Adverse Selection and Cost Functions in Insurance Markets
ECON 552/890

Luca Maini[*]

January 17, 2022

See most recent version here

CONTENTS

[*]These notes are a work in progress. Please check back frequently for an updated version, and email lmaini@email.unc.edu for corrections.

Insurance markets are unlike most other product markets (e.g., manufacturing) in that "production" costs depend on the characteristics of the customer who purchases the product. The cost of providing insurance is higher for individuals at high risk. As a result, car insurance is more expensive for drivers with a history of accidents, homeowner insurance costs more in areas at high risk of flooding or earthquakes, and life insurance is more expensive for older customers and customers with pre-existing medical conditions. Often, the insurance company cannot accurately assess all risk factors of a specific customer and will offer the same insurance contract to a group of customers with different risk profiles. As a result, products with different costs are sold at the same price.

The relationship between cost and agent characteristics is the result of two phenomena. The first one is the notion that different consumers have a different proclivity to generate insurance claims and that price changes will change the set of consumers purchasing insurance, thus affecting its cost. This often leads to a phenomenon known as **adverse selection**, which was formally introduced in economics for the first time in the seminal work of Akerlof (1970).[1] The second one is **moral hazard**: the notion that consumers can put a varying amount of effort into actions that will decrease or increase the probability of generating an insurance claim. Deductibles, copays, and coinsurances are the insurance company's instruments to fight moral hazard.

1 Adverse Selection in Insurance

1.1 A primer on adverse selection in insurance markets (Einav and Finkelstein, 2011)

I start by introducing the issue of adverse selection. This section is based on Einav and Finkelstein (2011). I focus on health insurance, but keep in mind that the same intuition will translate to any other type of insurance.

Consider a perfectly competitive market where many firms provide the same health insurance plan to individuals who vary in their ex-ante probability of getting sick. Consumers choose to either buy insurance or not, and firms only choose what price to set. (The assumption that all health plans are the same is not innocuous. Analyzing markets where different insurers can offer different plans (see, e.g., Rothschild and Stiglitz, 1976) is exponentially more complicated, though it does not significantly alter the implications of what we focus on here.)

You can think of the health insurance plan as a contract shielding the consumer from risk. If the patient gets sick, the insurer covers some of the patient's healthcare costs.

Figure 1 plots the demand curve in this market. Since consumers are purchasing a single unit of the product, the demand curve shows the willingness to pay of each consumer for insurance, which we can break down into two elements. First, the expected cost that the insurer is going to cover. This should be pretty intuitive. If I expect to get sick with probability $\frac{1}{2}$, and treatment costs $\$1,000$, the expected cost of the insurance contract is \$500, and the insurer will require me

[1] To be clear, the concept had been known for at least a century before Akerlof, with the words "adverse selection" appearing in print as early as the 1870s (see, e.g., https://books.google.com).

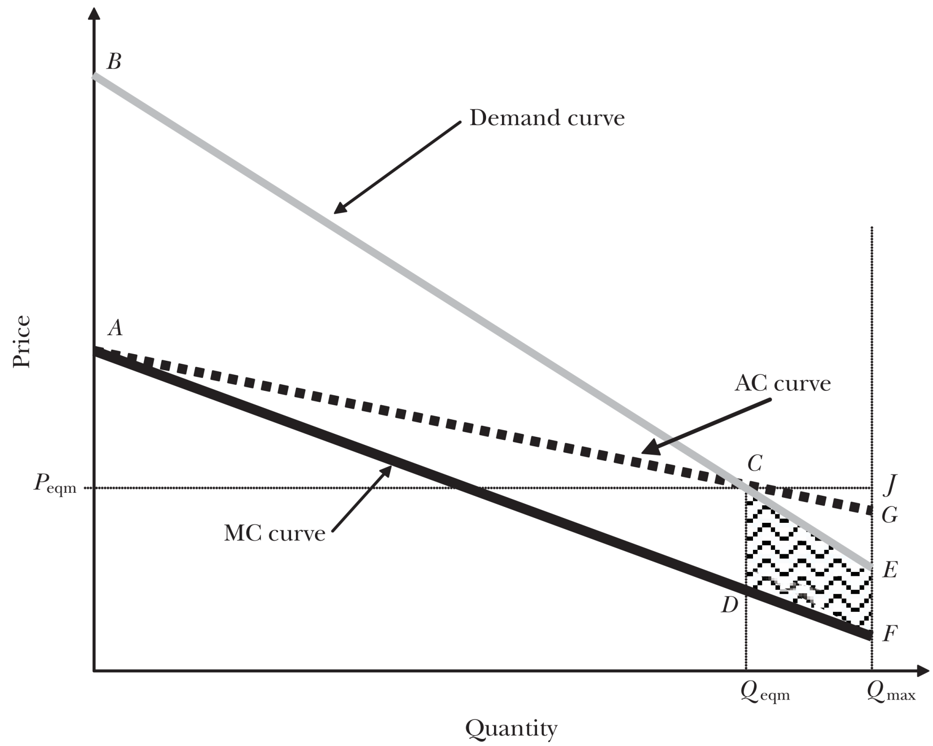

Figure 1: Adverse selection in a textbook setting (Einav and Finkelstein, 2011)

to pay at least that much in order to break even. Second, the risk premium. The risk premium exists because individuals are risk-averse. Given a choice between a lottery of two outcomes, or the expected value of that lottery, most people would choose the latter. As a result, willingness to pay for insurance is usually higher than expected cost.

The graph highlights an essential aspect of the problem: willingness to pay is higher for consumers who expect to be sick more often. This differential is reflected by the downward slope of the marginal cost curve. Consumers with a high probability of getting sick have a higher willingness to pay. Notice as well that the average cost function is also downward-sloping and above the marginal cost function. (Why? Because the average cost for a given quantity Q is averaged across all the consumers from 0 to Q, and the *marginal* consumer has the lowest cost of all.)

Since this is a perfectly competitive market, the competitive equilibrium price and quantity satisfy the zero-profit condition, i.e. price equals average cost. This leads to the equilibrium denoted by point C in the graph.

It should be immediately obvious that C is an inefficient equilibrium. Why? Because it leaves out consumers whose willingness to pay for insurance is above their marginal cost (the segment from Q_{eqm} to $Q_{\max}$ in the graph). In an efficient market, price is equal to marginal cost. This discrepancy is at the heart of the inefficiency caused by adverse selection. Because marginal cost is always below average cost, the competitive equilibrium leads to under-insurance.

Because even a perfectly competitive insurance market can be inefficient, insurance markets tend to be heavily regulated, both abroad and in the US. The two most common policy tools to deal with the problem of under-insurance are **mandates** and **subsidies**.

The main selling point for mandates is to ensure that the optimal equilibrium (i.e., full insurance) is reached. They have downsides, though: if you enforce a mandate, some consumers will have to pay a price for insurance that is above their willingness to pay (these are the consumers at the very right end of the x-axis, close to $Q_{\max}$). So even though the market is efficient, some consumers are worse off. In markets that are not perfectly competitive, where insurers can charge a premium on top of average cost, the distortion can be larger, and you may even have over-insurance.[2]

Relative to mandates, subsidies do not make anyone worse off. The effect of a subsidy is to push the demand function upward. The downside is that subsidies are expensive. Since many individuals face budget constraints, the subsidy required to achieve full insurance can be substantial. Additionally, it is hard to target subsidies towards the patients who need them the most, which means that a lot of funds will be given to individuals who would have purchased insurance anyway. If the government could raise money at no cost, this would not be a big issue. However, raising funds can be very costly when it is done through taxation. You also may not know in advance what subsidy you need.

If you look at health systems worldwide, they generally use a combination of mandates and subsidies. That is how the ACA was set up. Outside the US, most single-payer systems provide universal health insurance, more or less for free, but pay for it through taxes.

Some talking points:

- What happens to cost when a plan removes copays for office visits?
- What happens to costs if a plan stops covering a very good hospital (Shepard, 2020)?
- The ACA imposed a mandate, but did not enforce it in a strict way. As a result, a small, but significant percentage of patients remained uninsured. In the second year of the ACA, premiums increased, and a few insurers left the market. Do you think those reactions make sense? Why?
- Extreme adverse selection can result in complete market unraveling (sometimes also called a "death spiral"). This happens when demand is between marginal cost and average cost (Figure 2). It's hard to find evidence of market unraveling in the real world, because, well, if the market unravel there is no market to analyze. That being said, evidence of market unraveling exists: see for example Cutler and Reber (1998). Other examples outside health care include insurance markets for housing near areas at high risk of flooding.[3]

[2] Over-insurance in a perfectly competitive market requires risk-loving individuals who value an insurance contract below its expected value. While this is theoretically possible, it is unlikely to matter in practice. However, if premiums include a markup over expected cost, then some risk-averse individuals may optimally choose to forgo insurance if their risk-premium is lower than the markup.

[3] https://www.iii.org/fact-statistic/facts-statistics-flood-insurance

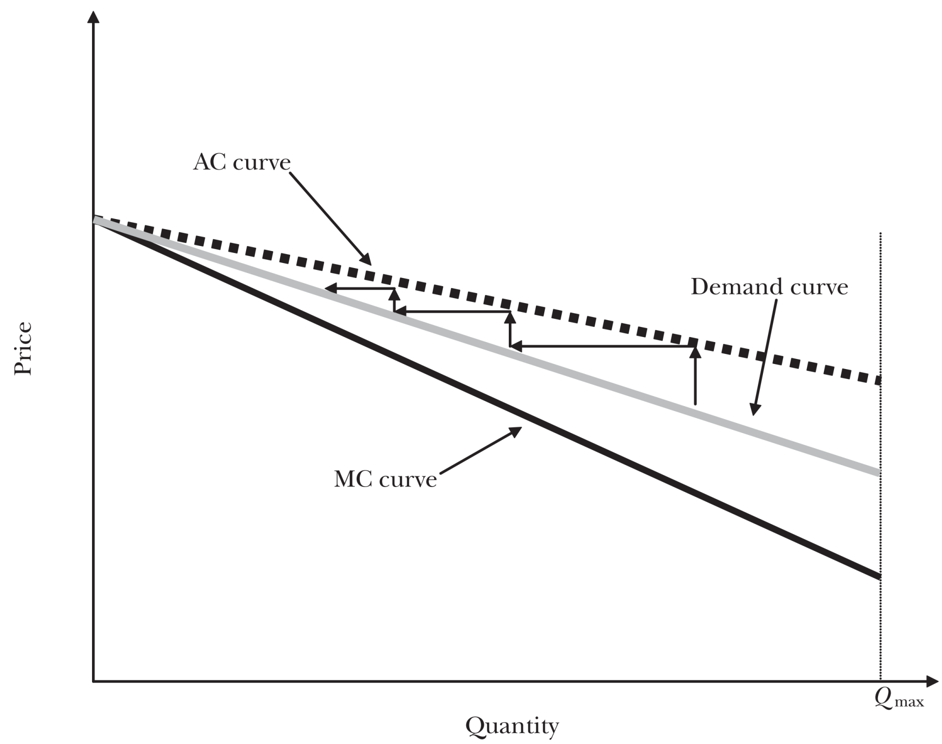

Figure 2: Complete market unraveling (Einav and Finkelstein, 2011)

1.2 Testing for Adverse Selection I (Cardon and Hendel, 2001)

Cardon and Hendel (2001) looks for evidence of adverse selection in the health insurance market. If adverse selection exists it means that individuals know their likelihood of filing an insurance claim better than the insurance company can predict, or price on. The authors estimate a model of insurance choice with an unobserved component, and test the correlation between this component and the claims/cost equation.

Theoretical Model

The model is purely about demand, and takes place over two sequential stages: insurance choice, and choice of health care consumption. Agents—indexed by i—find out their health state only after purchasing insurance, but have some prior knowledge of their future health state that the econometrician cannot observe. Health insurance plans are indexed by j.

Second stage Utility is written as

$$U(m_i, h_i) = U(m_i; x_i + s_i) \tag{1}$$

where m_i is a composite good, health h_i is the sum of a random health state s_i and health care x_i. Given a policy j and a health state s_i, agents maximize utility subject to

$$m_i + C\left(x_i, Z_j\right) = y_i - p_j \tag{2}$$

the price of the composite good is normalized to 1, $C\left(x_i; Z_j\right)$ is the cost of health care consumption, which depends on plan characteristics Z_j, and p_j is the premium of plan j. Combining 1 and 2 yields the following maximization problem

$$U_{ij}^{\star}\left(s_i\right) \equiv \max_{x_i} U\left(y_i - p_j - C\left(x_i, Z_j\right); x_i + s_i\right)$$

The first order condition of this maximization problem gives an expression for $x_i^{\star}\left(Z_j, s_i\right)$, i.e. optimal medical care choice given health status and plan characteristics.

First stage Assume that the agent's decision whether to buy insurance (and which insurance to buy) depends on

- observable characteristics;
- an unobservable error ν_{2i} that affects health outcomes;[4]
- another error ε_{ij} that affects plan choice but not health outcomes (and therefore becomes irrelevant in the second stage)

Under these assumptions, the expected utility from purchasing plan j is

$$\begin{aligned} V_{ij}\left(\nu_{2i}, \varepsilon_{ij}\right) &= \mathbb{E}\left[U_{ij}^{\star}\left(s_i\right)\middle|\, \nu_{2i}\right] + \varepsilon_{ij} = \\ &\int_{s_i} U\left(\left(y_i - p_j - C\left(x_i^{\star}\left(Z_j, s_i\right), Z_j\right); x_i^{\star}\left(Z_j, s_i\right) + s_i\right)\right) \mathrm{d}F\left(s_i \middle|\, v_{2i}, D_i\right) + \varepsilon_{ij} \end{aligned} \tag{3}$$

where D_i are demographic characteristics, and ε_{ij} is assumed to be EV(1) in order to obtain a simple analytic formula for market shares.

Statistical Model

In order to take the model to the data, one needs to make some functional and distributional assumptions. The authors of the paper approximate utility using a second-degree polynomial in m_i and h_i. They model out-of-pocket costs as a piece-wise linear function, where the consumer pays full cost up to the deductible and a fraction of the cost after that. This parametrization allows them to derive the first-order conditions analytically. The FOCs have two possible interior solutions (because of the second-degree polynomial assumption). To identify the global optimum

[4] The use of ν_2 to denote a "structural" error—i.e. an error term that the agent observes prior to making a choice, but that the econometrician does not observe—is a convention from the IO literature.

they check each point, plus the two corner solutions (no health care consumption or no composite good consumption).

Finally, they assume that the health shock s_j follows probability distribution function

$$s_i = -\exp\left\{K\left(D_i\right) + \nu_{1i} + \nu_{2i}\right\} \tag{4}$$

with mean $K\left(D_i\right) + \nu_{2,i}$. $K\left(D_i\right)$ is modeled as a linear function of demographics (age, age squared, sex, region indicators, race indicators, and an indicator for white-collar worker). ν_{1i} and ν_{2i} are both assumed to be normally distributed with variance $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$ respectively (this in turn implies that s_i follows a lognormal distribution). The variance of the shock to the agent is $\sigma^2_{\nu_1}$ (while to the insurance it is $\sigma^2_{\nu_1} + \sigma^2_{\nu_2}$). As an aside, notice how impactful this assumption is, and how important it is to specify it clearly. The error ν_2 is:

- observed by the agent, who takes into account when making choices,
- unobserved by the insurance company, who treats it as a random shock in their own calculations,
- unobserved to the econometrician, who must account for selection and endogeneity when estimating moment conditions.

Estimation

The data comes from the National Medical Expenditure Survey, and is divided among insured and uninsured individuals.[5] Uninsured tend to have lower incomes, be employed by smaller firms who don't offer health care, be younger and spend less on healthcare. At a first glance the data is consistent with an adverse selection story: those who are insured have higher expenditures, and more so when the cost of insurance is higher. However, those who do not have insurance also have lower incomes, so a story of selection based on income also makes sense.

They estimate the model using a method-of-moments estimator. The parameter of interest is θ, and contains:

- The five coefficients from the utility polynomial:

$$U\left(m_i, h_i\right) = \phi_1 m_i + \phi_2 h_1 + \phi_3 m_i h_i + \phi_4 m_i^2 + \phi_5 h_i^2 \tag{5}$$

- The coefficients of the function $K\left(D_i\right)$
- Variance of error terms ν_1 and ν_2

To build the moment conditions they proceed as follows:[6]

[5] The current version of this survey, which is publicly available, is Medical Expenditure Panel Survey (MEPS).

[6] My explanations of the estimation steps is different from the explanation you see in the paper. Therefore, it is possible (though not likely), that the steps followed by the authors may not match my description exactly.

- Given a guess for θ, predict health care consumption choices, and then simulate expected utility from each plan to predict choices. This is done in four steps:

  1. Predict optimal expenditures under each plan

  $$x_{ij}\left(\theta, \nu_{2,i}, D_i\right) = \int_{\nu_1} x_i^{\star}\left(Z_j, K\left(D_i\right) + \nu_{2,i} + \nu_{1,i}\right) \mathrm{d}F\left(\nu_{1,i}\right) \tag{6}$$

  (Notice that these expenditures are a function of $\nu_{2,i}$, so we only need to integrate over ν_1)

  2. Obtain the expected utility of each plan j as a function of $\nu_{2,i}$ and demographics:

  $$V_{ij}\left(\theta, \nu_{2,i}, D_i\right) = \int_{\nu_1} U\left(m_i, x_{ij}\left(\theta, \nu_{2,i}, D_i\right) + K\left(D_i\right) + \nu_{2,i} + \nu_1\right) \mathrm{d}F\left(\nu_1\right) \tag{7}$$

  3. Simulate market shares as a function of $\nu_{2,i}$ and demographics. Thanks to the logit error ε_{ij}, these can be written as

  $$MS_j\left(\theta, \nu_{2,i}, D_i\right) = \frac{\exp\left(\int_{s_i} V_{ij}\left(\theta, \nu_{2,i}, D_i\right) \mathrm{d}F\left(s_i \middle| v_{2i}, D_i\right)\right)}{\sum_k \exp\left(\int_{s_i} V_{ij}\left(\theta, \nu_{2,i}, D_i\right) \mathrm{d}F\left(s_i \middle| v_{2i}, D_i\right)\right)} \tag{8}$$

  If the functional form inside the integral is not familiar to you, go check the derivation of logit market shares (for example, here).

  4. Obtain market-level shares by integrating over $\nu_{2,i}$:

  $$MS_j\left(\theta, D_i\right) = \int_{\nu_2} MS_j\left(\theta, \nu_2, D_i\right) \mathrm{d}F\left(\nu_2\right) \tag{9}$$

  (You could base estimation on shares conditional on demographics, though I think they use market-level shares in practice).

- Given market shares, they can also calculate expected expenditure for an individual under policy j as

$$x_{ij}\left(\theta, D_i\right) = \int_{\nu_1}\int_{\nu_2} x_i^{\star}\left(Z_j, K\left(D_i\right) + \nu_2 + \nu_1\right) \times MS_j\left(\theta, \nu_{2,i}, D_i\right) \mathrm{d}F\left(\nu_2\right) \mathrm{d}F\left(\nu_1\right) \tag{10}$$

  (notice here that $MS_j\left(\theta, \nu_{2,i}, D_i\right)$ acts as a modifier that essentially gives you the conditional probability that an individual with characteristics $\nu_{2,i}$ and D_i chooses plan j).

- Obtain moment conditions by taking the difference between (i) the market share predictions $MS_j\left(\theta, D_i\right)$ and the market shares observed in the data, and (ii) the expected spending for each plan $x_{ij}\left(\theta, D_i\right)$ and the average spending observed in the data.

- The difference is the error term and they interact it with instruments to build extra moment conditions.

Identification

The authors use survey data, and do not have any policy change, or instrument that they can use to claim causal identification. This actually makes the paper a good example of how *structural* identification works.

Structural identification comes first and foremost from the model's parametric and distributional assumptions. In general, it's hard to test the viability of these assumptions, though it's occasionally possible to test things like error autocorrelation, and the presence of private information. Additionally, one can usually point to aspects of the data that drive the estimates of specific parameters.[7]

In this case, choice set variation is probably important. Different individuals have access to different plans. If plan offerings are uncorrelated with the characteristics of the patient population, this provides a nice source of exogenous variation. This is a big if. Plan characteristics are likely tailored to the population they are targeting. Still, targeting is imperfect, so choice set can provide some mix of exogenous and endogenous variation.

The main threat to identification is that more generous plans will have higher costs because of both adverse selection and moral hazard, and there is no clear way of disentangling the two here (see more on this in the final comments). The ideal experiment here would be to offer the same set of plans to two groups of individuals, but randomly change the premiums. Holding cost-sharing fixed implies that moral hazard should more or less be constant across the two groups, but differences in premiums will affect how adverse selection plays out.

Results

See the paper for more details. The bottom line is that they find little evidence of adverse selection, and a small expenditure elasticity with respect to price. The elasticity estimate appears consistent with estimates from the RAND Health Insurance Experiment (Manning et al., 1987).

Final Comments

The paper is quite useful as a methodological learning tool. It also provides a good, simple setup for a generic insurance choice problem. If you wanted to do something similar today however, you would want to make sure you think carefully about some potential empirical limitations:

- The modeling of heterogeneity across consumers is very limited. Consumers are only allowed to vary in terms of their unobserved health risk. There is no room to model heterogeneity in other relevant dimensions, such as (i) risk-aversion (more on this later), or (ii) price-sensitivity.

[7] I provide only an informal discussion of this issue here. For a formal treatment of sensitivity analysis see Andrews, Gentzkow and Shapiro (2017).

- More on price sensitivity. It is not clear that you can estimate price-sensitivity separately from adverse selection. The basic correlation that comes out of the data is that patients with more generous plans consume more healthcare. There are two possible explanations for this: one is that patients of more generous plans are sicker, and require more care. The second is that cost-sharing is lower in more generous plans, and therefore patients consume more care. There are a few ways around this problem:

  – The authors say that their data allows them to see similar (in terms of observable characteristics) individuals facing similarly generous policies with different premia. They then claim that price-sensitivity is identified from coinsurance variability across individuals, whereas adverse selection is identified off of plan choices of individuals facing different premiums. You might worry about this for two reasons. First, it assumes homogeneity in all the dimensions discussed in the previous point, which may not be realistic—e.g. individuals who pick more expensive plans tend to be higher income, and higher income individuals tend to be less price-sensitive. Second, it is essentially treating premiums as exogenous, which is also unrealistic.

  – If one were willing to assume that the distribution of health outcomes conditional on observables was the same across employers it would be possible to distinguish between moral hazard and adverse selection since different firms offer more or less generous plans. To understand why think of the extreme case in which each employer only offers a single plan. Then we have identical individuals (unobservables have the same distribution), and we can compare how they behaved under different plans. Of course in this case you'd worry that variation in plans offered may have something to do with the distribution of health outcomes. Possible escape routes out of this conundrum include cost variation (e.g. in the provision of medical care), or alternative sources of price variation (see for example the price variation in the paper we present next).

- The main test that the authors run may not be conclusive. Even though the econometrician does not know ν_{2i}, the insurance company might, and might be able to use that information to set premiums. So finding a correlation between ν_{2i} and health expenditure does not necessarily imply adverse selection.

- The small price elasticity is at odds with more recent research (see e.g. results from the Oregon Health Insurance Experiment).

1.3 Testing for Adverse Selection II (Einav, Finkelstein and Cullen, 2010)

Einav, Finkelstein and Cullen (2010) attempt to estimate the welfare loss from adverse selection from claims data. One of the key contributions of the paper is in introducing a very intuitive graphical test of adverse selection, which basically consists of estimating the slope of the cost function. A flat cost function implies that there is no selection in the data. An upward-sloping cost function implies adverse selection. A downward-sloping cost function would imply advantageous selection. The main caveat of this paper is that they can exploit (what they claim is) exogenous variation in price, which is rare. Therefore the method cannot be generalized easily. However, the intuition of the test remains.

The paper also has a separate section on welfare. We concentrate on the adverse selection framework here, and only mention the welfare section in passing.

Theoretical Model

Consider a stylized model where agents choose between a high-generosity and a low-generosity plan. Denote

$$\pi(\xi) = \mathbb{E}[U_H(\xi)] - \mathbb{E}[U_L(\xi)] \tag{11}$$

as the willingness to pay for the high plan as a function of consumer characteristics ξ. ξ's are ordered and follow distribution $G(\cdot)$. They denote any type of consumer heterogeneity that could reflect willingness to pay for higher coverage (this includes risk-aversion, price-sensitivity, etc.). The demand function is given by

$$D(p) = \int \mathbb{I}(\pi(\xi) \geq p)\, dG(\xi) = \mathcal{P}(\pi(\xi) \geq p) \tag{12}$$

Assuming a perfectly competitive Bertrand market with two or more firms, health plan premiums will equal average cost, which in this case is

$$AC(p) = \frac{1}{D(p)} \int c(\xi)\, \mathbb{I}(\pi(\xi) \geq p)\, dG(\xi) = \mathbb{E}[c(\xi) | \pi(\xi) \geq p] \tag{13}$$

The marginal cost curve can be backed out by differentiation the average cost curve:

$$MC(p) = \frac{\partial TC(p)}{\partial D(p)} = \frac{\partial (AC(p) \cdot D(p))}{\partial D(p)} = \left(\frac{\partial D(p)}{p}\right)^{-1} \frac{\partial (AC(p) \cdot D(p))}{\partial p} \tag{14}$$

Finally, assume that it is optimal to provide insurance to at least some individuals (i.e. no full unraveling), and that the marginal cost curve crosses demand at most once. This guarantees existence and uniqueness of equilibrium at the lowest break-even price.

Estimation

The authors have access to a very good datasets, which gives them data on a geographically varied set of employees of a large US company. Employees at different sites (over 300) pick plans from the same choice set, but with different premiums. The data includes (i) health plan premiums and market shares, (ii) cost data from health insurance claims, (iii) demographics. To estimate the AC curve one also needs data on the expected cost for all individuals with contract H, as well as data on subsequent risk realization and how it translates to insurer costs. Then, use the same variation in prices to trace the cost curve.

In their baseline specification the paper estimates linear demand and cost functions

$$D_i = \alpha + \beta p_i + \varepsilon_i \tag{15}$$
$$c_i = \gamma + \delta p_i + u_i \tag{16}$$

where D_i is a dummy for choice of contract H and c_i is the *incremental* cost between plan L and plan H. Using this specification one can write

$$MC(p) = \frac{1}{\beta}\left(\frac{\partial(\alpha+\beta p)(\gamma+\delta p)}{\partial p}\right) = \frac{\alpha\delta}{\beta} + \gamma + 2\delta p \tag{17}$$

Under the identification assumption of exogenous price variation (in other words $\mathbb{E}[\varepsilon_i | p_i]$ and $\mathbb{E}[u_i | p_i]$, one can estimate α, β, γ, and δ from equations 15 and 16 using simple OLS.

Identification

The entire empirical section of the paper rests on the assumption that price variation is exogenous. Some of the evidence they present does not quite align with the assumption (see, e.g., 2003 spending in Table I on p. 903), but most of it does. The big lesson here is not so much about the veracity of the assumption itself, but about the effort exerted by the authors in defending it. It is a very good example of what an economics paper should do when introducing unusual assumptions.

Results

Figure 3 shows the main results of the paper, with the estimated average and marginal cost curves. The authors find evidence of adverse selection, though the magnitude is relatively small.

Addendum: welfare loss from adverse selection

I present the welfare section of the paper for completeness.

Starting from where we left off in the theoretical model, we can write welfare as the sum of consumer and producer surplus, both of which are calculated as a certainty equivalents:

$$CS = \int \mathbb{E}[U_H(\xi) - p]\,\mathbb{I}(\pi(\xi) \geq p) + \mathbb{E}[U_L(\xi)]\,\mathbb{I}(\pi(\xi) < p)\,dG(\xi) \tag{18}$$

and

$$PS = \int \mathbb{E}[p - c(\xi)]\,\mathbb{I}(\pi(\xi) \geq p)\,dG(\xi) \tag{19}$$

together

$$TS = \int \mathbb{E}[U_H(\xi) - c(\xi)]\,\mathbb{I}(\pi(\xi) \geq p) + \mathbb{E}[U_L(\xi)]\,\mathbb{I}(\pi(\xi) < p)\,dG(\xi) \tag{20}$$

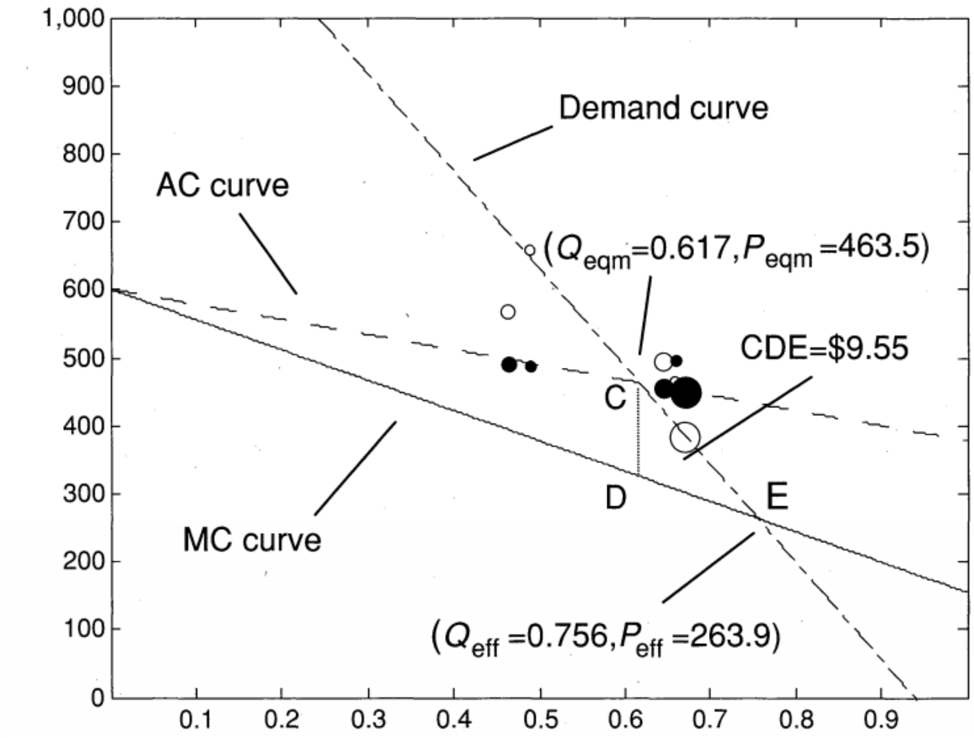

FIGURE V

Efficiency Cost of Adverse Selection—Empirical Analog

This figure is the empirical analog of the theoretical Figure I. The demand curve and AC curve are graphed using the point estimates of our baseline specification (see Table III). The MC curve is implied by the other two curves, as in equation (13). The circles represent the actual data points (see Table II, columns (3) and (4)) for demand (empty circles) and cost (filled circles). The size of each circle is proportional to the number of individuals associated with it. For readability we omit the one data point from Table II with only seven observations (although it is included in the estimation). We label points C, D, and E, which correspond to the theoretical analogs in Figure I, and report some important implied point estimates (of the equilibrium and efficient points, as well as the welfare cost of adverse selection).

Figure 3: Efficiency cost of adverse selection (Einav, Finkelstein and Cullen, 2010, Figure V)

Maximum TS is achieved if individuals buy insurance whenever $\pi(\xi) \geq c(\xi)$.

Graphically, refer to Figure 4 for an intuition of the efficiency cost of adverse selection. In particular, under the (very mild) assumptions of the paper, the cost and demand curves are the only thing needed for welfare analysis: the welfare loss is the area between the two.

The efficiency cost associated with competitive pricing (ref. Figure 4) can be computed algebraically by finding the two equilibria (one, the intersection of average cost and demand, the other, the intersection of marginal cost and demand). They find the loss to be approximately $10 per employee per year. So, not that large.

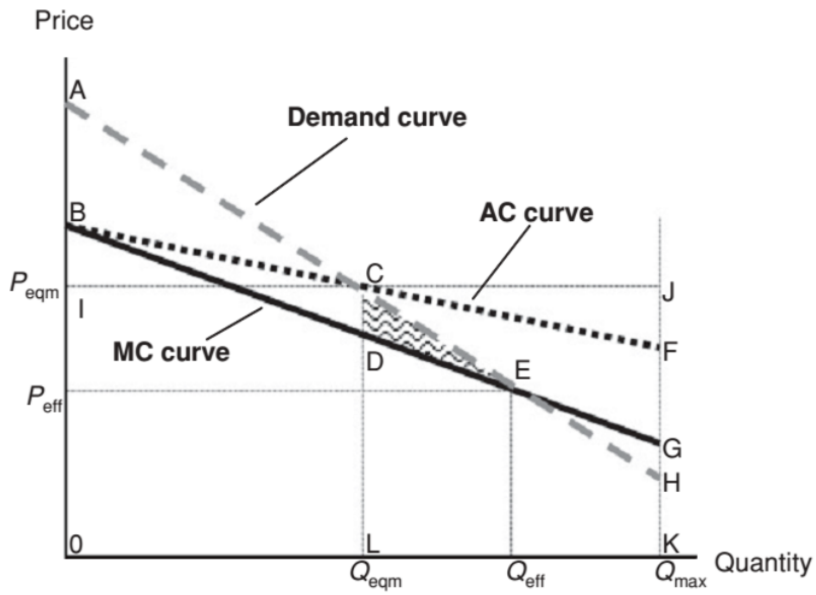

FIGURE I
Efficiency Cost of Adverse Selection

This figure represents the theoretical efficiency cost of adverse selection. It depicts a situation of adverse selection because the marginal cost curve is downward-sloping (i.e., increasing in price, decreasing in quantity), indicating that the people who have the highest willingness to pay also have the highest expected cost to the insurer. Competitive equilibrium is given by point C (where the demand curves intersects the average cost curve), whereas the efficient allocation is given by point E (where the demand curve intersects the marginal cost curve). The (shaded) triangle CDE represents the welfare cost from underinsurance due to adverse selection.

Figure 4: Efficiency cost of adverse selection (taken from Einav, Finkelstein and Cullen (2010))

Final Comments

The authors run auxiliary calculations to see what would be the optimizing price welfare (turns out to be 0), and what the impact of a mandate would be (welfare cost is three times as large, though this prediction takes them quite a bit out of sample).

The main limitation, other than the fact that truly exogenous price variation is hard to come by, is that not recovering underlying primitives (e.g. utility functions) limits the scope of counterfactuals (e.g. what would happen if we offered a different set of insurance contracts?). On the other hand, the estimation is a lot more robust because—unlike Cardon and Hendel (2001)—it does not need to rely on parametric assumptions. This is a common tradeoff in IO: more assumptions allow you to run more interesting counterfactuals, but restrict the generalizability of the findings.

1.4 Advantageous Selection in Insurance Markets

We conclude by mentioning the possibility that selection might actually work in the insurance company's favor occasionally. For example, consumers that are very risk-averse might both want to purchase generous insurance, *and* take extra steps to ensure that they never have to use it. This could be driven by the fact that most events that require insurance carry non-insurable loss. For example, accidents may cause lasting harm that car insurance does not cover, house insurance does not compensate for the loss of precious memorabilia, and health insurance does not prevent patients from suffering from disease.

Cutler et al. (2008)

This short paper argues that heterogeneity in risk-aversion is important. It finds that more risk-averse individuals are more likely to buy insurance and engage in activities that decrease the risk of using it. The analysis is simple and uses the equation:

$$\mathbb{I}_{\{\text{INSURANCE}\}} = \beta \{\text{BEHAVIOR}\} + X_1 \Gamma + \varepsilon \tag{21}$$

$$\text{RISK OCCURENCE} = \alpha \{\text{BEHAVIOR}\} + X_1 \Pi + \eta \tag{22}$$

where INSURANCE refers to a type of insurance (life, health, annuities, Medigap, etc.), BEHAVIOR refers to an action (smoking, drinking, receipt of preventive health care, seat belt usage, etc.), and RISK OCCURRENCE is an outcome variable related to the need for insurance such as (mortality, entering a hospital, transfer to a nursing facility, etc.).

The paper makes the point that we need a framework that allows for consumer heterogeneity in risk-aversion.

Einav et al. (2012)

This paper looks at the relative importance of risk-aversion over different types of risks. They compare the insurance choices of a cohort of employees over 5 different types of insurance. Find that there is high correlation between choices (though not approaching 1 in any case), and that choices in other domains have more predictive power than demographics, and measures of risk, and other individual characteristics. They conclude that despite the fact that there seems to be a common component across all domains, there is probably more than one risk-aversion parameter.

2 Cost Function Estimation in Insurance Markets

Here, I provide a brief overview of structural cost-estimation in insurance markets. Unlike Einav et al. (2010), I will not assume that I can rely on exogenous price variation. Instead, I will think about how to estimate cost from the ground up. I also provide some details on why cost function estimation is different in insurance markets (relative to more traditional consumer good markets).

2.1 The general cost-estimation setup

If we assume that we have recovered demand primitives, the general framework for cost estimation simply requires an assumption on the equilibrium played by firms in the market. The equilibrium condition (say, e.g. Nash in prices, or Bertrand) will translate to a vector of first-order conditions from each firm's optimization problem, which can be expressed in terms of demand and cost functions. For example, consider a simple oligopoly problem where firms compete over price and play a static Nash equilibrium. Each firm j solves

$$\max_{p_j} D_j\left(p_j, p_{-j}\right) \times p - C\left(D_j\left(p_j, p_{-j}\right)\right) \tag{23}$$

where we use the notation $-j$ to indicate the prices of other firms in the market. The moment condition is

$$\frac{\partial D_j}{\partial p_j} p_j + D_j\left(p_j, p_{-j}\right) = \frac{\partial C}{\partial D_j} \times \frac{\partial D_j}{\partial p_j} \tag{24}$$

Assuming the left-hand side of the equation is known, all that is needed for estimation of $c(\cdot)$ is a functional assumption, e.g. constant marginal cost:

$$c\left(D\left(p_j, p_{-j}\right)\right) = c \times D\left(p_j, p_{-j}\right) \tag{25}$$

which implies $\frac{\partial C}{\partial D} = c$.[8] (Generally our models are over-identified, so one would need to include an additional error term to fit the data, such as $\frac{\partial C}{\partial D} = c + \varepsilon$. You also need to specify assumptions on ε to make sure that its conditional expectation in the data is, in fact, equal to 0).

2.2 A two-stage model of insurance choice

To highlight the differences with the standard cost estimation setup, we present a two-stage model of insurance choice. The model is similar to the one in Cardon and Hendel (2001) (you'll also find various versions of this model in virtually any paper dealing with insurance choice). It includes two sequential choices: insurance, and healthcare expenditure. The sequence reflects not only the logical unfolding of events (patients must know their insurance policy before choosing how much healthcare to consume), but usually also the revelation of additional information. In this case, the information reflects patient learning about their health, usually modeled as the realization of a health shock.

I index agents using the subscript i. Each agent is defined by a set of characteristics X_i and is exogenously endowed with wealth y. Health insurance plans are indexed by j, and are defined by a premium p_j and characteristics Z_j. Characteristics here can be both financial (e.g., cost-sharing parameters), and non-financial (e.g., the network of providers that accept the insurance plan). In

[8]Notice that we don't have to worry about $\frac{\partial p_{-j}}{\partial p_j}$ because the conditions for Nash equilibrium hold behavior of other firms fixed. In some empirical settings, especially dynamic situations where firms have time to react to deviations, this type of simplification might not be credible.

the first stage, agents choose a health insurance plan. In the second stage, after learning their sickness status, they decide how much health care to consume.

Expenditure stage

Since it is a sequential model, I solve it recursively, starting from the last stage. Given the contract $\left(p_j, Z_j\right)$ chosen in the first stage, the agent solves

$$U_{ij}\left(X_i, Z_j, s_i\right) = \max_{e \geq 0} U\left(h\left(s_i, e\right), y - o\left(e, Z_j\right) - p_j, X_i\right) \tag{26}$$

where utility $U\left(h, x\right)$ depends on health status h and consumption income x. $h\left(\cdot\right)$ is a health function that depends on the realization of a sickness random variable s_i and expenditures on health care e. $o\left(\cdot\right)$ is out of pocket expenditures as a function of total expenditures and patient characteristics can affect the utility of the individual directly (e.g. think risk-aversion). In this model, moral hazard occurs if e is a function of Z_j.

This model generates restrictions based on observed choices. Since e is a continuous variable, it is usually possible to get point-identification from the first-order condition:

$$\frac{\partial U}{\partial h} \times \frac{\partial h}{\partial e} + \frac{\partial U}{\partial x} \times \frac{\partial x}{\partial e} = 0 \tag{27}$$

To estimate equation 27 one needs data on expenditures e and o (usually from claims data). Separate identification of $U\left(\cdot\right)$ and $h\left(\cdot\right)$ is usually done by functional assumption—not great, but hard to do better in this case.[9] The issue is that we don't know how health affects utility, and neither can be measured directly.

After estimating 27, the econometrician recovers optimal expenditure as a function of patient and plan characteristics, and the random health shock s_i: $e^{\star}\left(X_i, Z_j, s_i\right)$.

Plan choice stage

In the first stage, consumer i chooses plan j to solve

$$\max_j \int_S U_{ij}\left(X_i, Z_j, s_i\right) dP\left(s_i \middle| X_i\right) \equiv \max_j \mathbb{E}\left[U_{ij}\left(X_i, Z_j, s_i\right)\right] \tag{28}$$

where S represents the support of the random variable s_i.

Estimation of this stage requires data on plan and consumer characteristics, plan choices, and choice sets. Notice that the econometrician may not observe all components of X_i, which means allowing for unobserved heterogeneity is usually important. An additional complication in this setting is the well-documented presence of behavioral biases in these choices. Patients are known

[9] For example, Cardon and Hendel (2001) assumes that health is a linear function of expenditure plus a random shock, and that utility is a second-order polynomial of h and x.

to suffer from inertia (Handel, 2013), incomplete information (Handel and Kolstad, 2015), and occasionally may not even consider the full choice set (Barseghyan et al., 2020).

Since this is a discrete choice model, restrictions comes from a revealed-preference argument: agents maximize their utility, and therefore the utility of the chosen plan must be greater than that of all other plans. Revealed preference restrictions are inequalities, which occasionally means that we will not get point-identification. However, in many instances we can make distributional assumptions on the error term that allow us to get point identification from observed choice probabilities (i.e., market shares). For example, adding a logit error term to the utility function delivers the traditional logit formula for choice probabilities:

$$\mathcal{P}\left(i \text{ choose } j\right) = \frac{\exp\left(\mathbb{E}\left[U_{ij}\left(X_i, Z_j, s_i\right)\right]\right)}{\sum_{k \in \mathcal{J}_i} \exp\left(\mathbb{E}\left[U_{ik}\left(X_i, Z_k, s_i\right)\right]\right)} \tag{29}$$

where $\mathcal{J}_i$ denotes the set of all plans in the choice set of consumer i.

From the estimated choice probabilities, the econometrician can recover demand functions $q_j^{\star}\left(X_i, Z_j, p_j, Z_{-j}, p_{-j}\right)$ for each individual patient with characteristics X_i. These can then be aggregated over the distribution of X_i to obtain aggregated equilibrium demand functions

$$D_j^{\star}\left(Z_j, p_j, Z_{-j}, p_{-j}\right) = \int_{\mathcal{X}} q_j^{\star}\left(X_i, Z_j, p_j, Z_{-j}, p_{-j}\right) \mathrm{d}F\left(X_i\right) \tag{30}$$

for a given menu of plans.

Insurers' problem

We now consider the optimal premium-setting problem of the insurer. We treat Z_j as an exogenous variable and focus on the choice of p_j.[10]

Claims and costs. The cost function $C\left(\cdot\right)$ depends on plan and patient characteristics—including ones that may be unobservable to the econometrician! Remember that just because an insurer cannot price on certain characteristics it does not mean that they are unaccounted for. We can write the expected cost of an individual with characteristics X_i, who has chosen plan Z_j as

$$c\left(X_i, Z_j\right) = \int_s e^{\star}\left(X_i, Z_j, s_i\right) \mathrm{d}F\left(s_i \middle| X_i\right) \tag{31}$$

where $F\left(\cdot\right)$ is the distribution of s_i. The cost of insuring a group of patients depends on their characteristics as a group. A common simplification here is to assume that we can categorize patients in a finite number of groups based on their characteristics, numbered 1 to K and indexed by k. Patients within a group are assumed to be homogeneous and have characteristics X_k. Then,

[10] This is a reasonable assumption in some regulated markets, where the governments specifies strict coverage parameters (e.g. the ACA health exchanges, or Medicare Advantage), but does not work in many others. See comments at the end of this section for more details.

the total cost of the plan is

$$C\left(Z_j, p_j, Z_{-j}, p_{-j}\right) = \sum_{k=1}^{K} q_j^{\star}\left(X_k, Z_j, p_j, Z_{-j}, p_{-j}\right) \times c\left(X_k, Z_j\right) \tag{32}$$

Equilibrium. Finally, to close the model, we solve the equilibrium condition from the problem of the insurance company. The insurance company chooses p_j to solve

$$\max_{p_j} p_j \times D_j^{\star}\left(Z_j, p_j, Z_{-j}, p_{-j}\right) - C\left(Z_j, p_j, Z_{-j}, p_{-j}\right) \tag{33}$$

which yields a first-order condition that delivers an exclusion restriction on the cost function (usually also with the help of a functional assumption). This restriction is

$$\frac{\partial D_j}{\partial p_j} p_j + D_j\left(p_j, p_{-j}\right) = \frac{\partial C}{\partial p_j} \tag{34}$$

If we compare equation 34 with equation 24, the crucial difference is that in equation 24, cost depends on price only through the effect that price has on quantity demanded. In other words, price only affects cost because higher prices means fewer units, hence lower costs. When selection effects are present, price changes not only leads to changes in the quantity demanded, but also in the characteristics of consumers. This creates an additional effect.

There are a few subtle issues here. Since cost is a function of price, second order conditions are more likely to be non-standard, and multiplicity is a bigger concern. Indeed, many problems of competition in insurance markets do not have Nash equilibria (this was known as early as Rothschild and Stiglitz, 1976). This can be a pretty big problem because it means that multiple vectors of prices can satisfy the FOC.[11] Occasionally, you can get around these issues by looking for alternative equilibrium conditions (see e.g. Handel et al., 2015, which uses the notion of Riley equilibrium). In other settings, we have universal mandates (for example within the universe of a single large firms that self-insures) and the relevant condition is that price is equal to *average* cost.

Finally, the vector Z_j is also a choice variable in many markets. So one should ask why contracts Z_j were offered in the first place. This is a common critique to most papers that measure adverse selection: the only selection you can see is the one in markets that exist, and in contracts that exist. That does not mean that in absence of adverse selection we would not have seen different contracts. This is one of the questions at the frontier of the field. Ho and Lee (2020) is one of the most recent papers on this topic.

[11] This does not actually matter too much in estimation, but it invalidates most attempts at counterfactual analysis.

3 BIBLIOGRAPHY

Akerlof, George A., “The market for "lemons": Quality uncertainty and the market mechanism,” *Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro, “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *The Quarterly Journal of Economics*, 2017, *132* (4), 1553–1592.

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum, “Heterogeneous Choice Sets and Preferences,” 2020.

Cardon, James H. and Igal Hendel, “Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey,” *The RAND Journal of Economics*, 2001, *32* (3), 408.

Cutler, David M., Amy Finkelstein, and Kathleen McGarry, “Preference heterogeneity and insurance markets: Explaining a puzzle of insurance,” *American Economic Review*, 2008, *98* (2), 157–162.

_ and Sarah J. Reber, “Paying for health insurance: The trade-off between competition and adverse selection,” *Quarterly Journal of Economics*, 1998, *113* (2), 433–466.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen, “Estimating Welfare in Insurance Markets Using Variation in Prices,” *The Quarterly Journal of Economics*, 2010, *125* (3), 877–921.

_ , _ , Iuliana Pascu, and Mark R. Cullen, “How general are risk preferences? choices under uncertainty in different domains,” *American Economic Review*, 2012, *102* (6), 2606–2638.

_ and _ , “Selection in insurance markets: Theory and empirics in pictures,” *Journal of Economic Perspectives*, 2011, *25* (1), 115–138.

Handel, Ben, Igal Hendel, and Michael D. Whinston, “Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk,” *Econometrica*, 2015, *83* (4), 1261–1313.

Handel, Benjamin R., “Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts,” *American Economic Review*, 2013, *103* (7), 2643–2682.

_ and Jonathan T. Kolstad, “Health insurance for "humans": Information frictions, plan choice, and consumer welfare,” *American Economic Review*, 2015, *105* (8), 2449–2500.

Ho, Kate and Robin Lee, “Health Insurance Menu Design for Large Employers,” 2020.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz, “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment,” *American Economic Review*, 1987, *77, No 3* (May), 251–277.

Rothschild, Michael and Joseph Stiglitz, "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information," *Quarterly Journal of Economics*, 1976, *90* (4), 629–649.

Shepard, Mark, "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange," 2020.