sdsim.R1
24-March-2022

Supply, Demand and Selection in Insurance Markets: Theory and Applications in Pictures

ABSTRACT:

This paper recapitulates and extends the Einav, Finkelstein and Cullen (2010; EFC) supply and demand approach to modeling equilibrium in insurance markets. After describing the basic EFC model, we show that it can be extended to: markets featuring non-exclusive contracting and linear pricing; markets featuring selection on ex-ante moral hazard; and multiple-margin markets with price competition over two coverage tiers. In markets with non-exclusive-contracting-cum-linearly pricing, we show that the EFC model, graphs, and lessons all go through, essentially unchanged. In markets with selection on moral hazard, we show that two natural notions of adverse selection—notions that coincide in the basic EFC framework—may diverge. In a two-tier market, we extend Weyl and Veiga (2017) by showing that the lessons depend on whether contracts are offered "simultaneously" (as in Geruso et al., 2021) or "sequentially," where coverage is purchased in layers.

Supply, Demand and Selection in Insurance Markets:
Theory and Applications in Pictures

1. Introduction

Supply and demand analysis of competitive markets is one of the most basic tools of economic theory. One of the assumptions of the theory of competitive markets is that market participants have complete and symmetric information. Insurance markets frequently have incomplete and asymmetric information. Rothschild and Stiglitz (1976; RS) and Akerlof (1970), two of the seminal papers in the early literature on markets with asymmetric information, took radically different approaches to modeling equilibrium, with Akerlof sticking to a supply-demand framework (and describing the pathologies that emerge) and RS taking a contract-theoretic/endogenous contract approach. Until quite recently, the RS approach has been the predominant paradigm for insurance market research.[1] Einav, Finkelstein and Cullen's (EFC, 2010) and Einav and Finkelstein's (EF, 2011) elaboration of Akerlof's model has demonstrated the practicality of employing supply-demand analysis in insurance markets and opened up new research avenues.

The basic EFC approach takes as given an exogenously fixed insurance policy and assumes that firms compete on price in a perfectly competitive market for this policy. EFC show that price, quantity, and cost data can be used jointly to estimate the demand and cost curves for insurance. This allows for a direct test for selection and for straightforward estimates of the welfare cost of asymmetric-information-driven selection. Their approach has been widely used in analyzing selection markets, primarily in health insurance. Examples include Bundorf, Levin and Mahoney (2012), Einav, et. al, (2013), Einav Finkelstein and Schrimpf (2010), Finkelstein, Hendren and

[1] Although see Johnson (1977) who uses supply and demand to analyze a market with adverse selection.

Shepard (2019), Geruso, et. al (2021), Hackmann, Kolstad and Kowalski (2015), Handel, Hendal and Whinston (2015), Hendren (2013), Hendren (2019), Panhans (2018), and Weyl and Viega (2017), among others.[2] The purpose of this paper is to show that the supply and demand approach can be readily applied to analyze outcomes and policy interventions in a wide range of selection markets. We do so recapitulating the basic EFC framework and then describing how their basic supply and demand analysis can be extended to markets with features not previously analyzed in the literature.

In Section 2, we review the basic EFC framework. A key feature of this model is that it provides a simple way to characterize "adverse" versus "advantageous" selection. Specifically, a market is said to display adverse selection if the marginal cost curve (the resource cost of insuring additional individuals) is below the average cost (the cost to firms per insured individuals), or, equivalently, if the average cost curve is downward sloping. Conversely, a market is said to display advantageous selection if the average cost curve is upward sloping.

In Section 3, we show that this framework readily extends to markets with *ex ante* moral hazard (and selection thereon). In this extension, individuals may take unobservable (or non-contractible) actions to reduce the probability of a loss, the size of the loss, or both, and individuals differ in their relative cost (or effectiveness) of effort. These differences drive selection effects: individuals who find it more costly to exert effort find insurance more appealing. This cost heterogeneity implies a standard down-sloping demand curve for insurance. Moreover, since those individuals who buy insurance at high prices are more costly to insure than are those who buy insurance only at lower prices, the market features adverse selection at the individual level. Nevertheless, we show that this adverse selection at the individual level may not translate into

[2] Mahoney and Weyl (2017) similarly use graphical arguments and price-theoretic reasoning in a model of imperfect competition in selection markets. They emphasize the tradeoff between selection and market power.

adverse selection in the sense of EFC – i.e., at the market level. Specifically, we show that, despite the adverse selection at the individual level, the market-wide marginal cost curve may nevertheless lie above the market-wide average cost curve, and the latter may therefore slope upward. This apparent disconnect arises because the self-insurance or self-protection efforts of inframarginal buyers are endogenous to the price of insurance. As the premium falls (as it must to attract additional buyers) the inframarginal buyers exert less effort, raising the cost to insurers. We show, by means of an example, that this endogenous effort effect can outweigh the fact that the marginal buyer is less costly than the average buyer and can cause the average cost curve to be upward sloping. In other words, in markets with ex-ante moral hazard, advantageous selection in the EFC sense can arise even when individuals self-select into insurance purchases in a manner consistent with the common-sense notion of "adverse selection."

In Section 4, we then extend the EFC model to non-exclusive contracting cum-linear pricing, and show that essentially all of the insights of the basic EFC model go through unchanged. This extension allows us to apply the supply and demand approach, for example, to markets for annuities and life insurance, where contracting is typically non-exclusive and pricing is approximately linear.

Finally, in Section 5, we analyze an extension of EFC to markets in which there are two distinct exogenously fixed insurance policies that individuals can buy and over which firms compete. We show that the welfare effects of a subsidy in such markets hinge critically on the nature of the competition – effectively extending the basic one-margin conclusions of Weyl and Veiga (2017) to the case, a la Geruso et al (2021), where there are two choice margins for insurance consumers.

We first analyze a situation in which contract choice is "sequential." In this setting, coverage is layered: a consumer must purchase the primary policy in order to purchase a secondary "top up" policy. The canonical example of a sequential market is the market for Medigap insurance, which provides supplemental coverage over and above basic Medicare coverage. Earthquake coverage on top of a basic homeowner's policy is another example. In such a "sequential" setting, subsidies for basic coverage relieve adverse selection inefficiencies by improving take-up of the basic coverage, and they typically do not have any effect on the market for supplemental coverage.

We then analyze a situation in which contract choice is simultaneous, recapitulating Geruso, et. al. (GLMS, 2021). In such a simultaneous market, consumers make a binary choice over which of two policies, if any, to buy. The canonical example of simultaneous choice is an ACA exchange, where individuals can choose between (e.g.) a Silver or a Bronze plan. The choice between a low versus a high deductible policy in auto or home insurance also fits within this basic framework. As discussed in GLMS, subsidies for basic coverage in this setting improve takeup of insurance *overall*, but, in contrast to the sequential case, they also impact the market for generous coverage. Indeed, subsidies aimed at improving adverse selection on the "takeup" margin have the perverse effect of exacerbating adverse selection in the market for generous coverage.

Section 6 provides brief concluding remarks.

2. Supply and Demand in Selection Markets

In this section we review the supply and demand framework with exclusive contracting developed in EFC and EF. Insurers are risk-neutral expected profit maximizers and—contra the RS model—are assumed to offer a single exogenously fixed contract offering some fixed coverage. Risk averse consumers make a binary choice to either buy the insurance policy or not. Consumer heterogeneity is characterized by the parameter $\theta \in [0, \bar{\theta}]$ which is private information. We assume θ has distribution $F(\theta)$ and density $f(\theta)$.[3]

Each type θ is associated with a maximum willingness to pay for the contract, $\hat{p}(\theta)$, and an expected cost $c(\theta)$ to firms who provide them with insurance. Without essentially no loss of generality, we assume that $\hat{p}$ strictly decreasing in θ. The market is assumed to be perfectly competitive, with price-taking firms and consumers; we elaborate below.

Since consumers face a buy/don't buy decision, the aggregate quantity coincides with the fraction of the population that is covered. Letting $\hat{\theta}(P)$ be the largest value of θ such that the consumer buys the contract when the market price is P, the market quantity demanded is

$$Q(P) = \int_0^{\hat{\theta}(P)} dF(\theta) = F(\hat{\theta}(P)). \qquad (2.1)$$

The height of the market demand curve in (Q, P) space—depicted in both Figures 2.1 and 2.2, and necessarily downward sloping—reflects the willingness to pay of the marginal insurance buyer, i.e., the willingness to pay of the Qth most willing to pay buyer in the market.

Now suppose that the expected cost of coverage, $c(\theta)$, is decreasing in θ. Since the individuals with the highest willingness to pay are, in this case, the individuals with the highest expected cost, the marginal insurance buyer's cost to firms is lower than the average cost of insuring the inframarginal buyers. We then say that there is adverse selection. Figure 2.1 illustrates this case; it features the downward sloping marginal cost and average cost curves implied by a decreasing $c(\theta)$ curve.

It is instructive to consider the special case of the binary loss framework that has been a staple in insurance market research following RS. In this special case, we use θ to measure the

[3] We focus on this one-dimensional type case, but the EFC approach readily extends to allow for multiple dimensions of heterogeneity.

probability that a consumer will *not* experience a loss of size l out of wealth w, consumers are assumed to be expected utility maximizers with utility function $u(\cdot)$, and $c(\theta) = (1-\theta)l$ (i.e., there are no administrative or other costs). If the insurance policy provides full insurance, then $\hat{p}(\theta)$ solves $u\big(w - \hat{p}(\theta)\big) = \theta u(w) + (1-\theta)u(w-l)$. Hence, $\hat{p}(\theta)$ and $c(\theta)$ are decreasing in $\theta.$ Moreover, if consumers are risk averse—i.e., if u is strictly concave— then $\hat{p}(\theta) > c(\theta)$, and the demand curve always lies above the MC curve. Then it would be socially efficient for all consumers to be covered.

Towards understanding the supply side of the market, it is worth first recapitulating textbook model of perfect competition. In this textbook model, firms are price takers and believe that they can sell as many units of the good as they want at that taken price. If their marginal cost curve is downward sloping, they optimally choose to produce at the unique quantity where their marginal cost is equal to this price. Their supply curve thus coincides with their marginal cost curve. The same is true if their marginal cost curve is horizontal: they will then produce zero for prices below marginal cost, an infinite quantity for prices above it, and will be equally willing to produce any quantity if the price is equal to their marginal cost. As such, their perfectly elastic supply curve again coincides with their horizontal marginal cost curve.

The same basic logic goes through in the insurance context *except* that the marginal cost to a firm now depends on which buyers buy the good. For any given price P, firms (in this case insurers) still believe that they can sell as many units (of insurance) as they wish at that given price. If they make a sale, it will be to a buyer who is willing to buy at the given price – but the firm will not be able to tell which one. As such, the marginal cost *to the firm* of making an additional sale is a random draw of the cost associated with a consumer drawn at random from the set of all buyers who would be willing to buy at P. The expected marginal cost to a firm of an additional sale is thus the average across all of the buyers who are "in the market" at price P – i.e., the average cost associated with the $Q(P)$ buyers who would buy at P. In short, from the *firm's* perspective, each additional buyer is a random draw from the set of individuals who are willing to buy at the market price, and each firm's (expected) marginal cost curve is horizontal at a height equal to the height of *market's* average cost curve $AC(Q) \equiv \int_0^{F^{-1}(Q)} c(\theta)f(\theta)d\theta$.

There are two immediate implications of the observation that the market AC coincides with the firm's expected marginal cost curve. First, since the gap between $AC(Q)$ and $MC(Q)$ is equal to the gap between the private (firm level, expected) marginal cost and the social (market level) marginal cost curve, the observation implies that the gap can also be interpreted as an externality – in this case, an informational externality. Second, it implies that we can think of the average cost curve as the market *supply* curve. To see why, fix any market quantity Q^S. Then the height of the AC curve gives the expected marginal cost that a given firm perceives they will bear if they choose to sell an additional insurance contract. It thus the unique price at which the market is willing to provide Q^S units.

As in standard (information-externality-free) markets, market equilibrium price and quantity are determined by the intersection of supply (i.e., the average cost curve) and demand curves. This is the point C in Figure 2.1. Since price equals average cost at this point, firms earn zero profits. When the AC curve lies above the MC curve—i.e., when there is adverse selection everywhere, as in Figure 2.1—the AC curve intersects the demand curve at a quantity Q_{eq}, which is strictly less than the efficient quantity (Q_{max} in Figure 2.1). Because of the information externality, the individuals between Q_{eq} and Q_{max} are not insured. This under-provision of insurance is the well-known inefficiency associated with adverse selection. The welfare loss from this under-provision can be quantified via the deadweight burden quadrilateral $CDFG$ in Figure

2.1—i.e., the integral of the gap between the social benefit of insurance, as measured by the demand curve, and the social cost of insurance, as measured by the MC curve.

This welfare loss can be eliminated with an appropriate (Pigouvian) subsidy on insurance purchases. Specifically, a subsidy equal to the difference between average cost and wiliness to pay at Q_{max} will shift the demand curve up to the point where it intersects the AC curve at Q_{max} and restore efficiency. An alternative strategy to increase the proportion of the population covered is to impose a mandate for non-purchase with a penalty of the same size as the optimal subsidy. (In the case illustrated in Figure 2.1, where full coverage is optimal, any penalty above this level will also suffice.) These corrective interventions are illustrated in Figure 2.1. The amount of the subsidy or mandate is EG. The effect is to increase welfare by $CDFG$.

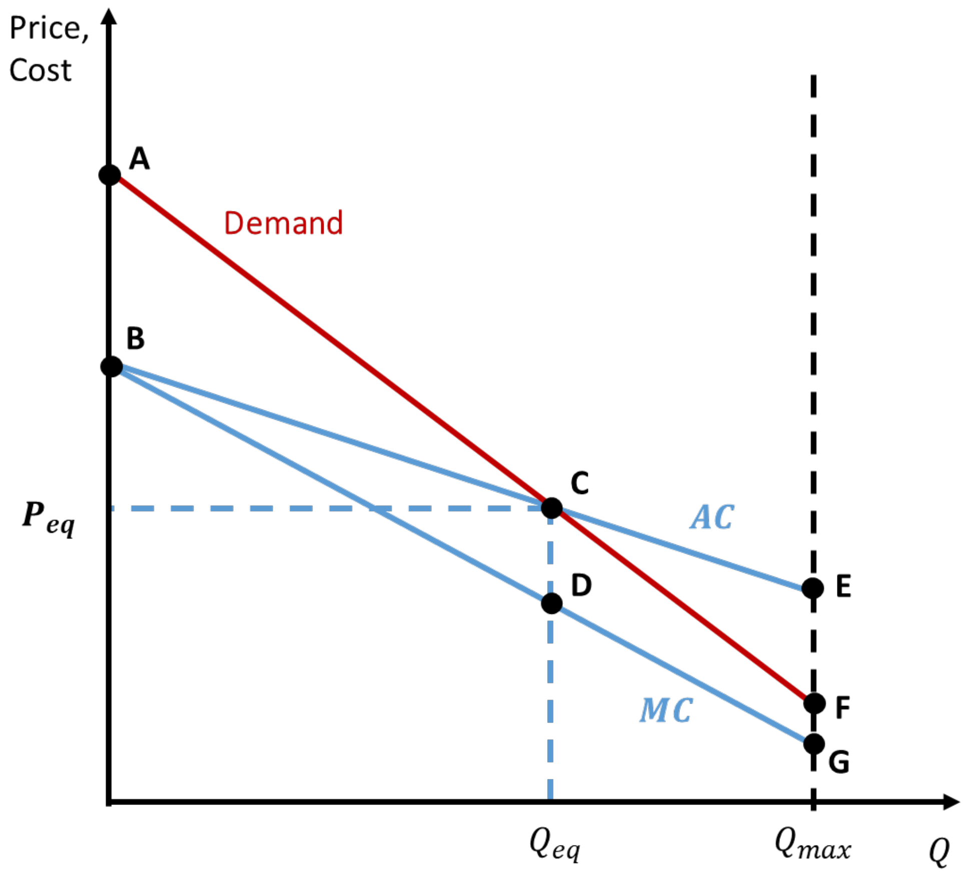
In the adverse selection case, there is a positive correlation between coverage and cost, since the average cost of the individuals who are covered is higher than the average cost of the individuals who are not covered. The positive correlation property is the basis of many empirical tests for adverse selection (e.g. Chiappori and Salanie, 2000, Cohen, 2005, Fang, 2008).

The analysis of the advantageous selection case, where $c(\theta)$ is instead *increasing* in θ, as illustrated in Figure 2.2, is entirely symmetric.[4] The efficient level of coverage, Q_{eff}, is given by the intersection of the demand curve and the MC curve (point C in Figure 2.2). The equilibrium price and level of coverage is given by the intersection of the demand curve and the AC curve, at point E in Figure 2.2. Since Q_{eq} is greater than Q_{eff}, advantageous selection leads to the over-provision of insurance. The welfare cost from the over-provision of insurance is the triangle CDE.

[4] The concept of advantageous (or propitious) selection was introduced by Hemenway (1990, 1992), and developed analytically by De Meza and Webb (2001) and a large subsequent literature. Advantageous selection could arise, for example, if intrinsic cautiousness is heterogeneous. Cautiousness is likely to be positively associated with perceived value of insurance and negatively associated with the likelihood and severity of accidents. So those who are more likely to purchase insurance will also, through this cautiousness channel, be less costly to insurers.

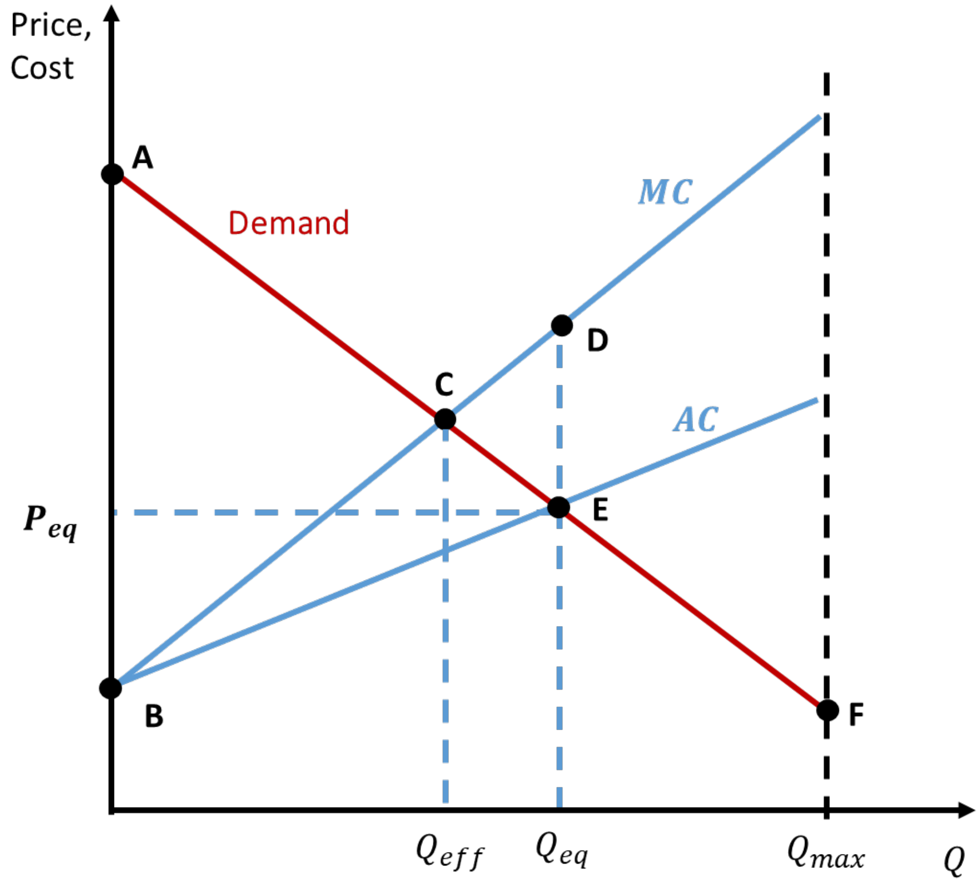
This welfare cost can be eliminated with a tax on purchase (or a subsidy on non-purchase) equal to the distance between the MC and demand curves at Q_{eff}.

Figure 2.1: Adverse Selection with Exclusive Contracts



Source: Adapted from Einav and Finkelstein (2011)

Figure 2.2: Advantageous Selection with Exclusive Contracts



Source: Adapted from Einav and Finkelstein (2011)

As EF point out, there are a number of important caveats. If willingness-to pay is large enough, then the equilibrium outcome is universal coverage. On the other hand, if willingness-to-pay is low enough the demand curve lies everywhere below the average cost curve and the market completely unravels—and this complete unravelling can occur even when it would be socially efficient for all individuals to have coverage.[5] Their point is that the welfare cost of adverse selection can vary widely. Even leaving aside these extreme cases, the magnitude of the welfare loss is sensitive to location and shape of the demand and cost curves.

Because of this, there are no simple, universal fixes. In canonical model of insurance, agents are all risk averse and are therefore willing to pay more for insurance than the cost of providing it. In such a world, an insurance mandate or government-provided universal coverage is social welfare maximizing. But if there are costs associated with selling insurance (such as administrative and claim processing costs), or if some individuals are unwilling to pay an actuarially fair price for insurance coverage (e.g., because the contract for sale is poorly designed, or because the individual is a risk seeker or faces liquidity constraint), then universal coverage may no longer be efficient. Under adverse selection there is still under-provision of insurance, but mandatory insurance may not even be welfare improving.

3. Selection on Moral Hazard

In this section we analyze general models with ex-ante moral hazard and show that the EFC analysis extends easily to these cases—with the caveat that modelers needs to be cautious in how they interpret "adverse selection." Throughout, we use the EFC framework where firms

[5] This observation was first made by Akerlof (1970). Hendren (2013) provides a simple sufficient condition for complete unraveling in the one-dimensional heterogeneity case. Hendren (2014) uses this condition to show that in a continuous-type generalization of the RS model, unraveling in the sense of Akerlof occurs if and only if an RS equilibrium exists: if an equilibrium in the sense of RS exists, then it must involve no individuals purchasing any insurance. For the case where losses exceed wealth, see Posey and Thistle (2019).

compete on the price p of a single exogenously fixed insurance contract which indemnifies a fraction α of an insurable loss l. Individuals take an unobservable or non-contractible action (or, in a trivial extension, a vector of actions) e which, depending on the setting, may affect the probability of a loss, the size of an insurable loss, or both. Individuals again differ in their type θ, again distributed with CDF F. Now, however, θ instead parameterizes the cost to the potential insurance buyer of choosing action e—with higher θ types having lower effort costs.

We first consider the classic binary risk setting of RS where there is an exogenously fixed (binary) loss out of initial wealth w. We allow both the size of the loss $l(e)$ and the probability of loss $\pi(e)$ to depend on the level of effort e. We take $\pi(\cdot)$ and $l(\cdot)$ to be decreasing and strictly convex. Specifically, preferences are given by:

$$U(e;\ \theta, \alpha, p)\ =\ (1 - \pi(e))u(w - p)\ + \pi(e)u(w - p - l(e)\ +\ \alpha\, l(e)) - (1 - \theta)e, \qquad (3.1)$$

where u is strictly increasing and strictly concave. In other words, individuals are risk averse expected utility maximizers over realized wealth and bear a (nonpecuniary) utility cost $(1 - \theta)e$ to reduce the accident risk. We discuss below what changes if the costs are instead pecuniary—e.g., associated with the purchase of safety goods.

Individuals can choose whether to buy insurance ($b = 1$) or not ($b = 0$), which we denote by the binary variable $b \in \{0,1\}$, and also can choose their level of effort e. Define $V(e, b;\ p, \theta) = U(e, p; \theta, \alpha)$ if $b = 1$ if the individual is insured, and $V(e, b;\ p, \theta,) = U(e, 0; \theta, 0)$ if $b = 0$ and the individual is uninsured. Then an individual's problem is to maximize V by choosing (e, b).
It turns out that this problem is extremely well-behaved. Straightforward calculations show that $V_{ep} \geq 0$ (and strictly so if $b = 1$), $V_{e\theta} > 0$, and $V_{p\theta} = 0$, so we have pairwise increasing (non-decreasing) differences in all pairs of the three continuous variables. Towards showing that we also have pairwise increasing differences of each of these three variables with the binary variable

b, define $\Delta V \equiv V(e, 0; p, \theta) - V(e, 1; p, \theta)$. Then $\Delta V_\theta = 0$ (trivially). Since $U_p(e, p; \theta, \alpha) < 0$, we have $\Delta V_p > 0$. Finally, differentiation and some algebra shows that $\Delta V_e > 0$ – which is intuitive, since the marginal benefit of loss and accident prevention efforts are smaller when insured than when uninsured. These features, taken together, imply that higher premiums (at least weakly) raise optimal efforts for all types θ and reduce the set of individuals who choose to buy insurance.[6]

An implication is that each type has a cutoff premium $\hat{p}(\theta)$ for which $b(\theta; p) = 1$ if $p < \hat{p}(\theta)$ and $b(\theta; p) = 0$ if $p > \hat{p}(\theta)$. For any premium there is a cutoff type $\hat{\theta}(p)$ below which all individuals purchase insurance and above which none do, and lower θ types put in less effort. In short: this is a model with "selection on ex-ante moral hazard," and it superficially appears to be a clear case of adverse selection, since the marginal buyers are lower risk than the inframarginal buyers.

In spite of these clean comparative statics, however, the market as a whole will *not* necessarily display "adverse selection" in the sense of EFC. That is, the marginal cost curve may lie above the average cost curve, and the later, consequently, may be upward-sloping *in spite* of the straightforward comparative statics suggesting adverse selection on moral hazard.

---

[6] Formally: these features imply that $V(e, b; p, \theta)$ is supermodular in $(e, -b, p, \theta)$. Since $V(e, b; p, \theta)$ is supermodular, Topkis's Theorem applies, and implies that the optimal choices $e^*(\theta; p)$ and $-b^*(\theta; p)$ are increasing in p and θ. See Milgrom and Shannon (1994) or Amir (2005) for treatments of supermodularity and Topkis's Theorem.

Figure 3.1: the average cost curve can be up or down-sloping

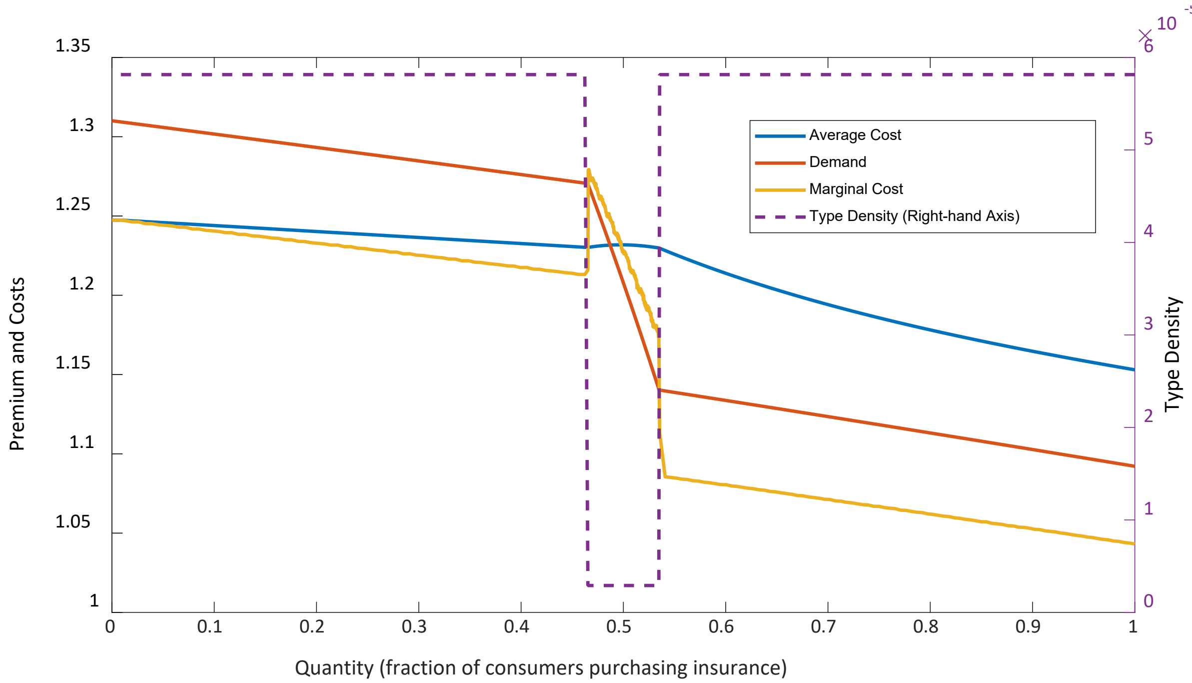


Figure 3.1 provides an illustrative example. It plots the demand and average cost curves for a simple case with log utility, $w = 9$, $l(e) = 7.5$, $\alpha = 6.2/7.5$, $\pi(e) = \frac{1}{e+2}$ and a stepwise uniform type distribution which has support on the interval [0.01,0.015] and which has with a density that is 1/20th as large on the sub-interval [0.011,0.014] as it is on the intervals [0.01,0.011] and [0.014,0.015] at either end of the support. (The dashed curve in Figure 3.1 plots the local density of types on the right-hand-axis.) As the diagram shows, the average cost curve is (modestly) up-sloping at intermediate quantities.[7]

[7] We generated this plot in Matlab as follows. First, we created three grids (vectors): a grid (vector) of types θ uniformly spaced on the interval [0.1,0.15] (and a vector of probability masses proportional to the densities f for each type in this grid), a grid of effort levels, and a grid of prices. Second, for each type, we computed the reservation utility associated with forgoing insurance. Using a binary search routine and numerical optimization over the effort grid, we then used the reservation utilities to compute each type's reservation price $\hat{p}(\theta)$ – i.e., the price at which the optimized-over-effort expected utility of purchasing insurance at that price would yield exactly the reservation utility. Third, for each price p, we computed the set of types with $\hat{p}(\theta) > p$, the total mass $Q(p)$ of those types (the demand), and the total costs $TC(p)$ and average cost $AC(p)$ associated with selling to that set of types. Finally, we used $TC(p)$ and $Q(p)$ to numerically approximate the marginal cost $MC(p)$. The AC, MC, and

The general intuition for this up-sloping portion is that there are two competing effects of selling an additional policy. The first is the standard "adverse selection" effect: the marginal insurance buyer is a higher-θ type than the average buyer. The marginal buyer therefore chooses a higher e^* than the average buyer, and hence lowers the average risk in the market. The second is a wealth effect that is due to ex-ante moral hazard and hence is absent from standard treatments in the literature following EF and EFC (which typically deal exclusively with selection markets and ex-post moral hazard). It arises because selling additional policies requires offering a lower price, and lower prices lower the preventative efforts of the inframarginal buyers. This inframarginal effect works to offset the effect of the marginal buyer on marginal cost.

The upward-sloping portion of the AC curve in Figure 3.1 shows, by example, that the inframarginal effect can dominate. The specific intuition for this example is as follows. As illustrated in the figure, the density of types in the middle is very low. This makes the demand curve slope downward steeply (as a decrease in price only brings a small mass of individuals into the market). So selling the marginal policy leads to a large price reduction and therefore induces a large negative effort response of the inframarginal types.

An alternative formulation of utility is the well-known Ehrlich-Becker (1972) model

$$U(p,e;\,\theta,\alpha) = \big(1-\pi(e)\big)u(w-p-(1-\theta)e) + \pi(e)u(w-\,p-\,l\,(e)+\,\alpha\,l(e)-(1-\theta)e), \qquad (3.2)$$

wherein individuals bear a monetary (instead of a utility) cost of investing in self-protection or self-insurance.[8] Comparative statics are more challenging in this formulation (formally: this is

demand curves in Figure 3.1 plot the loci of $\big(Q(p),AC(p)\big)$, $\big(Q(p),MC(p)\big)$, and $(Q(p),p)$ pairs over the elements of the p grid.

[8] Typical formulations re-parameterize preferences via the monetary investment $x\equiv(1-\theta)e$ and interpret $\frac{1}{1-\theta}$ as the "efficiency" of investment. The formulation here is equivalent but reparameterized in term of effective risk reduction, *e*. Since some people are less efficient than others, they need to spend more ((1-θ)*e*) to get the same effective risk reduction.

because preferences will no longer be supermodular), but this only makes the ambiguity in the sign of the slope of the average and marginal cost curves starker. Indeed, it is not hard to find examples with average cost curves that are either up or down-sloping (and, indeed, it is not hard to find examples with ugly comparative statics).

4. Non-Exclusive Contracts

Exclusive contracting is a natural assumption for many types of insurance, for example, health insurance and most property and liability insurance. It is a less reasonable assumption for life insurance, where individuals may buy multiple policies from multiple insurers, and, particularly in the case of annuities, pricing is approximately linear (*viz* Finkelstein and Poterba 2002). Insurers in a non-exclusive setting know how much coverage they have sold to an individual, but typically will not know the total amount of coverage the individual has bought. We follow a long literature in modeling such non-exclusive insurance markets by assuming that competition in such non-exclusive settings implies that insurers charge the same price, p, for each unit of coverage—i.e. that pricing is linear.[9] In these markets, individuals do not make a binary decision to buy coverage or not. Rather, individuals decide how much coverage they want to buy, at a unit price over which completion occurs.

The basic reason for using linear-prices to model non-exclusive insurance is simple: the logic of adverse selection in the Rothschild-Stiglitz pushes towards convex pricing schedules, where more generous policies have higher per-unit pricing. With non-exclusive contracting, however, individuals can circumvent convex pricing by purchasing multiple smaller policies from

[9] See, e.g., Abel, 1986, Villeneuve, 2003, Rothschild, 2015 in the annuity context, Hoy and Polborn (2000) and Polborn, Hoy, and Sadanand (2006) in the life insurance context, and Finkelstein, 2004 in the health insurance context. See in particular the discussion in Chiappori (2001).

many different insurers. If the "smallest" policy is sufficiently small, then the *effective* price schedule faced by a given buyer is linear.[10]

We assume each individual has a demand for insurance $q(p,\theta)$ that is decreasing in the price, p. Consumers may buy $q(p,\theta)$ from one firm or from several firms. The expected cost per unit of coverage for a consumer of type θ is $c(\theta)$, so that the total cost of selling q units to a θ type is $qc(\theta)$. The total demand for insurance at price p is

$$Q(p) = \int_0^{\overline{\theta}} q(p,\theta)dF(\theta). \tag{4.1}$$

The inverse demand curve is $P(Q)$. The total cost is

$$TC(Q) = \int_0^{\overline{\theta}} c(\theta)q(P(Q),\theta)dF(\theta). \tag{4.2}$$

The marginal cost is $MC(Q) \equiv \frac{dTC(Q)}{dQ}$, and the average cost is $AC(Q) = \frac{TC(Q)}{Q}$. As in the exclusive contracting case, and for the same reason, the average cost curve is the supply curve.

As in the exclusive contracting case, we define adverse (advantageous) selection as a case where the marginal cost curve lies below (above) the average cost curve, so that the latter is downward (upward) sloping. A sufficient condition for there to be adverse selection is an everywhere-decreasing marginal cost. Then:

$$AC(Q) = \frac{1}{Q}\int_0^Q MC(Q')dQ' > \frac{1}{Q}\int_0^Q MC(Q)dQ' = MC(Q). \tag{4.3}$$

There is a positive correlation between q and c if and only if the covariance

$$\int_0^{\overline{\theta}} q(p,\theta)(c(\theta)-\bar{c})\,dF(\theta) = \int_0^{\overline{\theta}} q(p,\theta)c(\theta)dF(\theta) - \bar{c}Q = Q(p)(AC(Q(p)) - \bar{c}) \tag{4.4}$$

[10] It is far from clear that this linear pricing assumption is ever a well-founded approximation of outcomes in competitive, non-exclusive markets – though it is, as noted a good description of actual pricing in annuity markets. In small-type-space models a la Rothschild and Stiglitz (1976), it does not seem to be well founded. See Attar, Mariotti, and Salanie (2014) and Dubey and Genakoplous (2019), who show that pricing is *not* linear in such models.

is positive (where $\bar{c} = \int_0^{\bar{\theta}} c(\theta)dF(\theta)$ is the average cost of coverage for the population as a whole). If the average cost curve is everywhere downward sloping, then we have $AC(Q) \geq \bar{c}$ for all Q. So an everywhere-downward-sloping average cost curve is sufficient for the positive correlation property to hold. The obverse is also true: an upward-sloping marginal cost curve implies an upward-sloping average cost curve, and hence a negative correlation between coverage and cost. In short, the relationship between marginal costs, average costs, and the positive correlation property are just as in the exclusive contracting case: if the market features a decreasing (increasing) marginal cost curve, then it also features a downward (upward) sloping average cost curve. And a downward (upward) sloping average cost curve implies that the positive (negative) correlation property also holds.

The case of adverse selection is illustrated in Figure 4.1. The only substantive difference vis-à-vis the exclusive contracting analog (Figure 2.1) is that, with non-exclusive contracts, the maximum quantity Q_{max} is determined by the total quantity demanded $Q(0)$ at a price of 0, rather than by the exogenously given population size. The socially efficient level of coverage, Q_{eff}, is where the MC curve intersects the demand curve, i.e. point F in Figure 4.1. Note that since marginal costs are strictly positive, it is necessarily the case that $Q_{eff} < Q_{max}$ – contra Figure 2.1, where the efficient quantity coincided with Q_{max}. As in Figure 2.1, the equilibrium price P_{eq} and quantity Q_{eq} are determined by the point (C in figure 4.1) where the AC curve intersects the demand curve, i.e., where the price is such that firms break even given the buyers who purchase at that price. Since there is adverse selection, the marginal buyer is less costly for firms than the inframarginal buyers are, and the AC curve lies strictly above the MC curve. It follows that $Q_{eq} < Q_{eff}$, and, as with exclusive contracting, there is a welfare cost due to the under-provision of insurance, equal to triangle CDF in Figure 4.1. As in the case of exclusive contracting, a subsidy can be used to increase welfare; specifically, a subsidy equal to the length EF in Figure 4 would implement the efficient level of coverage Q_{eff}. In principle, a mandate could also be used to increase welfare in this setting, but the fact that different individuals purchase different quantities of insurance complicates the practical implementation of – and the graphical depiction of – such mandates.

The case of advantageous selection is illustrated in Figure 4.2. Here the MC curve lies above the AC curve. As in the exclusive contract market, the equilibrium quantity, Q_{eq} is greater than the efficient quantity Q_{eff}. The welfare loss due to the over-provision of insurance is given by the triangle CDE in Figure 4.2.

Figure 4.1: Adverse Selection with Non-exclusive Contracts

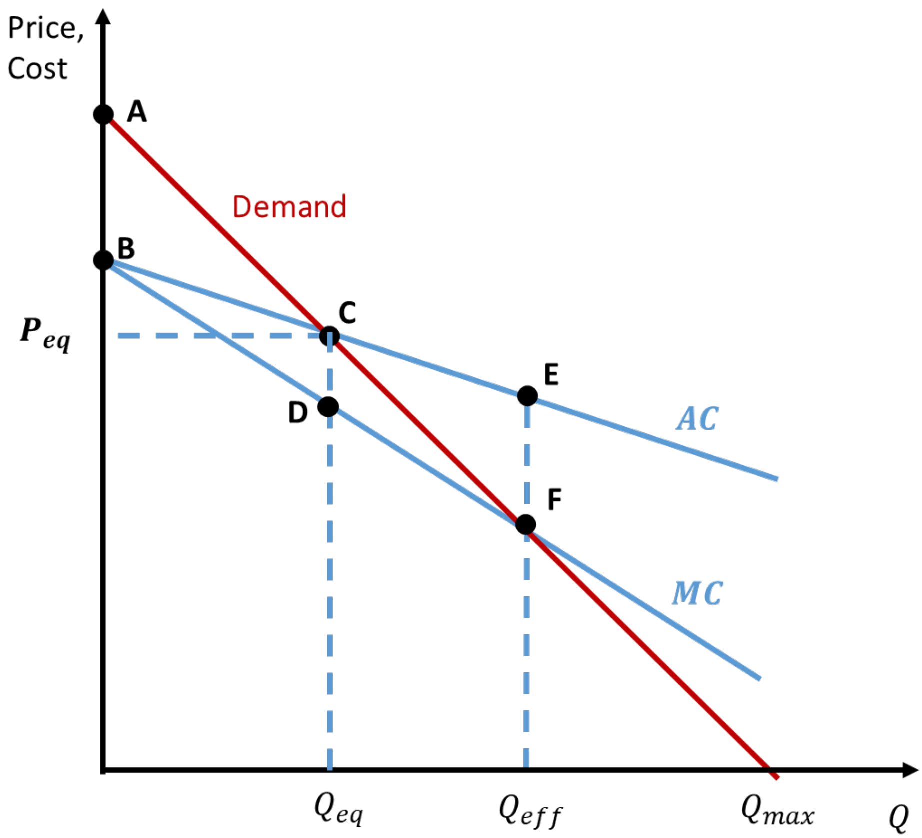
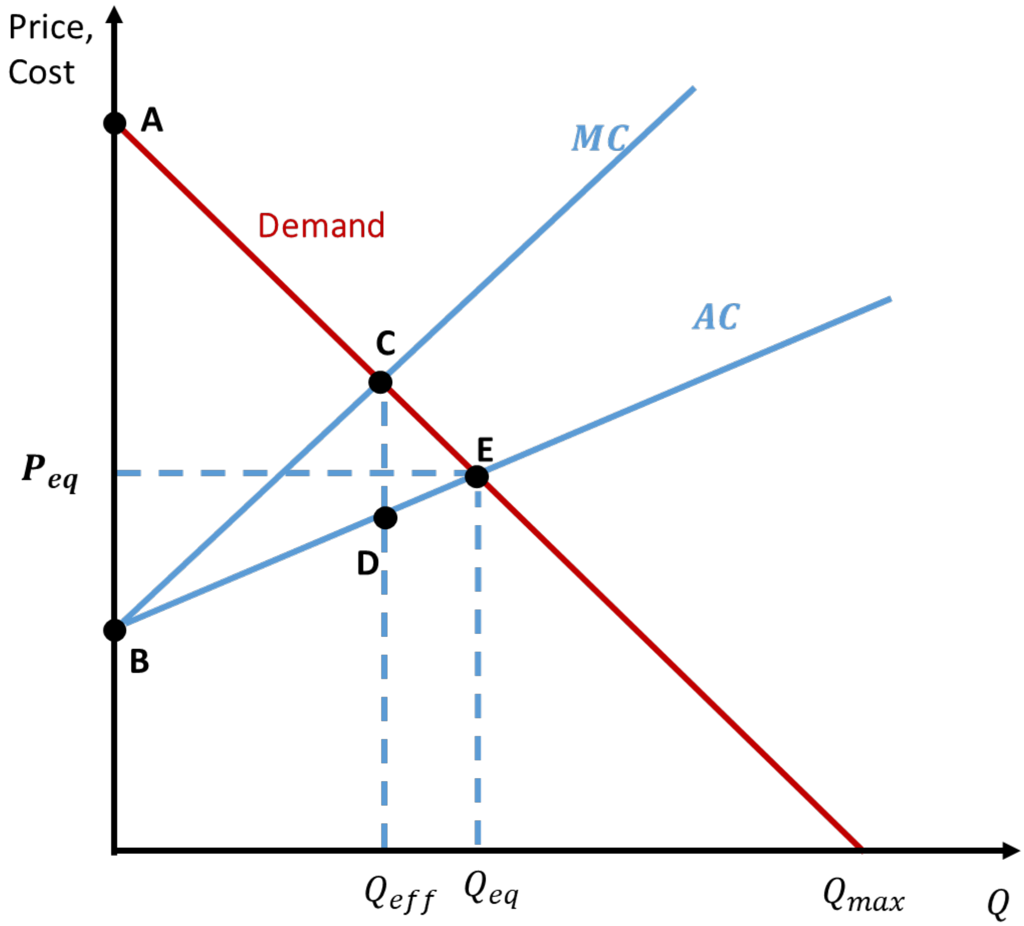


Figure 4.2: Advantageous Selection with Non-exclusive Contracts



It is straightforward to combine the preceding analysis of non-exclusive insurance markets with the moral hazard analysis of Section 3. Specifically, with non-exclusivity-cum-linear-pricing, one can treat α in equation 3.1 as a continuous choice variable which takes the place of the binary buy/don't buy decision in Section 3's analysis. It is not, in general, possible to get simple comparative statics in this model. But, as in Section 3, it is straightforward to come up with numerical examples where there is adverse selection at the individual level, but where the average cost curve is upward sloping so there is not adverse selection at the market level. That is: it is possible, as in Section 3, that both (a) the marginal insurance purchases that would occur in response to a small decrease in price would be by buyers who are less costly to firms then than the inframarginal purchases and yet (b) the average cost of provision rises when prices rise due to income effects on the optimal effort choices of the inframarginal buyers.

Section 5: Competition a la Einav-Finkelstein-Cullen (2010) with Multiple Products

In the standard EFC setting discussed in Section 2, individuals have a binary choice: buy an exogenously fixed level of coverage, or forego coverage. In this section we analyze models in which individuals can instead choose between two exogenous positive levels of coverage q_H and q_L, with the former providing strictly more coverage than the other. As in the EFC setting, they can also still choose to forego coverage. In other words, we discuss how to EFC can be extended to two exogenous coverage options.

Individuals have preferences over (coverage, premium) pairs (q, p). These preferences, $v(q, p; \theta)$, depend on the individual's privately known type $\theta \in [0, \bar{\theta}]$, again distributed with cdf F. We define the "willingness to pay" functions $\hat{p}_H(\theta)$ and $\hat{p}_L(\theta)$ via $v(q_i, \hat{p}_i(\theta); \theta) = v(0,0; \theta)$, i.e., as the premium that makes an individual indifferent between buying insurance contract i and forgoing insurance. We also define the function $\hat{p}_\Delta(\theta) \equiv \hat{p}_H(\theta) - \hat{p}_L(\theta)$. Whenever there are no income effects in consumer preferences, $\hat{p}_\Delta(\theta)$ can be interpreted as the willingness to pay to upgrade from q_L to q_H; we assume no income effects henceforth.[11]

We assume that $\hat{p}_H(\theta)$, $\hat{p}_L(\theta)$, and $\hat{p}_\Delta(\theta)$ are all strictly decreasing in Δ, so that higher (e.g., less risky) types both value insurance less *and* value incremental coverage less. This assumption on the $\hat{p}_i(\theta)$s ensures that types will be well ordered, in the following sense: if, given prices, an individual of type θ prefers coverage q_H over q_L then every type with $\theta' < \theta$ will also prefer q_H over q_L; similarly, if an individual of type θ prefers coverage q_L (or q_H) over no coverage, then every type with $\theta' < \theta$ will also prefer q_L (or q_H) over no coverage. An implication is that, in any equilibrium (which we will define below), there will be two cutoffs: $\hat{\theta}_H$ and $\hat{\theta}_L$, and individuals with $\theta < \hat{\theta}_H$ will purchase q_H, individuals with $\theta \in (\hat{\theta}_H, \hat{\theta}_L)$ will buy q_L, and individuals with $\theta > \hat{\theta}_L$ will forgo coverage. (These cutoffs could coincide, if, e.g., all types buy either q_H or all types forgo coverage.)

The costs of providing type θ with coverage q_H or q_L are exogenously given and denoted by $c_H(\theta)$ and $c_L(\theta)$. We define the incremental cost as $c_\Delta(\theta)$, and we assume that c_H, c_L, and c_Δ are all strictly decreasing in θ. The assumption that costs decrease with θ would hold, for example, if θ is the probability of not having an accident leading to a loss of l, and if the two contracts differ in the fraction of l covered by the policy. More generally, this assumption implies that the markets feature adverse selection. We additionally assume that $c_H(0) < \hat{p}_i(0)$ for $i = H, L, \Delta$; this ensures

[11] Zero income effects is consistent with quasilinear preferences $v(q, p; \theta) = \hat{v}(q; \theta) - p$, or with expected utility maximization for individuals with constant absolute risk aversion preferences. Without this assumption, the willingness to pay for incremental coverage will depend on the baseline wealth level. In practical terms, in the "sequential" case that we consider below, these income effects will mean that the willingness to pay for incremental coverage can depend on the premium for basic coverage. Abstracting from income effects therefore makes analysis of the sequential case significantly simpler.

that the lowest type is willing to pay their actuarially fair price to buy insurance, and therefore that there is scope for trade. It will be useful to define the average cost curves $c_i(\theta)$ (for each $i \in \{H, L, \Delta\}$) as $AC_i(\theta) = \frac{\int_0^\theta c_i(t)dF(t)}{F(\theta)}$, i.e., as the average cost of providing coverage i to all types below θ.

In the following two subsections, we contrast two distinct institutions for providing the different coverage levels. The first we call "sequential," and the second we call "simultaneous." In the sequential institution, we think of the coverage as being "layered:" individuals choose between being uninsured or buying a primary basic coverage policy. Then, *if* they have purchased the primary policy, individuals may additionally choose a secondary "supplemental" policy which provides additional coverage. The primary and secondary policies are sold and priced separately, with equilibrium prices determined via price competition a la EFC.

In simultaneous contracting, individuals instead make a single choice from among three mutually exclusive policy options: a low coverage contract, a high coverage contract, or a zero-coverage "contract" (i.e., remaining uninsured). In a simultaneous contracting, the high-coverage policy can be thought of as a bundle of the low coverage primary policy plus supplemental secondary coverage – but the primary and secondary policies are bundled together rather being sold separately, and firms compete over the price of the bundled policy.

In both contracting institutions, competition occurs over two distinct prices. The key difference is between *what* is being priced. In sequential contracting, firms compete over prices p_L and p_Δ for the primary and supplemental policies, respectively. In simultaneous contracting, firms instead compete over prices p_L and p_H. In both cases, we define an *equilibrium* to be a pair of prices such that each contract breaks even, given the set of consumers who choose to buy it at that price. We assume for expositional simplicity that the equilibrium is unique.[12] In fact, we make the stronger assumption that $AC_i(\theta)$ and $\hat{p}_i(\theta)$ cross at most once.

A few preliminary remarks and definitions will be useful before we turn, in subsections A and B, respectively, to comparing and contrasting properties of equilibrium in the sequential and simultaneous contracting environments. First, a comment on the relation to the literature. Our analysis is closely related to both Weyl and Veiga's (2017) and Geruso et al. (2021). Like us, Weyl and Veiga (2017) compare equilibrium properties in sequential and simultaneous contracting environments – which they refer to as "incremental pricing" and "total pricing" frameworks. Unlike in our analysis, however, they follow EFC and Handel et al. (2015) in assuming universal coverage – i.e., that all buyers *either* buy L or H coverage, so there is only one choice margin for insurance buyers. Geruso et al. (2021), focus specifically on the implications of having the second margin by allowing individuals to opt out of coverage entirely. They do so specifically in the simultaneous contracting case; our discussion of that case replicates the basics of their analysis. As in all of these papers, we focus exclusively on price competition over fixed contracts, and abstract from endogenous contracting in the Rothschild and Stiglitz (1976) tradition.

Second, equilibrium outcomes are always different in the two contracting environments. That is, the set of individuals who end up with each of the three different coverage levels will generically be different in the two contracting environments. (The exceptions is in trivial cases e.g., where there is universal purchase of high coverage.) We explain the simple intuition for this result at the end of subsection B.

[12] This definition of equilibrium is the natural extension of EFC to the two coverage-level case. As discussed in Geruso et al (2021), there may be multiple prices satisfying this condition, in which case it may be desirable to refine the equilibrium notion.

Third, there are three "critical" types, denoted by θ_L^*, θ_H^* , and θ_Δ^* which play an important role in our analysis of both institutions. These are the types for which $AC_i(\theta_i^*) = \hat{p}_i(\theta_i^*)$, assuming such a point exists (and, if not, we take $\theta_i^* = \bar{\theta}$) Note that θ_L^* is the lowest type who would just be indifferent to purchasing q_L in the EFC market *if* the only coverage options were q_L or null-coverage; the equilibrium price would be $\hat{p}_L(\theta_L^*)$ in this one-option EFC market. For example, if θ measures the probability of no loss, then θ_L^* is the highest θ (lowest risk) type who finds it desirable to purchase insurance at $\hat{p}_L(\theta_L^*)$ in equilibrium, and all—and only—lower types (higher risk types) buy insurance. Similarly, θ_H^* would be the cutoff type in an EFC market for which only q_H (or null coverage) were available. The third critical value, θ_Δ^*, has the following interpretation: it would be the equilibrium price in an EFC market in which all individuals were provided with (or were required to buy) q_L and competition was over supplemental policies (i.e., as the equilibrium price in the EFC (2010) setting).

The nature of equilibrium will hinge on whether $\theta_L^* > \theta_H^*$ or $\theta_H^* \geq \theta_L^*$. We will refer to the former as case 1 and the latter as case 2. Note that the average cost for H is equal to the average cost for L plus the average cost for Δ, and similarly for willingness to pay. It follows that, both in case 1 and in case 2, θ_H^* must lie between θ_L^* and θ_Δ^*. We now turn to our equilibrium analysis.

A. Sequential contracting

With sequential contracting, individuals face the two prices p_L and p_Δ for primary and supplemental coverage, respectively. Faced with these prices, a type θ will choose to buy both policies if both of the following hold

$$\hat{p}_L(\theta) + \hat{p}_\Delta(\theta) > p_L + p_\Delta \qquad (5.A.1)$$

$$\hat{p}_\Delta(\theta) > p_\Delta . \qquad (5.A.2)$$

The former of these ensures that an individual would rather buy both layers than forgo insurance; the latter ensures that the individual prefers to buy the supplement instead of just the primary policy.

A type θ will choose to buy only the primary policy if (5.A.2) is violated and

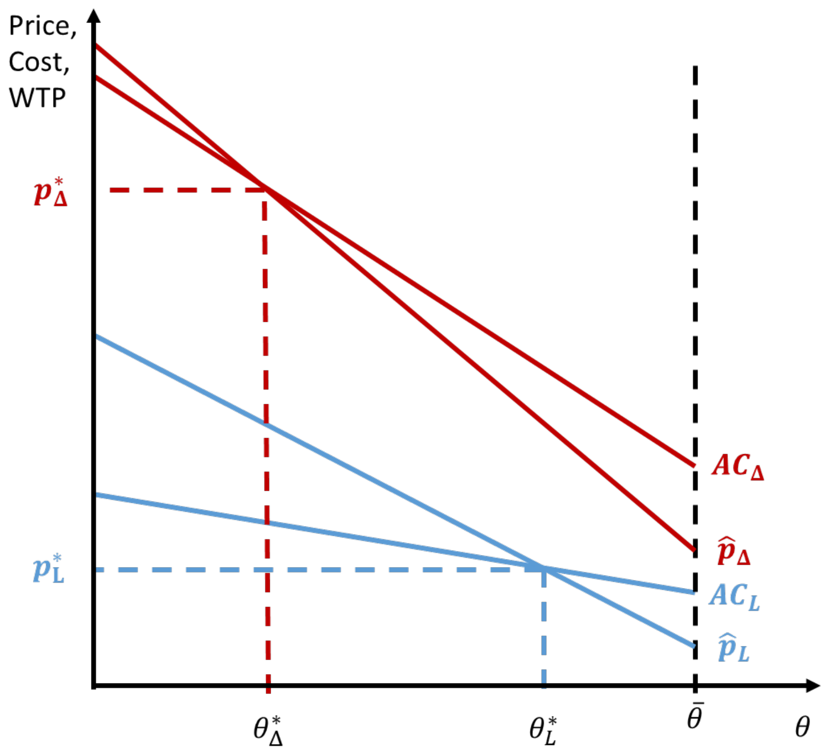
$$\hat{p}_L(\theta) > p_L. \qquad (5.A.3)$$

Observe first that if (5.A.1) and (5.A.2) hold for a type θ, then they also both hold for all types $\theta' < \theta$. So, for any given prices, there will be some cutoff type $\tilde{\theta}_H$, such that all types with $\theta < \tilde{\theta}_H$ (e.g, higher risk types) will choose high coverage. Observe second that, similarly, there is a cutoff type $\tilde{\theta}_L$ below which (5.A.3) is satisfied. Together, these two observations imply that for any pair of prices p_L, p_Δ there are three mutually exclusive possibilities for consumer demand. First, it is possible (e.g., if both prices are very low) that $\tilde{\theta}_H = \bar{\theta}$. Then all types will purchase high coverage. Second, it is possible that $\tilde{\theta}_H < \bar{\theta}$ and that $\hat{p}_L(\theta) < p_L$ for all $\theta > \tilde{\theta}_H$. In this possibility, the highest θ (lowest risk) types forgo insurance, and everyone else purchases *both* the primary and the supplemental policy. (This second possibility will arise, e.g., if the primary policy is expensive, but not too expensive, and the supplemental policy is cheap.) Third, it is possible that $\tilde{\theta}_H < \tilde{\theta}_L \leq \bar{\theta}$. In this third possibility: the lowest θ (highest risk) types buy both policies, and obtain coverage q_H; some intermediate θ (lower risk) types buy only q_L; and some of the highest θ (lowest risk) types may forgo insurance.[13]

[13] There is also a possibility in which prices are so high that no policies get sold at all; we omit this, as it will never occur in equilibrium, given our assumption that the willingness to pay of the lowest-indexed type exceeds the cost of providing that type with coverage.

Figure 5.A.1 illustrates an equilibrium in which the third possibility obtains. The price of the primary policy is p_L^*, where the average cost curve $AC_L(\theta)$ and the willingness to pay curve $\hat{p}_L(\theta)$ cross. The price of the supplemental policy is p_Δ^*, where the average cost curve $AC_\Delta(\theta)$ and the willingness to pay curve $\hat{p}_\Delta(\theta)$ cross. Note that this can only occur when $\theta_L^* > \theta_H^*$, i.e., in which we called "Case 1". In fact, the diagram indicates that whenever Case 1 obtains, equilibrium will take this "third possibility" form: low θ types will buy both policies and obtain coverage q_H, and an interval of higher types will buy only the primary policy and obtain coverage q_L.

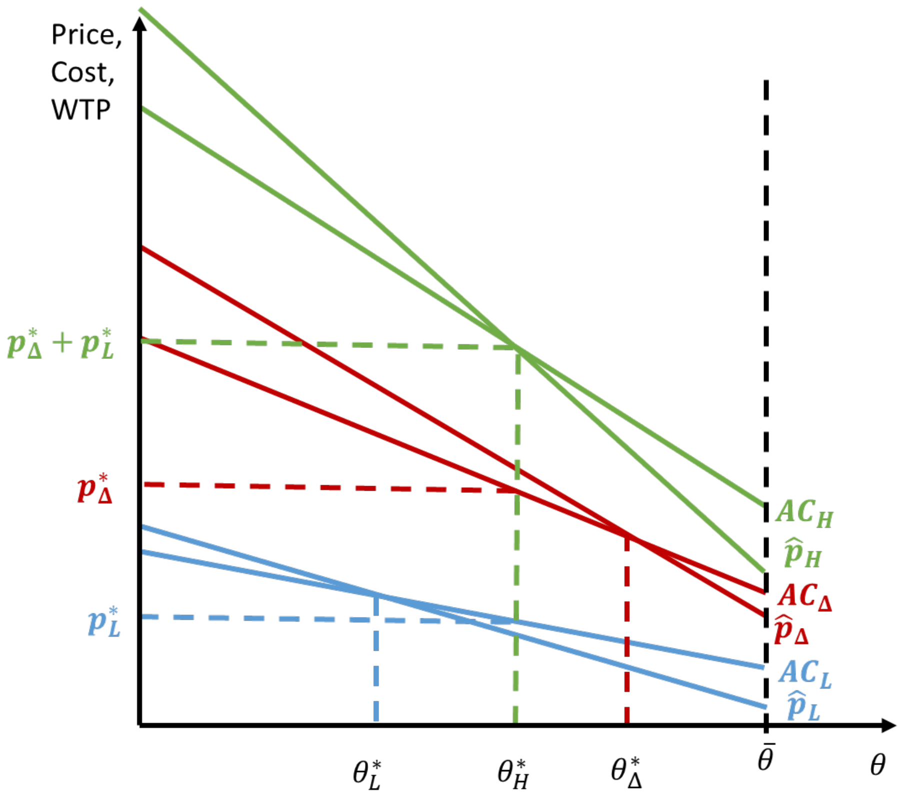
Figure 5.A.1: Equilibrium in Sequential Markets in Case 1



When Case 2 ($\theta_H^* \geq \theta_L^*$) obtains, in contrast, equilibrium can *never* feature types who buy only primary coverage. The argument for this is simple: suppose, by way of contradiction, both that types below some $\hat{\theta}_H$ buy both policies and that a non-trivial interval of types $(\hat{\theta}_H, \hat{\theta}_L)$ buy only the primary policy. Then equilibrium pricing would imply that $AC_\Delta(\hat{\theta}_H) = \hat{p}_\Delta(\hat{\theta}_H)$ and $AC_L(\hat{\theta}_L) = \hat{p}_L(\hat{\theta}_L)$, i.e., $\hat{\theta}_H = \theta_\Delta^*$ and $\hat{\theta}_L = \theta_L^*$ and hence that $\theta_L^* > \theta_H^* > \theta_\Delta^*$, inconsistent with Case 2.

Figure 5.A.2 illustrates equilibrium in Case 2. It is effectively the same as equilibrium in the EFC market when *only* q_H is available: Individuals below the critical value θ_H^* (where the average cost and willingness to pay curves for q_H cross) buy both policies, and individuals above θ_H^* buy no insurance at all. The only difference is that individuals buy two distinct contracts to achieve coverage q_H: the basic coverage level q_L, and the top-up policy $q_H - q_L$. As indicated in Figure 5.A.2, each of the two layers is priced so that it breaks even, given that all types below cutoff θ_H^* purchase both the basic coverage and the supplemental policy.

Figure 5.A.2: Equilibrium in Sequential Markets in Case 2



In case 2, as depicted in Figure 5.A.2, adverse selection implies inefficiency, just as in the basic EFC model with adverse selection (Figure 2.1): the marginal cost lies below the average cost curve here, so providing insurance to types just above θ_H^* would be efficiency-enhancing (deadweight loss reducing).

In case 1, as depicted in Figure 5.A.1, adverse selection implies inefficiency along *two* margins. Because the marginal cost for incremental coverage lies below the average cost of incremental coverage, it would efficiency-enhancing to for individuals with types just above θ_Δ^* (e.g., slightly lower risk types) to switch from q_L to $q_H = q_L + \Delta q$. And because the marginal cost for basic coverage lies below the average cost for basic coverage, it would *also* be efficiency-enhancing for individuals with θ just above θ_L^* (e.g., slightly lower risk types) to receive coverage.

In case 1, these two margins of inefficiency can be targeted (at least marginally) *separately*. Introducing a small subsidy for incremental coverage policies, e.g., will raise the equilibrium cutoff θ_Δ^* but leave θ_L^*, and all types who purchase only basic coverage, unaffected. Similarly, introducing a small subsidy for basic coverage will raise θ_L^*—efficiently inducing more individuals to get basic coverage—without having any effect on the cutoff θ_Δ^*. A subsidy on basic coverage will, of course, also increase the *welfare* of the (low risk) types below θ_Δ^*, since they will be purchasing a now-lower-priced basic policy.[14]

B. Simultaneous Contracting

The simultaneous contracting institution that we analyze here coincides with the model developed in Geruso et. al. (GLMS, 2021).[15] As emphasized by GLMS, there will potentially be adverse selection on two margins, much as in case 1 in the sequential contracting institution. First, there can be adverse selection on the choice between the two coverage levels. Second, there can be adverse selection on the extensive margin, that is, on the choice between becoming insured at all and an outside option of remaining uninsured. In contrast to the sequential contracting institution, however, these margins cannot easily be targeted separately. Indeed, a major point of emphasis in

---

[14] Because there are no income effects, a marginal subsidy on basic policies could be funded by a uniform, non-distorting tax. If the subsidy on basic coverage were funded instead by, e.g., a distorting tax on supplemental policies, then the margin θ_Δ^* would be also be distorted, with negative efficiency consequences.

[15] See also Finkelstein, Hendren and Shepard (2019), and Azevedo and Gottlieb (2017). The latter derive a fully general extension of EFC which allows for an arbitrary number of coverage levels and a universal type space. GLMS (and the parallel analysis here) can be seen as a special case of Azevedo and Gottlieb's analysis in which the coverage levels are restricted to two, and the type space is a one-dimensional subset of the universal type space. These restrictions impose enough structure that GLMS can analyze equilibrium, and equilibrium comparative statics, with relatively simple geometry.

GLMS is that there is a *tradeoff* between selection on the intensive and the extensive margins: attempts to increase the number of individuals receiving basic coverage by, e.g., subsidizing basic coverage policies will lower the price of those policies, and lead some low θ (high risk) types to switch, inefficiently, from high to low coverage.

GLMS focus on the two-margin case where the lowest θ (highest risk) types $\theta < \hat{\theta}_H$ types buy q_H and intermediate types $\theta \in (\hat{\theta}_H, \hat{\theta}_L)$ types buy q_L. The equilibrium conditions in this case are *conceptually* straightforward: we look (a) for a cutoff type $\hat{\theta}_H$ whose willingness to pay to upgrade from q_L to q_H exactly matches the average cost for firms to serve all types in the interval $[0, \hat{\theta}_H)$ and (b) for a cutoff type $\hat{\theta}_L$ whose willingness to pay for q_L exactly matches the average cost for firms to serve all types in the interval $(\hat{\theta}_H, \hat{\theta}_L)$. Practically, however, these equilibrium conditions are harder to identify than in the sequential contracting cases depicted in Figures 5.A.1 and 5.A.2. The equilibrium condition for $\hat{\theta}_H$ is complicated by the fact that the willingness to pay to upgrade to q_H depends on the surplus $\hat{p}_L(\hat{\theta}_H) - p_L$ they would get from buying q_L, and hence on the equilibrium price p_L^*. Specifically, the condition is:

$$\hat{p}_H(\hat{\theta}_H) - [\hat{p}_L(\hat{\theta}_H) - p_L^*] = AC_H(\hat{\theta}_H) = p_H^* \tag{5.B.1}$$

Equation 5.B.1 states that the willingness to pay to upgrade (the willingness to pay for q_H less the surplus from q_L)—the demand for H—has to equal the average cost of serving that demand (the "supply"), which is in turn equal to the equilibrium price.

The equilibrium condition for $\hat{\theta}_L$ is complicated by the fact that average cost

$$\widetilde{AC}_L(\hat{\theta}_H, \hat{\theta}_L) \equiv \frac{\int_{\hat{\theta}_H}^{\hat{\theta}_L} c_L(\theta) d\theta}{F(\hat{\theta}_L) - F(\hat{\theta}_H)} \tag{5.B.2}$$

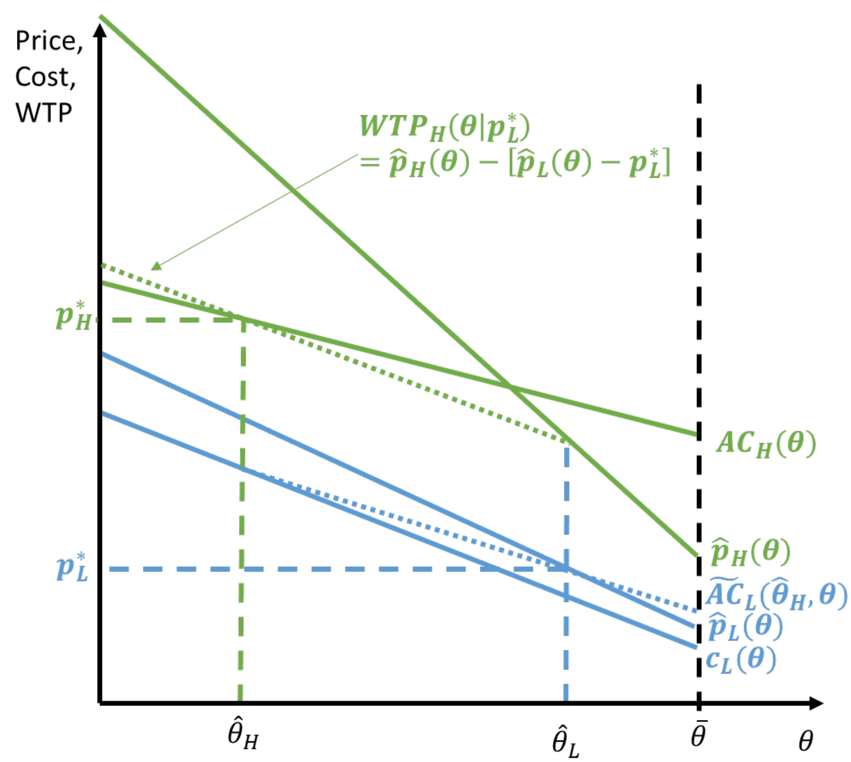
depends on *both* cutoffs. Consequently, so does the equilibrium condition:

$$\hat{p}_L(\hat{\theta}_L) = \widetilde{AC}_L(\hat{\theta}_H, \hat{\theta}_L) = p_L^* \tag{5.B.3}$$

The equilibrium conditions 5.B.1 and 5.B.3 are non-separable: they must be solved as a system.

GLMS describe a graphical algorithm for determining this solution. Figure 5.B.1 recapitulates.

Figure 5.B.1: Equilibrium in Simultaneous Markets



The green dotted line labeled $WTP_H(\theta|p_L^*)$ in Figure 5.B.1 depicts the (endogenous) willingness to pay for q_H given the equilibrium price p_L^*. Per equilibrium condition (5.B.1), the $WTP_H(\theta|p_L^*)$ curve crosses the $AC_H(\theta)$ curve at the equilibrium price p_H^*. Additionally,

$WTP_H(\theta|p_L^*)$ must strike $\hat{p}_H(\theta)$ at $\hat{\theta}_L$. This is because, in equilibrium, $\hat{\theta}_L$ is the type that gets exactly zero surplus from purchasing q_L.

The blue dotted line labeled $\widetilde{AC}_L(\hat{\theta}_H, \theta)$ depicts the (endogenous) average cost of providing the interval $(\hat{\theta}_H, \theta)$ with coverage q_L. Per equilibrium condition (5.B.3), $\widetilde{AC}_L(\hat{\theta}_H, \theta)$ crosses $\hat{p}_L(\theta)$ at the quantity $\hat{\theta}_L$ and the equilibrium price p_L^*. Additionally, it intersects the (marginal) cost curve $c_L(\theta)$ at $\hat{\theta}_H$. (This is because the marginal cost coincides with the average cost for the first unit sold.)

As in case 1 with sequential contracting (viz Figure 5.A.1), the equilibrium depicted in Figure 5.B.1 features two margins of inefficiency. First, the cutoff $\hat{\theta}_L$ is inefficiently low: adverse selection causes too many individuals (e.g., the lowest risks with $\theta > \hat{\theta}_L$) to forgo insurance, even when purchasing it would be efficient. In Figure 5.B.1, $c_L(\theta)$ lies everywhere below $\hat{p}_L(\theta)$, so efficiency in fact requires that *all* types buy insurance. Even if these two curves were to cross however, the equilibrium $\hat{\theta}_L$ would necessary be inefficiently low, since equilibrium features $\hat{p}_L(\hat{\theta}_L) = \widetilde{AC}_L(\hat{\theta}_H, \hat{\theta}_L) > c_L(\hat{\theta}_L)$, and thus the willingness to pay of types just above the cutoff exceeds the marginal cost of providing them with insurance. Thus, there is inefficiency at the extensive margin.

The second source of inefficiency is that $\hat{\theta}_H$ is *also* inefficiently low. This is harder to see in the figure, but follows from some simple algebra. Specifically, re-arranging equilibrium condition 5.B.1 shows that:

$$AC_\Delta(\hat{\theta}_H) + [AC_L(\hat{\theta}_H) - p_L^*] = \hat{p}_\Delta(\hat{\theta}_H) \qquad (5.B.4)$$

Because c_Δ is decreasing, we know that $c_\Delta(\hat{\theta}_H) < AC_\Delta(\hat{\theta}_H)$. The types in the average $AC_L(\hat{\theta}_H)$ are all lower θ (e.g. higher risk) than the types in the average $\widetilde{AC}_L(\hat{\theta}_H, \hat{\theta}_L)$, so we also know that $AC_L(\hat{\theta}_H) - p_L^* > \widetilde{AC}_L(\hat{\theta}_H, \hat{\theta}_L) - p_L^* = 0$. It follows from equation 5.B.4 that $c_\Delta(\hat{\theta}_H) < \hat{p}_\Delta(\hat{\theta}_H)$, i.e., that it would be efficiency-enhancing for individuals just above $\hat{\theta}_H$ to be shifted towards q_H. Thus, there is inefficiency at the intensive margin.

Both the sequential contracting (Figure 5.A.1) and simultaneous contracting (Figure 5.B.1) (can) thus feature a two-cutoff form, with types below $\hat{\theta}_H$ (high risks) buying q_H, types between $\hat{\theta}_H$ and $\hat{\theta}_L$ buying q_L and types above $\hat{\theta}_L$ (low risks) forgoing insurance. Moreover, both cutoffs are inefficiently low under both institutions. The two settings thus look qualitatively similar. It is important to note, however, that they do not – indeed, cannot – coincide.

To see why, consider the equilibrium depicted in Figure 5.A.1 for the sequential institution, in which types $\theta < \theta_\Delta^*$ get coverage q_H at total price $p_L^* + p_\Delta^*$, types $\theta \in (\theta_\Delta^*, \theta_L^*)$ get coverage q_L at price p_L^* and types $\theta > \theta_L^*$ get no coverage. For the simultaneous institution equilibrium depicted in Figure 5.B.1 to yield the same coverage levels for each type, consumers would have to be willing to make the same decisions in the two environments, given their pricing. This would require first that p_L^* in the two figures coincide (so that the same cutoff type θ_L^* is indifferent to coverage q_L and null coverage). But this is inconsistent with break-even pricing for firms: firms in the simultaneous contracting environment of Figure 5.B.1 face the average cost curve $\widetilde{AC}_L$, which averages the cost of all types $\theta \in (\theta_\Delta^*, \theta_L^*)$ and hence (since the marginal cost is decreasing in type) lies strictly below the AC_L curve for firms in the sequential contract environment of Figure 5.A.1. (Graphically: if the AC_L curve from Figure 5.A.1 were drawn in Figure 5.B.1, it would look like the dotted blue line $\widetilde{AC}_L$ but would "start" at c_L at $\theta =$

0 instead of at $\theta = \hat{\theta}_H$. It follows that *if* the cutoff types remained the same in the simultaneous contracting environment as they were in the sequential one, then firms selling coverage q_L would be strictly profitable. Similarly, one can show that *if* the cutoff types remained the same in the simultaneous contracting environment as they were in the sequential one, then firms selling coverage q_H would make losses. Intuitively: with sequential coverage, the baseline contract q_L provides cross-subsidies from the lower cost types to the higher cost types, a cross subsidy that disappears with simultaneous coverage.

For closely related reasons, we expect that the inefficiency at the $\hat{\theta}_H$ margin will be worse under simultaneous contracting, and the inefficiency at the $\hat{\theta}_L$ margin will be worse under sequential contracting.

The logic for why inefficiency will be worse on the "intensive" $\hat{\theta}_H$ margin under simultaneous contracting mirrors that in Weyl and Veiga (2017). Observe, in particular, that adverse selection inefficiencies in the sequential case stem from the fact that $AC_\Delta < c_\Delta$, and hence there is a wedge between p^*_Δ and c_Δ at that margin. Per Equation (5.B.4), the simultaneous case features this wedge plus the *additional* "wedge" $\left[AC_L\left(\hat{\theta}_H\right) - p^*_L\right]$. Intuitively, adverse selection is worse along the intensive margin with simultaneous contracting, because when the marginal individual shifts from q_H to q_L, they raise the average risk in the q_L pool (they are now the highest risk type in that pool) *and* raise the average risk in the q_H pool (as they were the lowest risk in that pool). When the marginal individual shifts in the sequential case, they raise the average risk in the q_Δ (supplemental policy) pool but leave the q_L pool unaffected.

To see why, on the other hand, inefficiency will tend to be worse on the "extensive" $\hat{\theta}_L$ margin in the simultaneous case, observe that, in both institutions, the inefficiency stems from the wedge between the marginal cost c_L and an average cost. In the simultaneous institution, the average is only over $[\hat{\theta}_H, \hat{\theta}_L]$ interval, and in the sequential market it is over the entire $[0, \hat{\theta}_L]$ interval. The latter includes the "extra" high cost types, so the wedge will tend to be larger.

A similar logic underpins the central theoretical observation of GLMS, namely that implementing purchase subsidies or non-purchase penalties in simultaneous markets has offsetting efficiency consequences on the two margins.[16] In Figure 5.B.1, implementing a subsidy for purchasing any insurance policy will shift up the $\hat{p}_L(\theta)$ curve, moving the equilibrium $\hat{\theta}_L$ to the right and lowering p_L^*, improving efficiency on the extensive margin. In so doing, however, it will shift down the WTP_H curve, and thereby *lowering* the cutoff $\hat{\theta}_H$, worsening the inefficiency on the intensive margin. Note that this is in contrast to the effects of a purchase subsidy (or non-purchase penalty) with sequential contracting as in Figure 5.A.1: such a subsidy will lessen inefficiency on the extensive margin without affecting the intensive margin.[17]

6. Conclusion

The approach articulated in Einav, Finkelstein, and Cullen (2010) and Einav and Finkelstein (2011) has reinvigorated research into insurance markets, and interventions therein, by re-orienting it towards a basic supply/demand approach. We provide a simple, user-friendly

[16] See also Azevedo and Gottlieb (2017), who simulate the consequences of a minimum coverage requirement and show that this requirement leads, via a similar channel, to significant and inefficient reductions in the level of coverage chosen in equilibrium by relatively low risk types.

[17] Note that a purchase subsidy in Case 2 (Figure 5.A.2) of the sequential contracting institution *will* affect the purchase of high-coverage contracts – it will lead to additional purchases of both the primary and supplemental policies.

summary of their approach, and we demonstrate how it can be extended to new settings—including markets with selection on ex-ante moral hazard and non-exclusive contracting environments—and show that it can be readily applied to policy-relevant problems.

The EFC approach has primarily been used to analyze markets with exclusive contracting and to analyze problems on the extensive margin. We show that the extension from exclusive to non-exclusive contracting is straightforward, and that the main results of the analysis go through. In other words, we show that—just as standard supply-demand analysis for "ordinary" markets applies equally well in settings (like housing) where the extensive margin is the primary driver of demand and in settings (like fast food) where the intensive margin is of first-order importance—supply and demand analysis in selection markets also applies equally well in markets where the intensive margin is important. This specifically opens the door to empirical analysis of markets for life insurance and annuities.

There are, however, some important differences between the textbook model and insurance markets. The main difference is the connection between demand and cost. In "ordinary" markets, demand is determined by preferences, cost is determined separately by the production technology and the firm's supply curve is its marginal cost curve—and, correspondingly, the market supply curve is the social marginal cost curve. In selection markets, the shape of the cost curves is determined by consumer selection on the demand side of the market. The individual's type determines demand and also directly (or, in the case of selection on moral hazard, indirectly) determines cost. With adverse selection, for example, the riskiest individuals have the highest demand (either willingness-to-pay or quantity demanded) and are also the highest cost customers to serve. This leads to another difference between product markets and insurance markets, namely, that, in competitive selection markets, it is the market's *average* cost curve, not the market's marginal cost curve which serves as the supply curve.

We also show that the EFC supply and demand approach can also be readily adapted to situations in which there is selection on ex-ante moral hazard. In such settings, there is still likely to be "adverse selection" on the extensive margin – at lower prices, the marginal individual entering the market is less costly to serve than inframarginal individuals (since they value insurance less because it is less personally costly for them to mitigate risk, but, by the same token, their risk mitigation efforts will reduce the costs to their insurer). But we show that—even when there is clear "adverse selection" on the extensive margin—the market-wide marginal cost of serving an additional buyer can still be higher than the market's average cost of serving the buyers, and the average cost curve can therefore slope upwards. In other words, even when there is clear adverse selection on the intensive margin, the market can look, in aggregate, exhibit behavior which would be interpreted as advantageous selection in the EFC framework.

Finally, we discuss extensions of the EFC approach to settings with multiple, vertically differentiated contracts. In such settings, there may be *tradeoffs* between the extensive and intensive margins. But whether or not there are tradeoffs depends critically on how the markets are structured, in particular, on whether the contract choice is sequential (i.e., layered) or simultaneous. In a market with simultaneous choice a subsidy to the low coverage policy expands coverage on the extensive margin, but reduces coverage on the intensive margin as the change in relative prices induces individuals to switch from the high to the low coverage contract. In a market with sequential choice there is no such intensive margin effect. This suggests that the sequential market architecture insulates the market for the secondary policy from interventions in the market for the primary policy.

References

Abel, A. B., 1986, Capital Accumulation and Uncertain Lifetimes with Adverse Selection, *Econometrica*, 54: 1079-1097.

Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanism, *The Quarterly Journal of* Economics, 89, 488-500.

Attar, A., T. Mariotti, and F. Salanie (2014). Nonexclusive competition under adverse selection, *Theoretical Economics* 9, 1-40.

Azevedo, E., and Gottlieb, D. (2017). Perfect competition in markets with adverse selection. *Econometrica*, *85*(1), 67-105.

Bundorf, M.K., J. Levin, and N. Mahoney, (2012). Pricing and welfare in health plan choice, American Economic Review, 102(7), 3214-3248.

Chiappori, P.-A., 2001, Econometric Models of Insurance Under Asymmetric Information, in: G. Dionne, ed., *Handbook of Insurance* (Boston: Kluwer Academic Publishers), pp. 365-393.

Chiappori, P.-A. and B. Salanie., (2000) Testing for asymmetric information in insurance markets, *Journal of Political Economy*, 56-78.

Cohen, A., (2005) Asymmetric information and learning in the automobile insurance market, *Review of Economics and Statistics*, 197-207

De Meza, D. and D. Webb (2001) Advantageous selection in insurance markets, *Rand Journal of Economics*, 32, 249-262.

Dubey, Pradeep and Geanakoplos, John (2019), Non-Exclusive Insurance with Free Entry: A Pedagogical Note, *Cowles Foundation Discussion Papers*, 91.

Ehrlich, I., and G. Becker (1972) Market insurance, self-insurance and self-protection, *Journal of Political Economy*, 80: 623-684.

Einav, L., and A. Finkelstein (2011). Selection in insurance markets: Theory and empirics in pictures, *Journal of Economic Perspectives*, 25(1), 115-138.

Einav, L., A. Finkelstein and M. Cullen (2010). Estimating welfare in insurance markets using variation in prices, *Quarterly Journal of Economics*, 125(3), 877-921.

Einav, L., A. Finkelstein, S. Ryan, P. Schrimpf, and M. Cullen (2013). Selection on moral hazard in health insurance, *American Economic Review*, 103(1), 178-219.

Fang, H., M. Keane, and D. Silverman, 2008 Sources of advantageous selection: evidence from the Medigap insurance market. *Journal of Political Economy,* 116:303–350

Finkelstein, A., 2004, The Interaction of Partial Public Insurance Programs and Residual Private Insurance Markets: Evidence from the U.S. Medicare Program, *Journal of Health Economics*, 23: 1-24.

Finkelstein, A., and J. Poterba, 2002, Selection Effects in the Market for Individual Annuities: New Evidence From the United Kingdom, *Economic Journal*, 112: 28-50.

Finkelstein, A., N. Hendren and M. Shepard (2019). Subsidizing health insurance for low income adults: Evidence from Massachusetts, *American Economic Review*, 109(4): 1530-1569.

Geruso, M., T. Layton, G. McCormack, and M. Shepard (2021). The two margin problem in insurance markets, *Review of Economics and Statistics*, forthcoming.

Handel, B., I. Hendel and M. Whinston (2015). Equilibria in health exchanges: Adverse selection versus reclassification risk, *Econometrica*, 83(4) 1261-1313.

Hemenway, D. (1990). Propitious selection, *Quarterly Journal of Economics*, 105(4), 1063-1069.

Hemenway, D. (1992) Propitious selection in insurance, *Journal of Risk and Uncertainty*, 5(), 247-251.

Hendren, N. (2013). Private information and insurance rejections, *Econometrica*, 81(5), 1713-1762

Hendren, N. 2014, Unravelling vs unravelling: A memo on competitive equilibriums and trade in insurance markets, *Geneva Risk and Insurance Review*, 39: 176-183.

Hendren, N. (2019). Measuring ex ante welfare in insurance markets, NBER working paper 24470.

Hoy, M., and M. Polborn, 2000, The Value of Genetic Information in the Life Insurance Market, *Journal of Public Economics*, 78(3): 235-252.

Johnson, W. (1977). Choice of compulsory insurance schemes under adverse selection. *Public Choice*, 31(Fall), 23-35.

Mahoney, N., and E.G.Weyl (2017). Imperfect competition in selection markets, *Review of Economics and Statistics*, 99(4), 637-651.

Panhans, M. (2019). Adverse selection in ACA exchange markets: Evidence from Colorado, *American Economic Journal: Applied Economics*, 11(1), 1-36.

Polborn, M., M. Hoy, and A. Sadanand, 2006, Advantageous Effects of Regulatory Adverse Selection in the Life Insurance Market, *Economic Journal*, 116: 327-354.

Posey, L. and P. Thistle (2019) Large losses and equilibrium in insurance markets, *Geneva Risk and Insurance Review*, 44(2): 222-244.

Riley, J. (1979). Informational equilibrium, *Econometrica*, 47(2) 331-359.

Rothschild, M., and J. Stigliz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *Quarterly Journal of Economics*, 90(1), 630-649.

Rothschild, C. (2015). Nonexclusivity, linear pricing, and annuity market screening, *Journal of Risk and Insurance*, 82(1), 1-32.

Villeneuve, B., (2003). Mandatory Pensions and the Intensity of Adverse Selection in Life Insurance Markets, *Journal of Risk and Insurance*, 70(3): 527-548.

Weyl, E.G. and A. Viega, (2017). Pricing institutions and the welfare cost of adverse selection. *American Economic Journal: Microeconomics* 9(2), 139-148