adults that have a risk score of zero. The moments are computed as the ratio of average covered expenditures within five-year age brackets for adults between 25 and 64 years old to the average covered expenditures of adults between 20 and 24 years old. The cost parameter on the risk score is identified using the ratio of average covered expenditures among adults with a positive risk score to those with a risk score of zero. This helps to separate sorting-related costs from firm-specific or product-specific costs. The parameter on actuarial value is identified using the ratio of experienced cost of each metal level to Bronze plans from the 2016 rate-filing data.

Conditional on these three cost parameters (ϕ_{AV}, ϕ_{Age}, ϕ_r), the firm-specific cost parameter, ϕ_f, is set to exactly match the projected average cost in the 2015 rate-filing data. See Appendix Section C.5 for more detail on the data.

When simulating moments that match data from the insurance firm rate filings, I use the reinsurance adjusted cost, c_{ijm}^{rein}. The moments from the Medical Expenditure Panel Survey are computed using total covered expenses across all insured individuals. Thus, I use the predicted cost c_{ijm} to compute these moments.

Cost is estimated using two-stage MSM to obtain the efficient weighting matrix. The estimated demand parameters are used to simulate the distribution of consumer age and risk scores throughout products in each market, using ACS data as the population of possible consumers (see Appendix Section C.2).

E.1 Identifying Assumption

The identifying assumption is that any unobserved cost variation is orthogonal to the idiosyncratic demand shocks. Given the specification of the demand function, this implies that the only mechanisms through which cost and preferences are correlated are through age and risk scores. If this assumption is violated and the remaining correlation is consistent with adverse selection, then the coefficient on actuarial value will be biased upward. For illustration, suppose I estimate $\hat{\phi}$ to solve for a single product and single observable type,

$$\frac{E[S_{ij}c_{ij}]}{S_j} - AC^{data} = 0$$
$$E[S_{ij}c_{ij}] = S_j AC^{data}.$$

This is equivalent to

$$S_j E[c_{ij}] - \text{cov}(S_i, c_{ij}) = S_j AC^{data}.$$

I assume that, conditional on age and risk score, this covariance term is 0. If there is an endogeneity problem consistent with adverse selection, this covariance term would be positive and increasing in plan generosity, leading to an upward bias in the estimated coefficient on adverse selection.

An alternative specification could treat expected total medical spending as a household characteristic. Then, I could allow preferences to vary with this characteristic instead of risk scores. Doing so has the advantage of circumventing this particular exogeneity assumption, but the principle concern that residual costs unobservable to the econometrician are correlated with demand errors would remain.

E.2 Reinsurance

In 2015, the ACA implemented a transitional reinsurance program, which mitigates a portion of the liability to insurance firms of very-high-cost enrollees. This policy was important in limiting the amount of realized adverse selection facing insurance firms and is included in cost estimation in order to match the post-reinsurance average firm costs. The federal government covered 45% of an insurance firm's annual liabilities for a particular individual that exceeded an attachment point, $\underline{c} = \$45,000$, and up to a cap, $\bar{c} = \$250,000$. For an individual with a cost c_{ijm}, the insurance firm is liable for the cost c_{ijm}^{rein} under the reinsurance policy.

$$c_{ijm}^{cov} = \min\left(\max(c_{ijm} - \underline{c}, 0), \bar{c} - \underline{c}\right)$$
$$c_{ijm}^{exc} = \max(c_{ijm} - \bar{c}, 0)$$
$$c_{ijm}^{rein} = \min(c_{ijm}, \underline{c}) + 0.45 c_{ijm}^{cov} + c_{ijm}^{exc}$$

E.3 Matching Firm Moments

Let $\bar{C}_f^{obs}$ be the observed projected firm-level average cost. The firm-specific cost parameters, $\tilde{\psi}(\phi)$, can be set such that these moments are matched exactly. Without incorporating reinsurance, $\tilde{\psi}(\phi)$ can be computed analytically.

$$\bar{C}_f^{obs} = e^{\psi_f} \frac{1}{\sum_{j \in J^f} S_j} \sum_{j \in J^f} \int_i S_{ij} e^{\phi_1 AV_{jm} + \phi_2 Age_i + \phi_3 r_i^{HCC}} dF(i)$$
$$\tilde{\psi}_f(\phi) = \log\left(\frac{1}{\sum_{j \in J^f} S_j} \sum_{j \in J^f} \int_i S_{ij} e^{\phi_1 AV_{jm} + \phi_2 Age_i + \phi_3 r_i^{HCC}} dF(i)\right) - \log(\bar{C}_f^{obs})$$

When incorporating reinsurance, the parameters ψ can no longer be separated from ϕ because they interact in determining how much reinsurance an individual receives. Instead, $\tilde{\psi}$ can be found by iteration.

$$\tilde{\psi}_f^{n+1} = \tilde{\psi}_f^n + \left[\log\left(\frac{1}{\sum_{j \in J^f} S_j} \sum_{j \in J^f} \int_i S_{ij} c_{ijm}^{rein}(\psi_f, \phi) dF(i)\right) - \log(\bar{C}_f^{obs})\right]$$

Without any reinsurance, this iteration method gives the analytic result at $n = 1$ given any feasible starting point, ψ^0. The reinsurance payments are not particularly sensitive to ψ which affects average payments and have less effect on the tails targeted by reinsurance. As a result, $\tilde{\psi}$ can be precisely computed with a small number of iterations.

E.4 Method of Simulated Moments

I will write the moments as $d(\phi)$ to represent the remaining moments on the cost ratios by metal level, age, and risk, incorporating the predicted parameters of $\tilde{\psi}(\phi)$. $\hat{\phi}$ is estimated by minimizing, for a weighting matrix W,

$$\hat{\phi} = \text{argmin}_{\phi}\ d(\phi)' W d(\phi)$$

The minimum of the function is found using the Neldermead method. I estimate $\hat{\phi}$ in two stages. In the first stage, I use the identity weighting matrix and obtain estimates of the variance of the moments, V. In the second stage, I use $W = V^{-1}$. Similar to the demand estimation, the moments do not necessarily apply to every observation of the data. I use the same procedure from Petrin (2002) to compute the variance of the moments.

E.5 Model Fit

Table A6 presents the targeted and estimated moments used in the cost estimation. The age and risk moments are matched more closely than the metal-level ratio moments. In particular, the cost specification leads to overestimates of the cost of covering individuals with Platinum coverage. The combination of ordered risk preferences, age preferences, and log-linear costs in actuarial value lead to the implication that the difference in average costs among expensive and generous plans (Gold and Platinum) is much greater than the difference in average cost among the less comprehensive options (Silver and Bronze).

In estimating the parameters of demand and marginal cost, I do not use the assumption that firms are optimally setting Nash-Bertrand prices to maximize profit. This approach allows the demand and cost parameters to be identified from data, and shielded from potential model mis-specification. In general, the demand model implies greater markups than the cost estimation. The median and mean implied markup from the demand estimation is 43 and 48 percent, respectively. The mean and median implied markups from the cost estimation are 34 percent and 28 percent, re-

Table A6: Cost Estimation Fit of Cost-Ratio Moments

	Data	Model Fit	
		GMM-1	GMM-2
Age ($r^{HCC} = 0$)			
18 - 24	1.0	-	-
25 - 29	1.34	1.32	1.30
30 - 34	1.44	1.53	1.56
35 - 39	2.08	2.38	2.35
40 - 44	2.98	2.08	2.07
45 - 49	1.74	2.56	2.57
50 - 54	3.49	2.74	2.84
55 - 59	2.98	3.75	3.78
60 - 64	3.57	3.75	3.82
Risk Score			
$r^{HCC} = 0$	1.0	-	-
$r^{HCC} > 0$	3.57	3.26	3.29
Metal Level			
Bronze	1.0	-	-
Silver	2.28	1.67	1.77
Gold	3.80	3.39	3.41
Platinum	4.28	7.39	7.47

Note: This table displays the targeted and estimated cost ratios that are used to identify the marginal cost estimation. In each category—age, risk score, and metal level—the ratios are defined relative to the first row. The first row of each category is equal to one by construction. The two columns of estimated moments represent the two demand estimation specifications used to simulate the moments. Marginal costs are not estimation for the final specification, GMM-3, because this specification cannot be used in counterfactual analyses.

spectively. Appendix Figure A1 plots the marginal revenue and marginal cost implied by estimated parameters under the baseline policy regime, which includes risk adjustment and reinsurance.

The fact that the demand model and Bertrand-Nash competition imply greater markups could be an indication that state insurance agencies are successful in negotiating lower markups on behalf of consumers. This mechanisms are outside of the scope of this article. To the extent that regulators can effectively discipline markups, the results that follow will underestimate the positive effects of consolidation.

F A Model of Risk Adjustment in the Affordable Care Act

The ACA includes a risk adjustment transfer policy specifically intended to mitigate between-firm adverse selection. The government administers a transfer between firms that is equal to the difference between the firm's own average cost and the implied average cost of the firm if it were to insure the same risk balance as the market as a whole (Pope et al. (2014)).[4] (For more details on the policy specifics, see Section 3.)

$$T_j(\boldsymbol{p}) = \underbrace{\frac{E[\sum_k S_{ik} c_{ik}]}{E[\sum_k S_{ik}]}}_{\text{Pooled Cost}} - \underbrace{\frac{E[S_{ik} c_{ij}]}{E[S_{ij}]}}_{\text{Average Cost}}$$

In the presence of risk adjustment transfers, the firm then faces a new average cost, $AC_j^T(\boldsymbol{p}) = AC_j(\boldsymbol{p}) - T_j(\boldsymbol{p})$. The equilibrium price can be written as

$$p_j^* + \frac{S_j}{S_j'} = \Psi_j \frac{E\left[(\sum_k \frac{\partial S_{ik}}{\partial p_j}) c_{ij}\right]}{\sum_k \frac{\partial S_j}{\partial p_j}} + \left(1 - \Psi_j\right) \frac{E[\sum_k S_{ik} c_{ik}]}{\sum_k S_k} \tag{13}$$

where,

$$\Psi_j = \frac{S_j}{\sum_k S_k} \frac{\sum_k \frac{\partial S_k}{\partial p_j}}{S_j'}$$

There are two important features of equilibrium under risk adjustment. First, the transfers adjust the private incentive of the firm according to how the marginal cost of its products deviates from the market-wide average cost. The policy-induced incentive is not the optimal sorting incentive in Equation (3) that penalizes or reward firms based on the profitability of their marginal consumers. Therefore, it does not eliminate the welfare cost of sorting.

Second, this particular policy converges to the firm's own private incentive as the market

[4] The implemented policy has to approximate this transfer using a risk-scoring system, but I will assume for theoretical simplicity that the regulator has full information about consumer risk.

share of a particular product increases or if one firm merges with others in the market. The policy follows the importance of the sorting distortion by fading out with market concentration.

G Effects of a Merger under Price-Linked Subsidies

Price-linked subsidies have two important effects in the context of mergers and market power. First, it allows firms a greater ability to exploit their market power. Greater prices are partially covered by the government rather than consumers, which reduces consumers' effective elasticity and leads to greater markups (Jaffe and Shepard (2020)).

Second, when two firms merge, the price effect is greater not only due to the reduced elasticity of consumers, but also due to the increased probability that the merged firm will control the linked product that governs the subsidy. The merged firm now internalizes more of the subsidy policy, and as a result, it is as if the merged firm faces *less elastic* consumers than pre-merger, even at the same prices and among the same consumers.

Without considering selection, the first-order effects of a merger when firms internalize the price-linked nature of subsidies are that government spending increases substantially, firms capture some of this increased spending as an increase in profits, and subsidized consumers are protected against—and in some cases can benefit from—higher prices. The key group that is harmed due to a merger are higher-income, un-subsidized consumers, which make up a smaller portion of the market.

These effects are important in the context of this article, because they greatly reduce the harm to consumers from increased markups. Consumers may benefit from mergers through less inefficient sorting without bearing the full burden of greater markups.

In this section, I repeat the main results of the article, allowing for the subsidies to adjust with prices and for firms to internalize their probability of controlling the price-linked product, i.e. the silver plan with the second lowest price. I follow the methodology of Jaffe and Shepard (2020) and require the equilibrium to be an ex-post best-response. The firm that controls the second-lowest-price silver plan sets the optimal price conditional on the knowledge that the plan is linked to the market-wide subsidy level.

In order to smooth the computation of equilibrium, I assume that firms have an expectation over the probability that the silver plan they offer in each market is the benchmark silver plan. Let p^{2lps} represent the second lowest-price silver plan. All silver plans in the market are assigned a probability that the plan is the benchmark plan,π_j, given by

$$\pi_j = \frac{e^{-\chi|p_j - p^{2slps}|}}{\sum_k e^{\chi|p_k - p^{2slps}|}}. \tag{14}$$

The parameter χ governs the certainty with which firms' know if they offer the benchmark premium. In the limiting case of a very large χ, this probability distribution collapses to certainty. In the results in this section, I set $\chi = 0.1$, which corresponds roughly to a firm knowing with 99% probability that its plan is the benchmark silver plan if the absolute price difference of the next closest silver plan is at least \$40. At the observed prices, the benchmark plans in 53 out of 107 markets are assigned probabilities greater than 70%, and in 88 markets the probabilities exceed 50%. With a greater certainty parameter, the equilibrium is more difficult to solve but it does not substantially alter the results of this section.

The price-linked model performs similarly in rationalizing the observed equilibrium. As shown in Figure A1, average marginal revenue and average marginal costs are similar in the baseline model used in the body of the article. The price-linked model has similar findings, but the marginal revenue variation matches slightly less of the marginal cost variation—33% versus 37%.

Table A7 displays the main results in the price-linked model. In both policy environments, more than half of all mergers are beneficial to consumer, and more than 1 out of 5 of the largest mergers are still beneficial to consumers.

I present two total welfare measures: one that excludes government spending (as in the body of the article) and another that includes spending. Because the primary costs of a merger in this environment are borne by the government, the vast majority of mergers increase the combined

Table A7: Many Mergers are Predicted to Improve Consumer Surplus and Social Welfare

	Number of Mergers	Fraction $\Delta CS > 0$	Fraction $\Delta SW > 0$ (Govt Spending Excl.)	Fraction $\Delta SW > 0$ (Govt Spending Incl.)
	Baseline			
Total	1186	0.53	0.83	0.07
Δ HHI				
<100	533	0.74	0.91	0.08
100 - 200	144	0.53	0.87	0.05
200 - 1000	319	0.36	0.77	0.07
>1000	190	0.21	0.66	0.03

Note: When accounting for price-linked subsidies, mergers are generally beneficial to consumers. This table displays the fraction of mergers with positive welfare effects. The top line displays the average across all mergers, and the following rows breakout the results by the size of the merger. The change in HHI is computed using pre-merger market shares to reflect pre-merger size of merging firms.

surplus accrued by consumers and firms. In the case that a dollar of government spending is associated with extra resource costs due to distortionary taxes, it may be the case that no mergers produce any welfare benefit. Similarly, if a dollar of government spending in this market is valued at less than a dollar due to preferences for redistribution, the fraction of mergers that produce a welfare benefit may be somewhere in between these two estimates.[5]

In these results, it is hard to disentangle the mechanisms created by adverse selection from those caused by the price-linked subsidies. Even in the absence of any adverse selection, mergers that lead to greater subsidies can be beneficial for both consumers—many prices fall in absolute terms due to greater subsidies—and for firms—due to greater profit at costs borne primarily by the government. These dynamics are important considerations for this market but outside of the primary scope of this article.

[5]As mentioned in Section 4, the total surplus generated by an additional dollar of government spending is less than a dollar.