

Could a Website Really Have Doomed the Health Exchanges?

Multiple Equilibria, Initial Conditions and the Construction of the Fine^{*}

Florian Scheuer
Stanford and NBER

Kent Smetters
Wharton and NBER

January 2014

Abstract

Public attention has focused on how the launch of the national health exchanges could impact the types of risks who initially enroll and thereby affect future premiums and enrollment. We introduce simple dynamics into a standard model of insurance under adverse selection to show that such “initial conditions” can indeed matter. When firms are price-takers, the market can converge to a Pareto-inferior “bad” equilibrium if there are at least three equilibria, which we suggest has empirical support. Strategic pricing eliminates Pareto dominated equilibria but requires common knowledge of preference and risk distributions. Changing the fine on non-participants from a fixed amount to a fraction of equilibrium prices increases the range of initial conditions consistent with reaching the “good” equilibrium while reducing the “badness” of the bad equilibrium — all without increasing the fine value in the good equilibrium. Allowing insurers to quickly change prices can encourage them to experiment with strategic pricing if market fundamentals are not perfectly known, increasing the chance of reaching the good equilibrium independently from initial conditions.

Keywords: Health Exchanges, Adverse Selection, Insurance Markets, Mandate
JEL-Codes: H3, I1, D4, D8

^{*}Email addresses: scheuer@stanford.edu, smetters@wharton.upenn.edu. We are grateful to Eduardo Azevedo, Liran Einav, Daniel Gottlieb, Nathan Hendren and Casey Rothschild for helpful comments.

1 Introduction

The Patient Protection and Affordable Care Act (“ACA” for short) limits the degree to which insurers can price discriminate based on age and preexisting conditions. Enrolling healthier people, especially younger individuals, is, therefore, widely regarded as important for avoiding high premiums. The government aims at enrolling 7 million individuals in the ACA during the first year, of which 2.7 million are between the ages 18 and 35.

The ACA rollout, however, contains initial conditions that may have discouraged healthy individuals from enrolling. First, there were issues with the main website at launch, limiting its ability to enroll participants, to transmit information accurately to insurers, and to securely store information.¹ Second, to prevent adverse selection, the ACA levies a fine (the “share responsibility penalty”) on individuals who do not enroll. But the size of that fine is quite small in 2014, equal to the greater of \$95 or 1% of income, even though increasing in future years. Third, because some households on the individual health insurance market lost their coverage, the government announced on November 14, 2013 that it is allowing individual state insurance commissioners to extend canceled policies by one year.

A potential counterbalancing effect is that fact that, if the initial enrollment deadline is missed, subsequent enrollment is delayed until the next “open enrollment season” even when a person gets sick. However, this effect in turn is weakened during the initial year of the ACA implementation since open enrollment occurs twice in 2014, roughly six months apart, in order to make its timing consistent with Medicare’s open enrollment season in future years.² Moreover, to fill even this short gap, a market for short-term policies (lasting 364 days or less) may emerge and has the potential to create cost savings for young adults, even net of subsidies and fines (Feldstein, 2013; Anderson and Cowan, 2013).³

Of these three initial conditions, policymakers and media outlets originally focused on the issues related to the website. As *The Economist* (November 23, 2013) put it:

¹Other usability issues that could have affected younger shoppers included the inability to compare prices before entering user information. Many young shoppers with new employers also have to separately submit payroll stubs, re-confirm their health exchange status at a later time, and then contact the insurer to make payment. These process breaks may reduce conversions as well, especially for consumers who are not strongly attached to the outcome.

²For 2014 only, there will effectively be two open enrollment periods, the first one from October 1 - March 31, 2014 and the second one from October 15 - December 7 to coincide with Medicare open enrollment. In future years, enrollments will occur once per year, coinciding with the Medicare dates.

³The ACA also provides exemptions for households enrolled in an existing “health care sharing ministry,” which allows its members with a common ethical or religious belief to pool health care costs; these risk pools are typically cheaper than many insurance plans due to lower coverage provided.

Insurers have set their premiums on the assumption that lots of young, healthy people would be compelled to buy their policies. But if it takes dozens of attempts to sign up, the people who do so will be disproportionately the sick and desperate. Insurers could be stuck with a far more expensive pool of customers than they were expecting, and could have no choice but to raise prices next year. That would make Obamacare even less attractive to the young “invincibles” it needs to stay afloat. (p. 15)

Even if many younger people are willing to give an improved website another chance, the small magnitude of the fine in 2014 may encourage many of them to simply pay it rather than larger premiums. A survey conducted by the Harvard University Institute of Politics (2013) finds that less than one third of uninsured Millennials (ages 18 - 29) plan to enroll in an ACA policy. The Pew Research Center (2013) claims that one third of all uninsured households, including older households, plan to not enroll. According to Project 2017 (2013), the premiums for single individuals younger than age 40 are two to nine times larger than the statutory fine in 2014, depending on income. In their estimates, the premiums were calculated using the cheapest “bronze” level plan and include any applicable premium subsidies provided by law.

Their calculations, if anything, are likely to understate the premium-to-fine ratio for two reasons. First, in response to private insurers canceling policies that were not ACA compliant, the government announced in December 2013 that it will grant a “hardship exemption” from the fine in 2014 for anyone who saw their prior policies canceled and who claim that the policies on the exchanges available to them are now “unaffordable.” Second, as the ACA law is currently constructed, the fine can only be levied on individuals who make enough money to file taxes and are owed a *refund* (IRS 2013).⁴ In fact, about 52.4% of single filers do not make sufficient income to file taxes (Tax Policy Center 2011), a percentage that increases for younger filers. We suspect that this feature of the ACA is another initial condition that may change in the future.⁵

The potential importance of allowing individual state insurance commissioners to ex-

⁴Unlike other tax debts, the ACA explicitly prohibits the IRS from enforcing the fine by garnishing wages or confiscating property. Instead, the fine must be levied only against tax filers who are owed a *refund* (IRS 2013). While over 75% of filers were owed a refund in tax year 2012 (IRS 2013B), a sophisticated filer could postpone payments, reducing their value in present value. While it is too early to know if the administrative tax courts would allow the IRS to assess interest and penalties for postponed payments, such assessments have historically been used in cases of illegal evasion rather than legal deferrals.

⁵According to final rule issued by the Health and Human Services on February 22, 2013, student health plans are also exempt from the key features of the ACA. Whether this feature is temporary or permanent is unclear. However, given the rule making process, it appears that HHS has discretion within the ACA to make this determination without additional Congressional approval. The HHS does not, however, have the ability to change the tax filing provisions without Congressional approval.

tend canceled policies by one year is unclear. But it has certainly raised concerns among insurers that ACA compliant plans could fail to attract younger participants. For example, Bloomberg (December 20, 2013) reports that “[i]nsurers said the exemptions may keep younger, healthier people from buying new coverage through Obamacare, a demographic that is needed to bring balance to the new government-run insurance marketplaces.”⁶ However, 12 states have already rejected the extension.⁷ Most other states are allowing for the one-year extension, but, in a few cases, private insurers are refusing to reissue the pre-ACA plans (Associated Press, 2013).

Interestingly, much of the public discussion about the role of initial conditions in the ACA rollout has been based on the standard terminology from the textbook treatment of adverse selection going back to Akerlof (1970). However, the textbook theory of insurance unraveling is not specified in terms of initial conditions but as the equilibrium of a static system of insurance cost and demand equations across risk types. Moreover, most of the insurance literature has implicitly focused on either linear demand and cost curves (see e.g. Cutler and Reber, 1998, and Einav and Finkelstein, 2011) or strategic insurers (Pauly, 1974, and Einav et al., 2010), both of which give rise to a unique equilibrium that is reached no matter which initial conditions we start from. That equilibrium emits a degree of risk pooling ranging from full pooling to no pooling (“unraveling”), or something in-between (where only some of the lower risk types drop out of the market), but there is not much role for initial conditions in affecting the eventual outcome. In the context of the ACA, the standard textbook model suggests that either the policy is destined to be “successful” (i.e., it pools risk across many risk types) or it is destined to “fail” (it does not successfully pool risk across many types). For example, Handel et al. (2013) analyze a model of the ACA health exchanges with a unique equilibrium and conclude that it may eventually involve limited degrees of risk pooling.

So is the media frenzy and the concern by policymakers about the ACA’s initial conditions much to do about nothing? Or, is it indeed possible that the ACA’s initial conditions could actually affect its eventual performance?

In this paper, we add simple dynamics to the textbook model of competitive insurance markets under adverse selection building on Akerlof (1970), Wilson (1977, 1980) and the recent work of Einav et al. (2010) and Einav and Finkelstein (2011), without assuming

⁶See also *The Economist* (November 23, 2013): “A bigger risk is that the “fix” harms the rest of Obamacare. The insurance lobby points out that Mr Obama’s plan will dissuade healthy people from buying more generous, costly coverage on the exchanges. This will leave insurers with a more sickly pool than they expected. That could drive up prices for 2015.” (p. 31).

⁷These are California, Indiana, Massachusetts, Minnesota, Montana, New York, Rhode Island, Nevada, Vermont, Virginia and Washington and West Virginia.

a shape for cost and demand curves. We show that initial conditions can indeed matter for the eventual outcome if (i) insurers are competitive price-takers, and (ii) there exist at least three competitive equilibria. In the case of exactly three equilibria, one equilibrium is unstable and other two are stable. The “good” stable equilibrium provides more coverage at a lower price relative (in fact, is Pareto superior) to the “bad” stable equilibrium. Using data on average costs from the Medical Expenditure Panel Survey, we provide a simple calibration of insurance demand that suggests that the presence of three competitive equilibria is consistent with moderate levels of risk aversion. The reason is that medical expenditures are fairly concentrated within the pre-Medicare population, thereby producing multiple inflection points in the willingness to pay curve across risk types. Multiple crossings with the average cost curve of insurance, therefore, can emerge.

The construction of the fine itself plays a potentially important role in the presence of multiple equilibria. The ACA uses an “absolute” fine, which, as noted earlier, is equal to the greater of 1% of income or \$95 in 2014, growing by pre-determined amounts over the subsequent two years.⁸ In contrast, the Massachusetts health care reform law, enacted in 2006, levied a “relative” fine equal to 50% of the smallest yearly premium for qualified plans available in the market.⁹ The “relative” fine, therefore, effectively grows with the amount of adverse selection. We show that the relative fine increases the likelihood of reaching the good equilibrium, even when the absolute and relative fines are normalized to be equal in value at the point of the good equilibrium. Moreover, even if the bad equilibrium is reached under the relative fine, its “badness” is strictly less than under the absolute fine. Put differently, a re-construction of the fine toward a relative basis is more likely to expand coverage (achieve the good equilibrium) while reducing the badness of the worst outcome, all without costing non-insured consumers anything more in the desired, good equilibrium with a high number of individuals covered at a low premium.

However, if firms set premiums strategically in a Nash equilibrium and have common knowledge (about the distribution of risk types and individual attributes like preferences, loss amounts, and wealth), then the popular media concerns about the system unraveling over time because of a bad first year are without merit: only a single equilibrium exists in this case (see e.g. Einav et al., 2010) and the market converges to it independently from initial conditions. In particular, if premiums were set at a high value, consistent with a Pareto dominated equilibrium under price-taking, then a profitable deviation would exist for any insurer that lowers its premium.

⁸In 2015, the fine grows to greater of 2% of income or \$325; by 2015, it is the greater of 2.5% of income or \$695. Thereafter, the minimum dollar penalty grows with the general inflation level.

⁹This fine has since been updated to be consistent with the ACA requirements.

The required assumption of perfect information about the full set of fundamentals of the market, though, may be quite strong in this context. Cutler and Reber (1998) and Monheit et al. (2004), for instance, provide evidence of repeated marginal price changes that suggest that insurers do not a priori know the entire shape of the demand and cost curves in the market and locally adjust premiums in response to profits or losses they experience. Experimenting with prices to find out about market fundamentals could be quite expensive in this setting, especially if insurers are unable to quickly adjust prices in response to pricing mistakes. In practice, prices are legally rigid in two ways. First, premiums under the ACA can only be adjusted infrequently, typically to coincide with the annual “open enrollment” period. Second, premium price increases must be justified under existing laws in many states and, at the federal level, under new rules implemented by the ACA itself.

The rest of this paper is organized as follows. Section 2 presents a simple insurance model within the price-taking setting; extensions of this model are presented in Appendix B. Section 3 generalizes this price-taking model to a dynamic setting and demonstrates how initial conditions can matter when multiple equilibria exist. This section also presents some quantitative evidence about the potential for multiple equilibria as well as examples of pricing dynamics from some previous efforts by individual states to reform their health insurance systems. Section 4 discusses the important role of the construction of the fine that is levied on consumers who do not purchase health insurance. Section 5 considers the setting where insurers set premiums strategically. Section 6 concludes.

2 Model with Price-Taking Insurers

For the sake of expositional simplicity, we start by considering a simple insurance model that incorporates many of the key features highlighted in Akerlof (1970), Wilson (1977, 1980), Einav et al. (2010) and Einav and Finkelstein (2011). Our most parsimonious model assumes a continuum of risk types, that risk is the only source of consumer heterogeneity, and losses are binary. In Appendix B, we demonstrate that our key results extend to a model setting with discrete risk types, richer forms of heterogeneity and multiple loss sizes.

2.1 Consumers

A unit measure of consumers have wealth $w > 0$ and face a potential loss of size $0 < l < w$ in the presence of limited liability. Consumers only differ in the probability $\pi \in [0, 1]$

of the loss occurring, which is distributed throughout the population by the continuous cumulative distribution function $H(\pi)$ with support $[0, 1]$.¹⁰ Let the random variable Π be H -distributed, and denote a realization of Π by π . Agents are risk-averse with a concave Bernoulli utility function $u(c)$ over consumption, so the expected utility of type π is given by

$$\pi u(w - l) + (1 - \pi)u(w)$$

when there is no insurance.

We assume that individuals can choose from exactly two available insurance contracts that differ exogenously in how much of the loss l they cover. Following Einav et al. (2010), and without loss of generality, we normalize the low coverage contract to be no insurance at a zero premium, and the high coverage contract to be full insurance at some endogenous premium p . Abstracting from moral hazard, we take l and each individual's risk π as exogenous and, therefore, independent of the insurance choice. The demand for insurance, therefore, is only a function of the price p .

The assumption of fixed coverage levels places our analysis in the spirit of Akerlof (1970) rather than Rothschild and Stiglitz (1976), who endogenize coverage levels as well.¹¹ As discussed in Einav et al. (2010) and Einav and Finkelstein (2011), this assumption is a reasonable characterization of many insurance markets. It becomes an even more appropriate assumption for the ACA health exchanges which place regulatory bounds on minimum coverage, despite allowing for a range of plans that differ in copayments made by consumers (see Handel et al., 2013, for a model of insurance markets with two fixed (non-zero) coverage levels, covering 90% and 60% of an individual's cost, respectively).

2.2 Insurers

There are many identical, risk-neutral insurers that maximize their respective expected profits. In this section, we assume that insurers act as price-takers, as in Akerlof (1970). In Section 5, we demonstrate how the results change when insurers set premiums strategically in a model of Bertrand competition.

We assume throughout that an individual's risk type π is private information, so insurers cannot offer different premiums to different individuals. Even if insurers could observe risk type, the ACA does not permit pricing based on pre-existing conditions. One

¹⁰The full support assumption can be viewed as a limiting case where, as in Hendren (2012), even the most extreme risk types, for whom the loss never or always occurs, exist with arbitrarily small but positive density. Our substantive results do not depend on this assumption. Appendix B relaxes this and other assumptions, as noted earlier.

¹¹See e.g. Netzer and Scheuer (2013) for a recent treatment.

can, therefore, think of our analysis as applying to a set of individuals who are otherwise identical in terms of characteristics that insurers are allowed to price, such as smoking status.

2.3 Competitive Equilibria

The following definition of a competitive equilibrium is consistent with the informational assumptions outlined above:

Definition 1. *With unobservable risk types π , a competitive equilibrium is a premium p^* and a critical type π^* such that*

$$u(w - p^*) \geq \pi u(w - l) + (1 - \pi)u(w) \quad \forall \pi \geq \pi^*, \quad (1)$$

$$u(w - p^*) < \pi u(w - l) + (1 - \pi)u(w) \quad \forall \pi < \pi^* \quad (2)$$

and

$$(1 - H(\pi^*))p^* = \int_{\pi^*}^1 \pi dH(\pi)l. \quad (3)$$

The first two conditions characterize consumers' demand for insurance, given the equilibrium premium p^* . At that premium, individuals of risk type $\pi \geq \pi^*$ are just indifferent or strictly prefer to buy insurance, whereas all other types $\pi < \pi^*$ prefer to stay uninsured. The third condition then requires insurers to make zero profits at the policy premium p^* on the pool of risk types who demand insurance when the premium is p^* , which includes all types $\pi \geq \pi^*$. In particular, the left-hand side of (3) equals the total premium revenue collected from these agents while the right-hand side is equal to their expected losses. This zero profit condition can be simply rewritten as $p^* = \mathbb{E}[\Pi | \Pi \geq \pi^*]l$, i.e. the equilibrium premium must equal the expected loss of the pool of insurance buyers induced to buy the policy.

To characterize the set of competitive equilibria, a graphical representation following Einav and Finkelstein (2011) is useful. For any critical buyer $\pi \in [0, 1]$, the *average cost* of insuring everyone with risk equal to or greater than π is

$$\Gamma(\pi) \equiv \mathbb{E}[\Pi | \Pi \geq \pi]l. \quad (4)$$

Our assumptions ensure that $\Gamma(\pi)$ is continuous, increasing in π , and satisfies $\Gamma(0) = \mathbb{E}[\Pi]l$ and $\Gamma(1) = l$. In words, when $\pi = 0$ is the critical type, the average cost of the entire population is just the unconditional expected loss. On the other hand, with critical type $\pi = 1$, their losses are certain, and so their expected loss is simply l .

On the demand side, we can define the *willingness to pay* $\Omega(\pi)$ for insurance of each type π implicitly by solving

$$u(w - \Omega) \equiv \pi u(w - l) + (1 - \pi)u(w). \quad (5)$$

Since the right-hand side is decreasing in π , there is a unique solution $\Omega(\pi)$ for each π , which is also continuous, increasing in π , and satisfies $\Omega(0) = 0$, $\Omega(1) = l$.¹² In words, the lowest risk type $\pi = 0$ never experiences a loss and, therefore, has no willingness to pay for insurance. In contrast, the highest risk type $\pi = 1$ experiences the loss l for sure and is, therefore, willing to pay a premium up to l . We can also interpret $\Omega(\pi)$ as an *inverse demand curve*: with a premium $p = \Omega(\pi)$, insurance will be demanded by all types higher than π (so that the inverse function $\Omega^{-1}(p)$ identifies the marginal buyer when the premium is p).

A competitive equilibrium, therefore, is any π^* such that $\Gamma(\pi^*) = \Omega(\pi^*)$, so that the average cost and willingness to pay curves intersect. For the remainder of the paper, we confine attention to the generic case where all intersections are proper intersections rather than tangency points of the two curves. A simple illustration is provided in Figure 1. Obviously, $\Gamma(1) = \Omega(1)$, so there always exists a competitive equilibrium where nobody buys insurance except for the very highest risk types with $\pi = 1$, who are just indifferent between buying or not buying when faced with the fair premium $p^* = l$ for this pool. In the situation depicted in Figure 1, this outcome is, in fact, the only equilibrium, corresponding to the case of complete unraveling emphasized in Akerlof (1970). Specifically, for any $\pi < 1$, the average cost curve is above the demand curve, so insurers would make losses at any premium $p < l$.

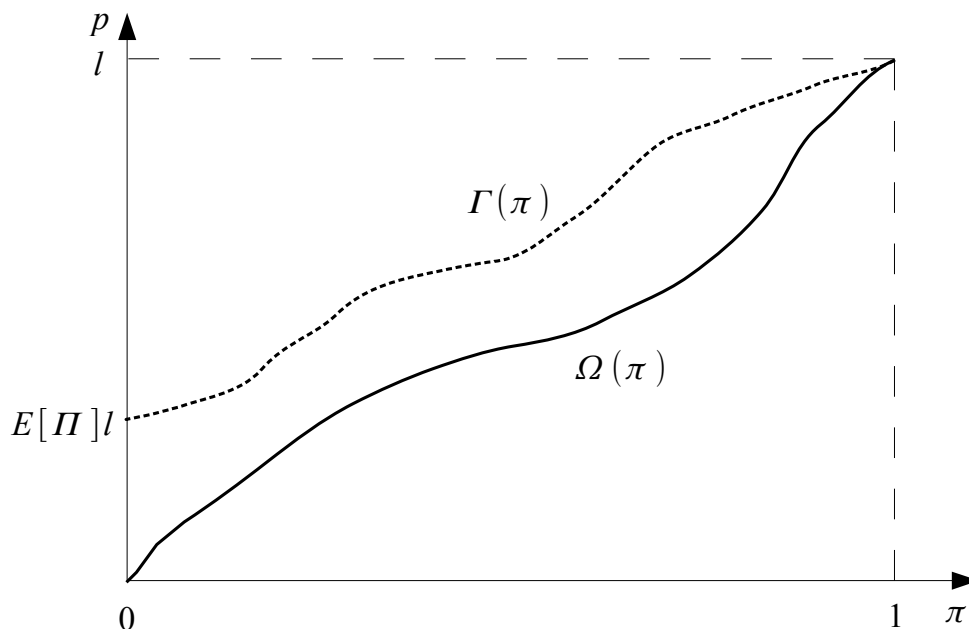
3 Equilibrium Multiplicity and Initial Conditions

Since most of the insurance literature has implicitly focused on either linear demand or cost curves (see, e.g., Cutler and Reber, 1998, and Einav and Finkelstein, 2011) or strategic insurers (see Section 5 and Einav et al., 2010), the possibility of multiple competitive equilibria has not received much attention in this context. As a result, there is no real role for the type of “initial conditions” discussed in Section 1.

Equilibrium multiplicity, however, can arise naturally in our framework because the average cost and demand curves are upward sloping, but their shapes are otherwise unre-

¹²As is standard, we are assuming that the loss size l does not exceed available wealth w . As discussed more below in our calibration exercise, in the presence of limited liability, it is possible for actual losses to exceed wealth for some types, and so $\Omega(\pi)$ is only weakly increasing in π and $\Omega(1) < l$.

Figure 1: Unique Competitive Equilibrium with Complete Unraveling



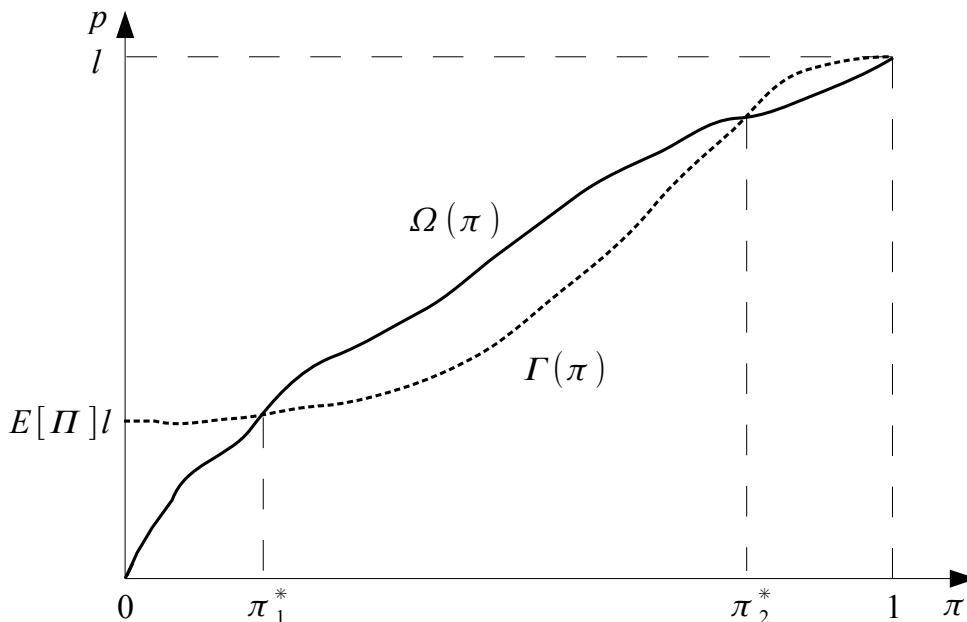
stricted. A simple example is depicted in Figure 3. There are three competitive equilibria in total, namely the one with unraveling located at $\pi^* = 1$ as well as two additional equilibria with critical buyers π_1^* and π_2^* .

As is well known in other contexts (see e.g. Mas-Colell et al., 1995, for a labor market model with unobservable productivities), whenever there are multiple equilibria, they are Pareto ranked: The equilibrium at π_1^* is Pareto better than the equilibrium at π_2^* , which is Pareto better than the equilibrium with complete unraveling at $\pi^* = 1$. For example, compare π_1^* against π_2^* . In the equilibrium at π_1^* , all types $\pi \geq \pi_2^*$ are better off than in the equilibrium at π_2^* because they pay a lower premium for their insurance (i.e., $\Gamma(\pi_1^*) < \Gamma(\pi_2^*)$). Moreover, the types $\pi \in [\pi_1^*, \pi_2^*)$ prefer to buy insurance at π_1^* rather than staying with their endowment, which would be their choice at π_2^* . So, they are also better off. The types $\pi < \pi_1^*$ are indifferent because they don't buy insurance in either case. Moreover, insurers earn zero profits in all equilibria.

Indeed, the “good” equilibrium π_1^* — which offers the lowest premium and entices the most consumers to buy insurance — Pareto dominates all the others.¹³ The other equilibria arise because of a *coordination failure*: When only a few individuals purchase insurance, they will be the highest risk types, and so the premium that breaks even for

¹³In fact, the good equilibrium is constrained Pareto efficient under the restriction to full insurance contracts.

Figure 2: Multiple Competitive Equilibria



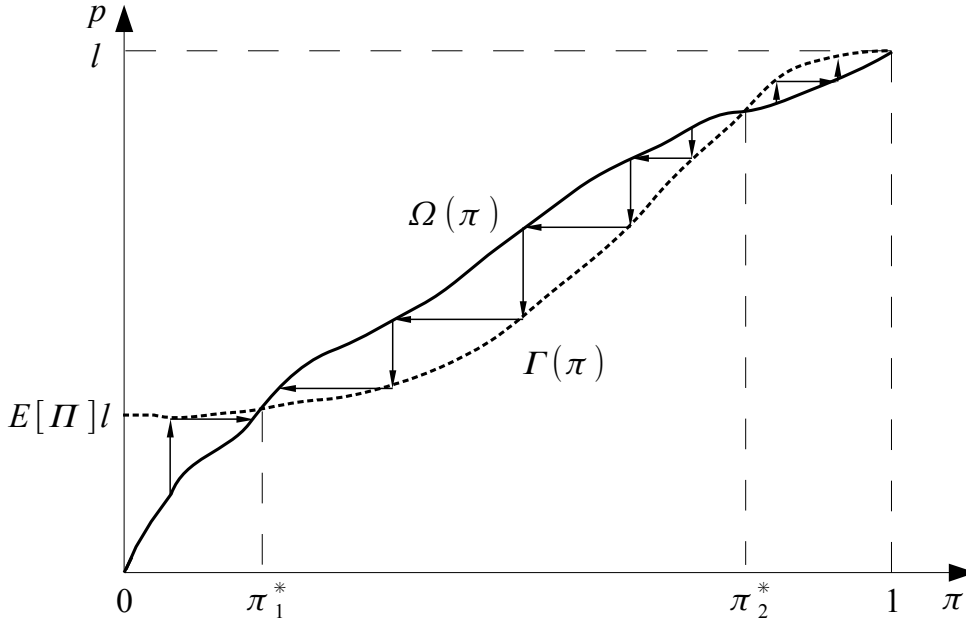
this pool will also be high. At the same time, only the riskiest types find it worthwhile to sign up for insurance because the premium is so high.

3.1 Introducing Dynamics

To study the circumstances under which initial conditions could affect which competitive equilibrium is reached, a simple dynamic version of this static model is required. The most straightforward way to introduce dynamics is to assume that, in each period, premiums reflect the average cost of the pool of individuals who purchase insurance, thereby allowing insurers to always break even. But, given this premium, consumers decide whether to enroll for insurance the next period. Cutler and Reber (1998) and Monheit et al. (2004) provide some evidence for this pattern of price and demand adjustments.

These dynamics can be conveniently illustrated graphically using the same type of diagram as before. Recall that, for any critical type π_t in period t , we can read the premium from the average cost curve by setting $p_t = \Gamma(\pi_t)$. The consumers' reaction in $t + 1$, therefore, can then be read off the demand curve to obtain a new marginal buyer $\pi_{t+1} = \Omega^{-1}(p_t)$, and so forth. This leads to the recursion $\pi_{t+1} = \Omega^{-1}(\Gamma(\pi_t))$ for the evolution of marginal buyers, as illustrated in Figure 3. It immediately implies that π increases (i.e. there is unraveling where the premium increases and fewer consumers

Figure 3: Dynamics and Equilibrium Stability



sign up for insurance) whenever $\Gamma(\pi) > \Omega(\pi)$ while π falls otherwise (more consumers demand insurance, so the premium falls).

We can see that, of the three competitive equilibria here, only two are stable, whereas the intermediate one with the marginal buyer π_2^* is unstable. Which competitive equilibrium is eventually reached depends on the initial value of π , as formalized in the following proposition:

Proposition 1. (i) Initial conditions $\pi \in [0,1)$ matter for which competitive equilibrium is reached only if there exist at least three competitive equilibria.

(ii) When there are exactly three equilibria with critical types $\pi_1^* < \pi_2^* < 1$, the intermediate equilibrium with marginal buyer π_2^* is generically unstable while the other two are stable.

(iii) In this case, π_2^* is the critical threshold for initial conditions: for any initial $\pi > \pi_2^*$, there is unraveling to the “bad” stable equilibrium where $\pi^* = 1$. For any $\pi < \pi_2^*$, the “good” stable equilibrium with critical type π_1^* is reached.

Proof. See Appendix. ■

With three equilibria, unraveling occurs starting from initial conditions to the left of π_1^* and to the right of π_2^* , whereas the dynamics imply falling premiums and more individuals enrolling otherwise. Evidence for such dynamics have been documented in states which, before the ACA, which placed restrictions on adjusting premiums based on age

and preexisting conditions. Writing about the New Jersey Individual Health Coverage Program (IHCP) started in 1993, Monheit et al. (2004) found dynamics similar to those shown in Figure 3 for values of $\pi > \pi_2^*$ (or $\pi < \pi_1^*$). In particular, between the end of 1995 and the end of 2001, enrollment fell from 186,130 individuals to just 84,968, with premiums rising by 200% to 300%. Three other states — Kentucky, New York and Vermont — tried health care reforms with similar consequences (Cohn, 2012). Cutler and Reber (1998) provide evidence for gradual unraveling of high coverage plans in a setting with employer provided insurance.

As widely reported in the popular media reports, the three initial conditions described in Section 1 would quite reasonably discourage lower-risk consumers from enrolling in the health exchanges relative to higher-risk consumers. In the context of our model in this section, we would expect consumers with large values of π to be the first to enroll, potentially leading to being trapped in the bad stable equilibrium. Of course, by Proposition 1, these mechanics only matter if multiple equilibria actually exist in the first place, a topic to which we now turn.

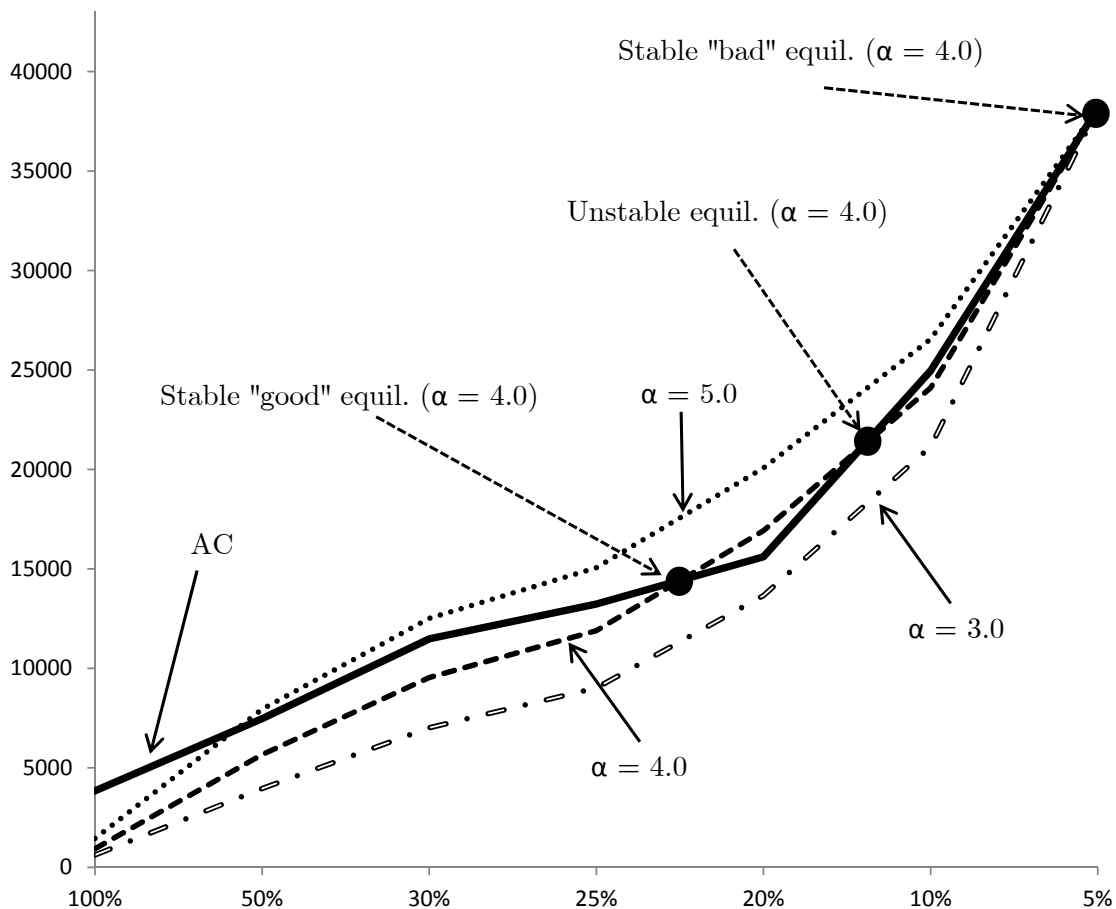
3.2 A Calibration based on the Medical Expenditure Panel Survey

To examine the potential for the type of multiple equilibria shown in Figure 3, we now present a simple quantification of the model. This subsection provides calculations corresponding to our model developed above, where household types differ in their probability of an identical loss. Appendix B presents a calibration based on the case where household types face the same loss probability across different loss amounts. The key lessons are the same in both sets of calculations. Moreover, the generalization in Appendix B demonstrates that similar quantitative exercises could be performed allowing for richer forms of heterogeneity.

Figure 4 shows the average cost curve as well as the willingness-to-pay curves at different levels of risk aversion. The horizontal axis corresponds to the top $X\%$ percent of spenders, where X (the “rank”) is the shown value. The vertical axis is denominated in dollars. The average cost curve (AC) is calculated from the Medical Expenditure Panel Survey for the pre-Medicare population (ages 18 - 64) using data provided to us by Cohen and Uberoi (2013). The average cost curve simply sorts medical spenders by percentile.¹⁴ For example, the mean health expenditure per person in the top 100% of the population (i.e., the entire population mean) is equal to \$3,844, increasing to \$7,476 for population in

¹⁴We are grateful to both of these authors for providing some additional data related to the pre-Medicare (ages 18 - 64) population that are not shown in their paper.

Figure 4: Willingness to Pay and Average Costs: Median Net Worth



Source: Medical Expenditure Panel Survey (Cohen and Uberoi 2013).

Explanation: Wealth equals net worth (assets less liabilities), including net housing wealth.

the top 50%, and climbing to \$38,147 for population in the top 5%.

We calibrate the marginal willingness to pay for insurance for each of the shown percentiles as follows. First, we assume throughout a constant relative risk aversion utility function $u(c) = c^{1-\alpha} / (1 - \alpha)$, where α is the level of risk aversion. Second, for this calibration, the constant loss value l is derived from equation (4) by using the average cost of the top 5% of the population from the MEPS and setting with $\pi_{5\%} = 1$ for them. Third, given this fixed loss value, the (marginal) value of π is then calculated recursively (from the top) at each value of $X\%$ by solving equation (4) for π .¹⁵ (Hence, the value of π increases as the shown value of $X\%$ decreases.) Finally, for a given value of α , the demand

¹⁵For instance, to compute the probability $\pi_{10\%}$ of the loss for the top 10-5% of spenders, we solve $(0.5\pi_{10\%} + 0.5\pi_{5\%})l = AC_{10\%}$, where we take the average costs $AC_{10\%}$ and $l = AC_{5\%}$ from the MEPS data and set $\pi_{5\%} = 1$. $\pi_{20\%}$ is then obtained from solving $(0.5\pi_{20\%} + 0.25\pi_{10\%} + 0.25\pi_{5\%})l = AC_{20\%}$.

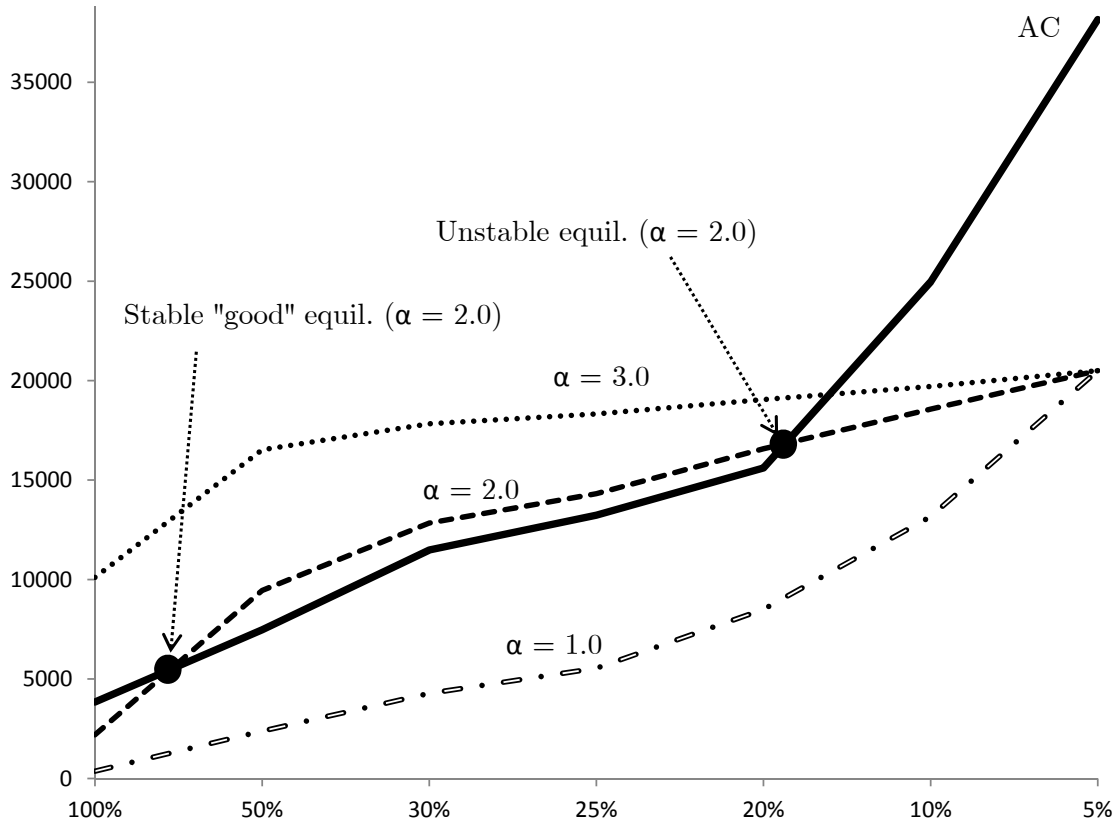
curve is then calculated by solving equation (5) for the value of Ω for each value of π , and hence $X\%$. The value of wealth w in equation (5) is initially set equal to the median net worth found in the 2010 Survey of Consumer Finances (Board of Governors 2012), which assumes that the probability of a loss is independent of the household's wealth.

Consider first the case of $\alpha = 3$. Notice that the willingness-to-pay curve is always smaller than the AC curve (i.e., $\Omega(\pi) < \Gamma(\pi)$) except at the top rank (the smallest shown value of $X\%$) where both curves join. This outcome corresponds to the full unraveling case shown previously in Figure 1. Intuitively, at this comparatively small level of risk aversion, agents with a smaller loss probability π (located at *larger* values of $X\%$ on the horizontal axis) are willing to forgo insurance, pushing up its average cost, thereby leading to unraveling as the value of $X\%$ gets smaller. Now, consider the case of $\alpha = 5$. In this case, the willingness-to-pay curve intersects the AC curve just once before again joining the AC curve at the smallest value of $X\%$. Since at the the good equilibrium (close to the top 50% of spenders), the demand curve intersects the cost curve from below, we know from Section 3 that it is stable, whereas the equilibrium at the top 5% is unstable. Starting from any initial condition, the dynamics will bring the market to the stable equilibrium located at the larger rank, with more than half of the population being covered, consistent with a comparatively large level of risk aversion.

Finally, consider the in-between case where $\alpha = 4$. Notice that the willingness-to-pay and AC curves now intersect at three places, at the shown stable "good" and "bad" equilibria and an unstable intermediate equilibrium. Notice that the sharply rising AC curve as rank grows smaller plays a critical role in causing these multiple intersections. More generally, multiple equilibria are more likely to be produced as losses become more concentrated. Indeed, health costs are much more concentrated than most other types of insurable losses. By Proposition 1, initial conditions matter here: if we start from a situation to the right of the unstable equilibrium (covering somewhere between the top 10 and 20% of the population in terms of spending), the dynamics converge to the worst equilibrium with only the top 5% covered. Otherwise, we eventually reach the best equilibrium with the critical quantile between the top 20 and 25%.

To check the robustness of our results to various assumptions, Figure 5 repeats the same calculations assuming that wealth w is now set equal to the median *liquid assets* reported in the 2010 Survey of Consumer Finances. (Hence, the AC curve remains unchanged.) Liquid assets are a potentially more accurate measure of the relevant amount of wealth when illiquid assets, mainly housing, cannot be legally confiscated to pay for medical bills. Because the value of liquid assets is smaller than the median net worth, we can consider relatively smaller values of α in our comparisons.

Figure 5: Willingness to Pay with Median Liquid Assets



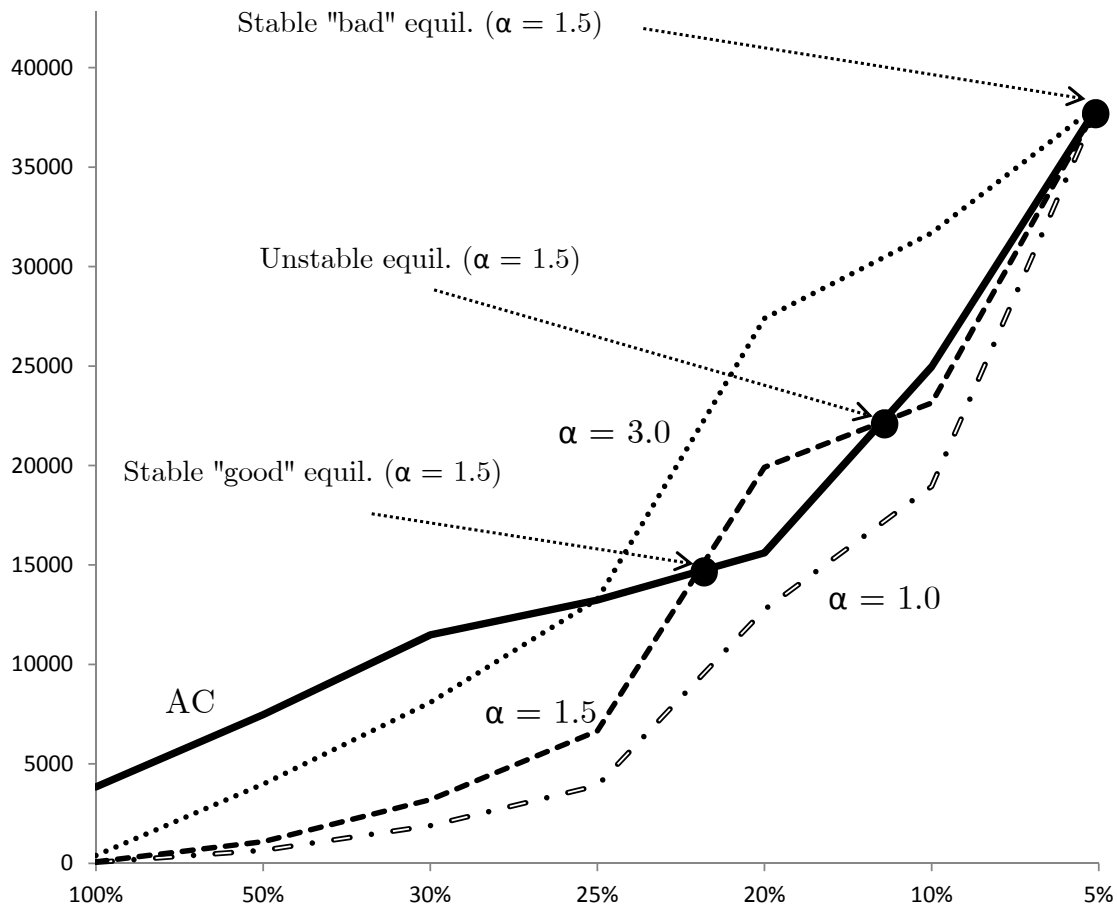
Source: Medical Expenditure Panel Survey (Cohen and Uberoi 2013).

Explanation: Wealth equals liquid assets only.

As before, notice that the smallest value of α , now set equal to 1, produces a willingness-to-pay curve that is always below the AC curve, corresponding to the case of full unraveling. The largest value of α , now set equal to 3, produces just one intersection (corresponding to an unstable equilibrium), as well as a stable corner equilibrium where everyone gets insurance (since the calibrated willingness to pay for insurance of the bottom 50% of spenders still exceeds the average cost of the entire population). The in-between value of α , now set equal to 2, again produces three equilibria. While the willingness to pay of the top 5% is less than their average cost, we can think of a bad equilibrium corresponding to full unraveling (covering even less than the top 5%). The reason why the highest willingness to pay no longer also joins the AC curve at its highest point is due to *limited liability*. The calibrated loss amount l now exceeds the wealth level w , and so the maximum potential loss is capped at w .¹⁶ This demonstrates that the presence of limited

¹⁶In the actual simulations, we cap the loss at w less \$1,000. Not only does this threshold avoid “almost” infinite marginal utility states, it roughly corresponds with Medicaid qualifications as well.

Figure 6: Willingness to Pay with Median Liquid Assets and Risk Increasing by Age



Source: Medical Expenditure Panel Survey (Cohen and Uberoi 2013).

Explanation: Wealth equals liquid assets only.

liability can enhance the likelihood of equilibrium multiplicity by reducing the slope of the willingness-to-pay curve in the same neighborhood where the slope of the AC curve is increasing.

By focusing on median wealth for all cost quantiles, our estimates, however, have not accounted for the fact that both the size of wealth and the probability of loss tend to increase in age. For additional robustness, Figure 6 shows the effect of assuming that rank now grows linearly in age, where 18-year olds are now effectively located at the 100% mark on the horizontal axis while 64-year olds are located at the 5% mark. We can now also use the median values of liquid assets at each age from the 2010 Survey of Consumer Finances. Notice that allowing for this relationship has very little impact on our results. Relative to Figure 5, the effects of limited liability are now absent because older people tend to have both more assets and a higher probability of loss.

Appendix B presents additional estimates that relax the assumption that the size of loss is fixed across different risk types. In particular, the probability of loss is now held fixed across households, and a risk type is now defined in terms of the heterogeneous size of loss. As before, a low level of risk aversion α leads to unraveling. Moreover, the in-between level of risk aversion produces multiple stable equilibria. However, unlike before, even large values of risk aversion now produce multiple stable equilibria. Overall, therefore, the empirical evidence is consistent with the potential for multiple equilibria, especially for moderate values of risk aversion and when loss sizes are not fixed.

Of course, we view these calibrations as merely a first step towards demonstrating the possibility of multiple equilibria in insurance markets. In a more complete exercise, which is beyond the scope of the current paper, the full MEPS data could be exploited to obtain the entire average cost curve on a fine percentile grid, and the marginal willingness to pay curve could be calibrated for each quantile using data on both wealth and medical spending. Moreover, we have considered data from the entire population aged 18-65 rather than just the subset of individuals most likely to demand insurance in the ACA health exchanges. Finally, our calculations have abstracted from the various fines and subsidies in place, to which we turn in the next section.

4 Mandate Enforcement through Fines

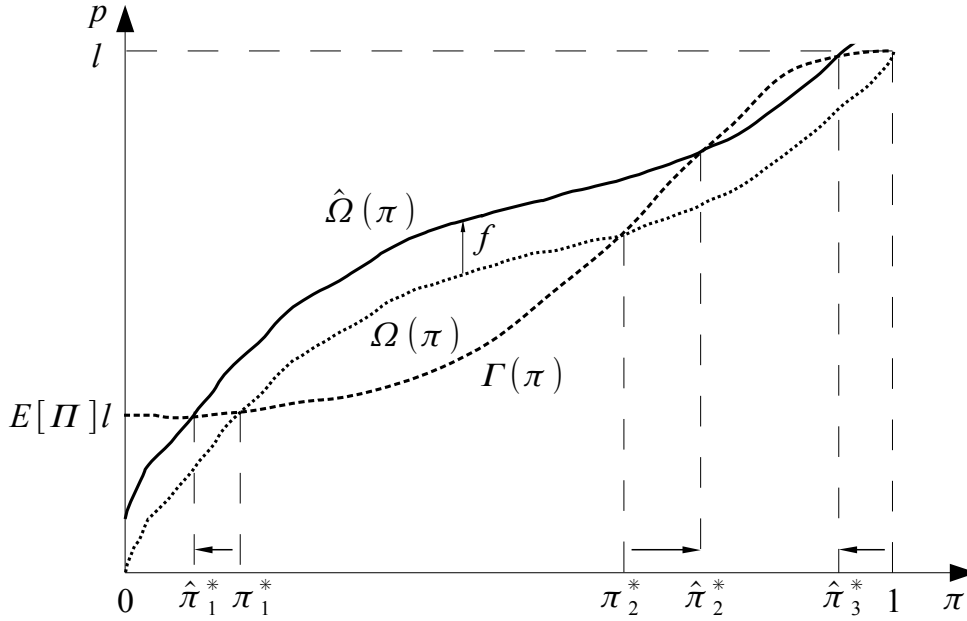
Given the previous experience in the states, the mandate — in reality, the associated *fine* that gives the mandate its force — is widely viewed as critical to the success of the ACA. Indeed, the mandate was the focus of the challenge to ACA heard by the Supreme Court in *National Federation of Independent Business v. Sebelius*.¹⁷ Chandra et al. (2011) present evidence that the phase-in of the mandate in the Massachusetts plan encouraged healthier consumers to enroll.

In this section, we show that the actual *form* of the mandate, and not just its size *per se*, plays an important role in the presence of equilibrium multiplicity. In addition to the fine, the ACA also makes subsidies available to households with lower income.

We begin by demonstrating how the introduction of a fine f for not having insurance as well as a subsidy s for having insurance can be captured in our graphical framework. Of course, neither the fine nor the subsidy affects the average cost curve $\Gamma(\pi)$. However, it affects the construction of the willingness to pay for insurance, now denoted as $\hat{\Omega}(\pi)$,

¹⁷The mandate was upheld as a valid exercise of Congress' power under the Taxing Clause. The case, heard along with *Florida v. Department of Health and Human Services*, also considered the expansion of the Medicaid system.

Figure 7: Enforcement of a Mandate through a Fine f



through the modified indifference condition

$$u(w - \hat{\Omega} + s) = \pi u(w - f - l) + (1 - \pi)u(w - f). \quad (6)$$

Notice that the subsidy s and the fine f both *shift up* the inverse demand curve $\hat{\Omega}(\pi)$, in fact, in a parallel manner in the case of a subsidy. We now focus on the construction of the fine f given its greater importance in the recent debate.

Unsurprisingly, a fine can give rise to better equilibria with more individuals insured. In the example shown in Figure 1 where only the complete unraveling equilibrium exists, shifting up the $\hat{\Omega}$ -curve will induce the emergence of equilibria with a positive mass of individuals getting coverage. In fact, this upward shift could induce multiple equilibria where previously there was only one. For the example shown in Figure 3 with three equilibria, Figure 7 illustrates how shifting up the $\hat{\Omega}$ -curve can shrink the range of initial values $(\hat{\pi}_2^*, 1]$ for which unraveling to the bad stable equilibrium π_3^* occurs. At the same time, it also shifts both the good stable equilibrium π_1^* and the bad stable equilibrium π_3^* to the left and, hence, leads to a greater number of individuals being covered at a lower premium.

Of course, we can always set the fine to be large enough such that there exists a unique equilibrium where everyone buys insurance and there is no risk of unraveling. However,

this outcome may be both inefficient and politically challenging. Indeed, the peculiar nature of the fine’s construction under the ACA — namely, its assessment only on tax filers who are owed a refund — reflects the sensitivity that Congress felt it faced in creating a fine that causes too much hardship.

A more interesting question is whether there exist another fine mechanism that eliminates the possibility of equilibrium multiplicity *without* imposing a higher fine *in the best equilibrium* $\hat{\pi}_1^*$. As Figure 7 makes clear, it is actually not necessary to impose a higher fine *everywhere* in order to eliminate the bad stable equilibria. A large fine value is only necessary in situations where few individuals enroll for insurance, that is, where the critical value of π^* and, hence, the premium are both large.

In contrast to the absolute dollar fine — or a percentage of income, whichever is greater — that exists under the ACA, a *relative* fine that is tied to the actual equilibrium premium in the market achieves exactly this. As noted in Section 1, a relative fine was part of Massachusetts health care reform law enacted in 2006. Using the dynamics developed in Section 3, suppose that in each period t , the fine that must be paid by uninsured consumers is set equal to kp_t , where $k > 0$ is some constant that can be interpreted as the percentage of the current premium p_t .¹⁸ Since $p_t = \Gamma(\pi_t)$, the resulting willingness to pay for insurance, now denoted as $\tilde{\Omega}(\pi)$, for each risk level π is then defined implicitly by

$$u(w - \tilde{\Omega}) = \pi u(w - l - k\Gamma(\pi)) + (1 - \pi)u(w - k\Gamma(\pi)).^{19} \quad (7)$$

The benefit of the relative fine is that the fine value — and, hence, consumers’ demand for insurance — automatically increase as the market unravels towards a bad stable equilibrium. This outcome occurs *even if* we choose k such that $k\Gamma(\hat{\pi}_1^*) = f$, so the relative and the absolute fines take exactly the same value in the best equilibrium $\hat{\pi}_1^*$. These dynamics are illustrated in Figure 8, which shows how the inverse demand curve under the relative fine $\tilde{\Omega}(\pi)$ is a counter-clockwise rotation at point $\hat{\pi}_1^*$ relative to the inverse demand curve under the absolute fine $\hat{\Omega}(\pi)$. Proposition 2 formalizes the advantages of a relative fine compared to an absolute fine with this normalization.

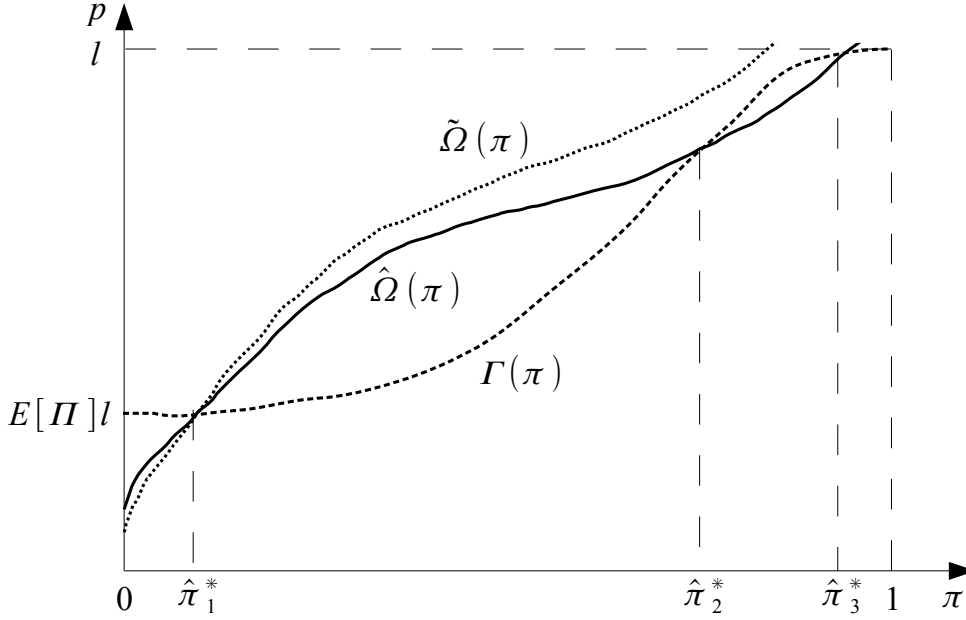
Proposition 2. *Let $\hat{\pi}_1^* < 1$ be the best equilibrium under an absolute fine $f > 0$, and set the relative fine such that $k\Gamma(\hat{\pi}_1^*) = f$ (i.e., equal fine values at the best equilibrium). Then:*

(i) *for any number $N \geq 1$ of equilibria, the worst equilibrium $\tilde{\pi}_N^*$ under the relative fine is Pareto better (more coverage at a lower price) than the worst equilibrium under the absolute fine $\hat{\pi}_N^*$, i.e. $\tilde{\pi}_N^* \leq \hat{\pi}_N^*$.*

¹⁸In the case of Massachusetts, the corresponding value of k would roughly equal 1/2.

¹⁹Notice that, since $\Gamma(\pi)$ is increasing, the right-hand side of (7) is still decreasing in π , and so $\tilde{\Omega}(\pi)$ remains well-defined and increasing.

Figure 8: Relative versus absolute fine with $k\Gamma(\hat{\pi}_1^*) = f$



(ii) The best equilibrium under the absolute and relative fine are identical, i.e. $\hat{\pi}_1^* = \tilde{\pi}_1^*$.

(iii) The interval of initial conditions $[0, \tilde{\pi}_1)$ from which we converge to the best equilibrium $\tilde{\pi}_1^*$ under the relative fine is larger than the range of initial conditions $[0, \hat{\pi}_1)$ from which we converge to the best equilibrium $\hat{\pi}_1^* = \tilde{\pi}_1^*$ under the absolute fine, i.e. $\tilde{\pi}_1 \geq \hat{\pi}_1$.

Proof. See Appendix. ■

In sum, re-construction of the fine toward a relative basis is more likely to expand coverage, by moving the market to the good stable equilibrium, without costing non-insured consumers anything more in the good equilibrium. However, even if the bad stable equilibrium does emerge (which may still be possible under a small enough relative fine), its “badness” is also reduced (more coverage at a lower price).

It is also worth pointing out that, in the context of this model, introducing a fine for non-participation (or other forms of enforcing a mandate) can never lead to a Pareto improvement. This is because there are always individuals with sufficiently low π who, in any equilibrium without a fine, prefer to demand no insurance. With a fine in place, they will either remain uninsured and pay the fine or, if the fine is high enough, buy insurance at a premium that is higher than their original willingness to pay. In either case, they will be worse off. Hence, a fine or mandate can increase coverage but not in a Pareto

improving way.²⁰

Even when relaxing the assumption that π has full support on $[0, 1]$ as in Appendix B, a Pareto improvement from introducing a fine is possible only in the presence of multiple equilibria, when the fine induces a shift from a bad equilibrium with low coverage to a good equilibrium where in fact *everyone* gets covered, and everyone being covered is an equilibrium *even without the fine*. Only this guarantees that nobody ends up paying the fine and even the lowest risk types actually prefer buying insurance when everyone does so, so they are better off compared to the bad equilibrium (see e.g. Figure 10 in Appendix B). This again demonstrates the importance of explicitly accounting for equilibrium multiplicity, which turns out to crucially underlie standard arguments for Pareto improving mandates or fines in the context of adverse selection.

5 Strategic Insurers

So far, we have focused on the notion of competitive equilibrium where insurers act as price-takers, as in Akerlof (1970). A natural question, therefore, is whether equilibrium multiplicity and, hence, the role of initial conditions also extend to situations in which insurers strategically set premiums rather than taking them as given.

Following Einav et al. (2010) in the insurance market setting and Mas-Colell et al. (1995) in the labor market setting, suppose there are at least two insurers who set premiums in a two-stage Bertrand game. In the first stage, insurers simultaneously announce their premiums. In the second stage, individuals decide whether to purchase insurance and, if so, from which insurer.²¹ Of course, each consumer's risk level π is still private information. But we now assume that there is common knowledge of the distribution $H(\pi)$, consumer wealth w , the loss amount l , and the form of utility $u(c)$. Hence, the shapes of $\Gamma(\pi)$ (the average cost curve) and $\Omega(\pi)$ (the willingness to pay curve) are common knowledge as well.

Returning to the setting without any fines, the following proposition, which is easily adapted from Mas-Colell et al. (1995) and Einav et al. (2010), shows that equilibrium multiplicity disappears.²²

²⁰This argument goes through unaffected when the revenue from the fine is returned lump-sum to all individuals, or when the fine on non-participants is replaced by a subsidy for participation that is financed by a lump-sum tax.

²¹As usual, to break a tie (since actual currency denominations are technically a countable set to the penny level), if multiple insurers announce exactly the same premium levels, individuals then randomize among them with equal probabilities.

²²See also Pauly (1974) for an informal argument.

Proposition 3. *With strategic insurers, the unique subgame perfect equilibrium outcome of the above two-stage game involves the critical type*

$$\pi_1^* = \min \{ \pi \in [0, 1] \mid \Omega(\pi) = \Gamma(\pi) \}.$$

In words, when insurers set premiums strategically, only the best competitive equilibrium (with the lowest premium and the most people covered) survives. This result holds even if there are more than one *competitive* equilibrium, that is, multiple equilibria in the price-taking model.

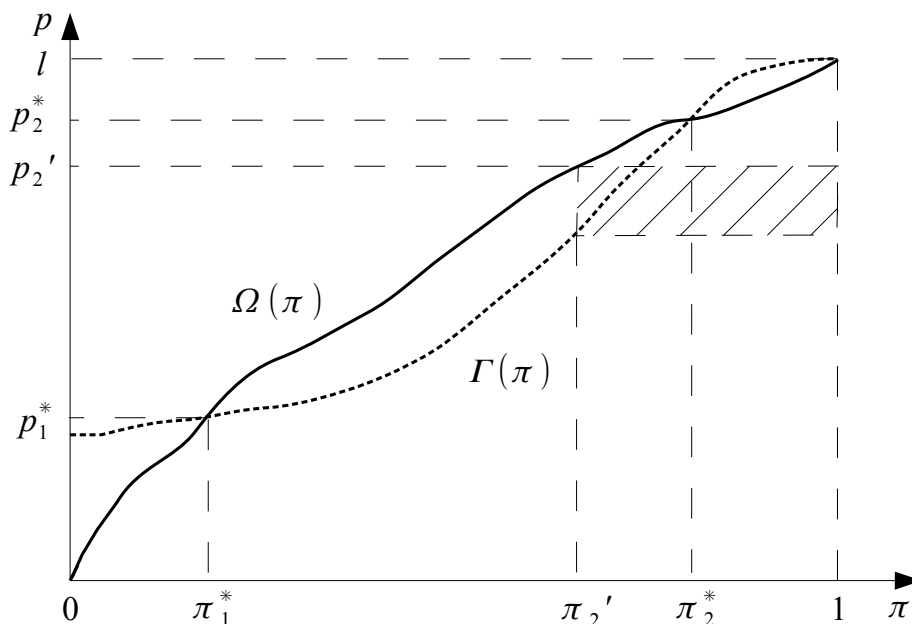
Intuitively, consider again the setting shown earlier in Figure 3 with exactly three competitive equilibria. Figure 9 illustrates the mechanics when firms now behave strategically. Suppose we are in the intermediate equilibrium with critical type π_2^* and premium p_2^* . If all insurers set premium p_2^* , all types $\pi \geq \pi_2^*$ demand insurance, and the average cost of this pool is $\Gamma(\pi_2^*) = p_2^*$, and so all insurers make zero profits. Hence, this outcome is a *competitive* equilibrium in the sense of Definition 1.²³ But it also emits a profitable deviation by any strategic insurer. In particular, suppose that an insurer deviates and sets a premium $p'_2 < p_2^*$. As drawn in Figure 9, this insurer will capture the entire market with demand from all types $\pi \geq \pi'_2$ and $\pi'_2 < \pi_2^*$. Moreover, at π'_2 , the average cost curve is below the demand curve, so $\Gamma(\pi'_2) < p'_2$. Hence, offering the premium p'_2 will result in strictly positive profits for the deviating insurer corresponding to the dashed area in Figure 9. The only competitive equilibrium from where there is no such profitable deviation is the good equilibrium with marginal buyer π_1^* and the lowest premium p_1^* .

Of course, if, alternatively, we started from the setting with just a single equilibrium with full unraveling, as shown earlier in Figure 1, that unique equilibrium in the price-taking setting remains the unique equilibrium in the strategic setting. Hence, assuming that premiums are set strategically does not necessarily imply that many households obtain insurance.

Strategic premium setting in a framework of Bertrand-like competition may seem like the more relevant case than price-taking for insurance markets, since many insurers are not atomistic and do actively set premiums taking into account their competitors' and customers' responses to their actions. However, as e.g. Mas-Colell et al. (1995) emphasize in the related setting of labor markets with adverse selection, the outcome in Proposition 3 relies on the rather strong assumption that firms have common knowledge about all market fundamentals, including the entire shape of the demand and cost curves Ω and Γ .

²³As shown earlier, this particular equilibrium is unstable in the context of dynamics. But this fact is immaterial for our argument here to show that this equilibrium does not survive under strategic premium setting. The same argument can be made for any stable equilibrium that is worse than the best equilibrium.

Figure 9: Equilibrium Uniqueness with Strategic Insurers



In contrast, in the competitive equilibria with price-taking considered in Sections 2 to 4, insurers only need to know the average cost of those who buy insurance at the going premium, but they do not need to know anything about the preferences or risk distribution underlying this or away from the current situation.

In the strategic setting, a *mistaken* attempt at a profitable deviation could lead to substantial losses. For instance, suppose we start from the worst equilibrium $\pi^* = 1$ in Figure 9. If an insurer deviates by offering a marginally lower premium $p = l - \varepsilon$, this will lead to losses since $\Gamma(\pi) > \Omega(\pi)$ for π close to one. To make profits, a deviating insurer would have to offer a discretely lower premium $p < p_2^* < l$, but also not too low, since losses would be incurred again if $p < p_1^*$. This demonstrates that rather precise knowledge on the side of insurers about market conditions potentially far away from the current situation is required to make the argument underlying Proposition 3 work.

In the absence of perfect information about the market structure, insurers may have to experiment with their pricing to find out about market fundamentals, or may prefer local adjustments to premiums in a backward looking rather than strategic way, as documented in Cutler and Reber (1998).²⁴ As noted in Section 1, transitory losses from pricing mistakes or experimentation could be magnified by the fact that regulations prevent insurers from changing prices frequently or sharply increasing premiums. Once set, ACA

²⁴See also the discussion in Section 3. Such behavior would effectively make them price-takers again.

plan premiums are generally viewed as locked until the next open enrollment period (Kaiser 2013) .

Moreover, under the McCarran-Ferguson Act of 1945, individual states typically regulate the business of insurance, and most states already require some steps before rates can be increased (National Conference of State Legislatures 2013).²⁵ However, because rules vary between states, Title I (Subtitle A, Sec. 1003) of the ACA creates a more uniform standard around rate increases. These rules include requiring states to collect premium information and determine if plans should be excluded from the health exchange based on unjustified premium increases.²⁶ If an insurer requests a premium increase above 10%, a more detailed explanation must be provided and posted on their and the HHS website. The ACA also makes \$250 million available to states to take action against insurers requesting unreasonable rate increases. According to the Centers for Medicare and Medicaid Services (2010) , “[t]his funding will help assure consumers in every state that any premium increases requested by their insurance company, regardless of size, is justified.”

In sum, an insurer who strategically guessed a wrong price could be stuck with that decision for a lengthy amount of time and then face resistance to large, corrective price increases. As a result, the consumer-protection legislation intended to protect consumers from large price increases could essentially force insurers into price-taking behavior that includes the potential of getting stuck in the bad equilibrium with high prices. In the absence of common knowledge, insurers may have more incentives to strategically seek profitable deviations — thereby work toward the good equilibrium with lower prices — if they could more easily experiment with price changes.

6 Conclusion

This paper introduces simple dynamics into a standard insurance model with adverse selection and characterizes when the “initial conditions” of a new insurance market—like a website failure at launch—could have permanent consequences. We show that initial conditions can be material if (i) insurers are competitive price-takers, and if (ii) there exist at least three competitive equilibria. In the case of exactly three equilibria, one equilibrium is unstable while the other two are stable. A “good” stable equilibrium is Pareto

²⁵At least two dozen states require that the insurer receives prior approval from the state insurance commissioner or department before increasing health insurance premiums (National Conference of State Legislatures 2013).

²⁶For a few states—Alabama, Louisiana, Missouri, Oklahoma, Texas, and Wyoming—these determinations will be made by the federal government since these states do not have effective review processes in place.

superior—by offering more coverage at a lower price—to a “bad” stable equilibrium. We provide some suggestive empirical evidence that the presence of three equilibria is indeed consistent with moderate levels of risk aversion. Multiple equilibria are more likely to emerge when losses are very concentrated (as in the case of health care) and in the presence of limited liability. Future work can provide a more detailed empirical analysis using a wider range of data sets.

As is well known, the Affordable Care Act (ACA) levies a fine on non-participants in order to try to prevent unraveling. Without the presence of multiple equilibria, a fine cannot expand coverage in a Pareto improving manner, but it effectively redistributes from low to high risk types. Equilibrium multiplicity is a necessary (but not sufficient) condition for a fine to achieve a Pareto improvement. The ACA’s fine, however, is constructed as an absolute amount, equal to the greater of a fixed dollar amount or a fixed fraction of income. In contrast, the 2006 Massachusetts plan, on which the ACA is modeled, levied a fine on a relative basis, equal to a fraction of the equilibrium premium. The relative fine, therefore, grows with the amount of adverse selection. We show that changing the fine from an absolute to a relative amount—normalized to be equal in the desired, good equilibrium—increases the range of initial conditions consistent with reaching the good equilibrium, while also reducing the severity of the bad equilibrium, if it still exists.

If insurers price strategically, rather than acting like price takers, only the good equilibrium emerges. In particular, any attempt to price at the bad equilibrium emits profitable deviations. However, strategic pricing also requires insurers to fully know the distribution of risk types and participant characteristics (preferences, loss amounts, and wealth). Incorrect deviations can be costly across a wide range of guesses: If an insurer reduces prices too little, they will continue to be stuck with high-risk types who are now paying below the fair rate required for zero profits. But if the firm reduces prices too much, it still can lose money even if it favorably changes the risk pool. Existing evidence from previous health care reforms at the state level and from some employer-based plans suggest that insurers instead update their prices more consistently with the price-taking model. While there could be good reasons for limitations to the frequency of price changes and the amount of increases, such as consumer protection, an unintended consequence could be that they further discourage price discovery, thereby increasing the potential for reaching a Pareto dominated equilibrium.

References

- AKERLOF, G. (1970): "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84, 488–500.
- ANDERSON, J., AND S. COWAN (2013): "How to get Health Insurance while Avoiding Exchanges," Discussion Paper Vol. 19, No. 16, The Weekly Standard, http://www.weeklystandard.com/articles/escape-obamacare_771532.html#.
- ASSOCIATED PRESS (2013): "Policy Notifications and Current Status, by State," Discussion paper, <http://news.yahoo.com/policy-notifications-current-status-state-204701399.html>.
- BLOOMBERG (2013): "Obama Lifts Health Mandate for those with Canceled Plans," Discussion paper, <http://www.bloomberg.com/news/2013-12-20/obama-aides-say-more-to-gain-coverage-under-aca-than-canceled.html>.
- BOARD OF GOVERNORS (2012): "Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances.," *Federal Reserve Bulletin*, 98, 1 – 80.
- CENTERS FOR MEDICARE & MEDICAID SERVICES (2010): "Shining a Light on Health Insurance Rate Increases.," Discussion paper, <https://www.cms.gov/CCIIO/Resources/Fact-Sheets-and-FAQs/ratereview.html>.
- CHANDRA, A., J. GRUBER, AND R. MCKNIGHT (2011): "The Importance of the Individual Mandate - Evidence from Massachusetts.'," *New England Journal of Medicine*, 364, 293–295.
- COHN, J. (2012): "The New Jersey Experience: Do Insurance Reforms Unravel Without An Individual Mandate?," Discussion paper, <http://www.kaiserhealthnews.org/Stories/2012/March/21/nj-ind-mandate-case-study.aspx>.
- CUTLER, D., AND S. REBER (1998): "Paying for Health Insurance: The Tradeoff Between Competition and Adverse Selection," *Quarterly Journal of Economics*, 113, 433–466.
- EINAV, L., AND A. FINKELSTEIN (2011): "Selection in Insurance Markets: Theory and Empirics in Pictures," *Journal of Economic Perspectives*, 25, 115–138.
- EINAV, L., A. FINKELSTEIN, AND M. CULLEN (2010): "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 125, 877–921.

- EINAV, L., A. FINKELSTEIN, AND J. LEVIN (2010): "Beyond Testing: Estimating Empirical Models of Insurance Markets," *Annual Review of Economics*, 2, 311–336.
- FELDSTEIN, M. (2013): "Obamacare's Fatal Flaw," Discussion paper, <http://www.project-syndicate.org/commentary/martin-feldstein-on-how-america-s-health-care-reform-could-unravel>.
- HANDEL, B., I. HENDEL, AND M. WHINSTON (2013): "Equilibria in Health Exchanges: Adverse Selection vs. Reclassification Risk," *NBER Working Paper No. 19399*.
- HARVARD UNIVERSITY INSTITUTE OF POLITICS (2013): "Survey of Young Americans' Attitude Toward Politics and Public Service: 24th Edition," Discussion paper, <http://www.iop.harvard.edu/survey-young-americans%E2%80%99-attitude-toward-politics-and-public-service-24th-edition>.
- HENDREN, N. (2012): "Unraveling Versus Unraveling: A Memo on Competitive Equilibria and Trade in Insurance Markets," *Mimeo, Harvard University*.
- INTERNAL REVENUE SERVICE (2013): "Questions and Answers on the Individual Shared Responsibility Provision," Discussion paper, <http://www.irs.gov/uac/Questions-and-Answers-on-the-Individual-Shared-Responsibility-Provision>.
- (2013B): "Filing Season Statistics May 10, 2013," Discussion paper, <http://www.irs.gov/uac/Newsroom/Filing-Season-Statistics-May-10,-2013>.
- KAISER FAMILY FOUNDATION (2013): "Frequently Asked Questions about Health Care Reform," Discussion paper, [Link:http://kff.org/health-reform/faq/health-reform-frequently-asked-questions/](http://kff.org/health-reform/faq/health-reform-frequently-asked-questions/).
- MAS-COLELL, A., M. WHINSTON, AND J. GREEN (1995): *Microeconomic Theory*. Oxford University Press, Oxford.
- MONHEIT, A., J. CANTOR, M. KOLLER, AND K. FOX (2004): "Community Rating And Sustainable Individual Health Insurance Markets In New Jersey," *Health Affairs*, 23, 167–175.
- NETZER, N., AND F. SCHEUER (2013): "A Game-Theoretic Foundation of Competitive Equilibria with Adverse Selection," *International Economic Review*, *forthcoming*.
- PAULY, M. (1974): "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," *Quarterly Journal of Economics*, 88, 44–62.

PEW RESEARCH CENTER (2013): “As Health Care Law Proceeds, Opposition and Uncertainty Persist,” Discussion paper, <http://www.people-press.org/2013/09/16/as-health-care-law-proceeds-opposition-and-uncertainty-persist/>.

ROTHSCHILD, M., AND J. STIGLITZ (1976): “Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Incomplete Information,” *Quarterly Journal of Economics*, 90, 629–649.

TAX POLICY CENTER (2011): “T11-0225 - Tax Units Paying No Income Tax, by Filing Status and Cash Income Level; Baseline: Current Law, 2011-13,” Washington, DC.

WILSON, C. (1977): “A Model of Insurance Markets with Incomplete Information,” *Journal of Economic Theory*, 12, 167–207.

——— (1980): “The nature of Equilibrium in Markets with Adverse Selection,” *Bell Journal of Economics*, 11, 108–130.

A Appendix: Proofs

A.1 Proof of Proposition 1

(i). We show that initial conditions cannot matter when there are only one or two competitive equilibria. If $\pi^* = 1$ is the only equilibrium, we must have $\Gamma(\pi) > \Omega(\pi)$ for all $\pi < 1$. Otherwise, by continuity of Γ and Ω and since $\Gamma(0) = \mathbb{E}[\Pi]l > \Omega(0) = 0$, there would have to exist at least one intersection of Γ and Ω at some $\pi < 1$ and hence another equilibrium. Since $\Gamma(\pi) > \Omega(\pi)$ for all $\pi < 1$, the dynamics imply unraveling to $\pi^* = 1$ for any initial π and therefore initial conditions do not matter.

If there are two equilibria with $\pi^* = 1$ and some $\pi_1^* < 1$, it must hold that $\Gamma(\pi) > \Omega(\pi)$ for all $\pi \in [0, \pi_1^*)$ and $\Gamma(\pi) < \Omega(\pi)$ for all $\pi \in (\pi_1^*, 1)$ by an analogous argument as above. Hence, for any $\pi \in [0, 1)$, we converge to the constrained efficient equilibrium $\pi_1^* < 1$. As a result, initial conditions again do not matter for the equilibrium that is eventually reached except in the non-generic case where the initial $\pi = 1$.

(ii) and (iii). Note first that, for any number N of equilibria $\pi_1^* < \dots < \pi_{N-1}^* < 1$, the best equilibrium π_1^* must be stable generically. This is because $\Gamma(\pi) > \Omega(\pi)$ for all $\pi \in [0, \pi_1^*)$ and, since there is a proper intersection of Γ and Ω generically, $\Gamma(\pi) < \Omega(\pi)$ for all $\pi \in (\pi_1^*, \bar{\pi}_1)$ and $\bar{\pi}_1 > \pi_1^*$ sufficiently close to π_1^* . With exactly 3 competitive equilibria $\pi_1^* < \pi_2^* < 1$, this implies that in fact $\bar{\pi}_1 = \pi_2^*$, and since again there is a proper intersection of Γ and Ω at π_2^* generically, $\Gamma(\pi) > \Omega(\pi)$ for all $(\pi_2^*, 1)$, as illustrated in

Figure 3. Hence, the intermediate equilibrium π_2^* is unstable and the other two are stable. Moreover, we converge to π_1^* for any initial $\pi < \pi_2^*$ and to $\pi^* = 1$ for any $\pi > \pi_2^*$.

A.2 Proof of Proposition 2

Note first that, for any absolute fine $f > 0$, $\hat{\Omega}(1) = l + f > \Gamma(1) = l$, so for any number N of equilibria, both the best equilibrium $\hat{\pi}_1^*$ and the worst equilibrium $\hat{\pi}_N^*$ under f must satisfy $\hat{\pi}_1^* \leq \hat{\pi}_N^* < 1$. We next observe that, comparing the definitions (6) and (7) and using the normalization that $k\Gamma(\hat{\pi}_1^*) = f$ and the fact that $\Gamma(\pi)$ is increasing in π , $\tilde{\Omega}(\pi) < \hat{\Omega}(\pi)$ for all $\pi < \hat{\pi}_1^*$ and $\tilde{\Omega}(\pi) > \hat{\Omega}(\pi)$ for all $\pi > \hat{\pi}_1^*$. We use this repeatedly to prove claims (i) to (iii) in the proposition.

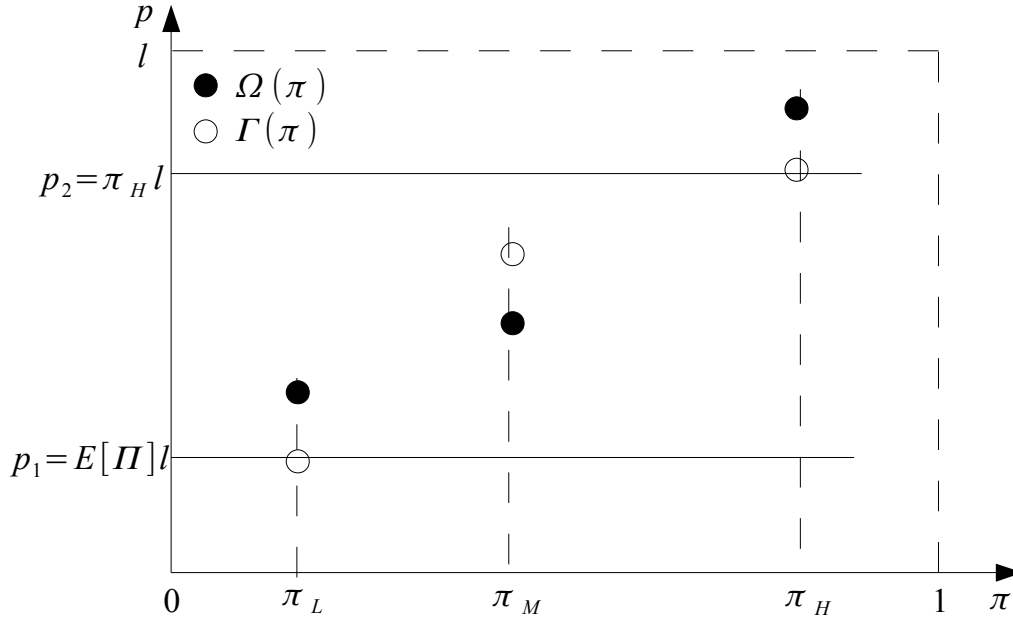
(i). Since $\hat{\Omega}(1) > \Gamma(1)$ under $f > 0$, the worst equilibrium $\hat{\pi}_N^* < 1$ must be such that $\hat{\Omega}(\pi) > \Gamma(\pi)$ for all $\pi > \hat{\pi}_N^*$. Since $\hat{\pi}_N^* \geq \hat{\pi}_1^*$, the above result that $\tilde{\Omega}(\pi) > \hat{\Omega}(\pi)$ for all $\pi > \hat{\pi}_1^*$ a fortiori implies $\tilde{\Omega}(\pi) > \hat{\Omega}(\pi) > \Gamma(\pi)$ for all $\pi > \hat{\pi}_N^*$. This immediately rules out $\tilde{\pi}_N^* > \hat{\pi}_N^*$.

(ii) and (iii). Note first that the best equilibrium $\hat{\pi}_1^*$ is always such that $\hat{\Omega}(\pi) < \Gamma(\pi)$ for all $\pi < \hat{\pi}_1^*$. Moreover, since we observed that $\tilde{\Omega}(\pi) < \hat{\Omega}(\pi)$ for all $\pi < \hat{\pi}_1^*$, we also have $\tilde{\Omega}(\pi) < \Gamma(\pi)$ for all $\pi < \hat{\pi}_1^* \leq \tilde{\pi}_1^*$. Hence, the range of initial values from which we converge to the best equilibrium always takes the form of an interval with lower bound zero and upper bound $\tilde{\pi}_1 \geq \hat{\pi}_1^*$. Since $\tilde{\Omega}(\pi) > \hat{\Omega}(\pi)$ for all $\pi > \hat{\pi}_1^*$ and $\tilde{\Omega}(\hat{\pi}_1^*) = \hat{\Omega}(\hat{\pi}_1^*)$, we also have $\tilde{\pi}_1^* = \hat{\pi}_1^*$, as claimed in (ii).

Suppose first that the best equilibrium $\hat{\pi}_1^* < 1$ is the unique equilibrium under the absolute fine f , so $\hat{\Omega}(\pi) > \Gamma(\pi)$ for all $\pi > \hat{\pi}_1^*$ and vice versa. Then by the above observation that $\tilde{\Omega}(\pi) > \hat{\Omega}(\pi)$ for all $\pi > \hat{\pi}_1^*$ and vice versa, this immediately implies that $\tilde{\pi}_1^* = \hat{\pi}_1^*$ is also the unique equilibrium under the relative fine. It also implies that, in both cases, the best equilibrium is globally stable, so we converge to it for any initial conditions and thus $\tilde{\pi}_1 = \hat{\pi}_1 = 1$.

Otherwise, since the best equilibrium $\hat{\pi}_1^*$ generically corresponds to a proper intersection of $\hat{\Omega}(\pi)$ and $\Gamma(\pi)$ and $\hat{\Omega}(\pi) < \Gamma(\pi)$ for all $\pi < \hat{\pi}_1^*$, we must have $\hat{\Omega}(\pi) > \Gamma(\pi)$ for some interval $(\hat{\pi}_1^*, \hat{\pi}_1)$ with $\hat{\pi}_1 > \hat{\pi}_1^*$. Hence, under the absolute fine, we converge to $\hat{\pi}_1^*$ for any initial π in the interval $[0, \hat{\pi}_1)$. Then the above observation that $\tilde{\Omega}(\pi) > \hat{\Omega}(\pi)$ for all $\pi > \hat{\pi}_1^*$ immediately implies $\tilde{\Omega}(\pi) > \Gamma(\pi)$ for some interval $(\hat{\pi}_1^* = \tilde{\pi}_1^*, \tilde{\pi}_1)$ with $\tilde{\pi}_1 > \hat{\pi}_1$. The range of initial values for π for which we converge to the best equilibrium under the relative fine is therefore $[0, \tilde{\pi}_1)$ with $\tilde{\pi}_1 > \hat{\pi}_1$.

Figure 10: Equilibrium Multiplicity with Three Types



B Appendix: Generalizing the Price-Taking Model

This Appendix generalizes the model of Section 2 to allow for the presence of discrete risk types, a richer amount of heterogeneity between consumers and more general variation in the size of losses.

B.1 Allowing for Discrete Risk Types

We now show multiple equilibria can emerge even when we relax the assumption that the distribution of types $H(\pi)$ is continuous with full support on $[0, 1]$. For example, consider a case with three risk types, $0 < \pi_L < \pi_M < \pi_H < 1$, of low (L), medium (M) and high (H) risk, respectively. Their willingness to pay for insurance $\Omega(\pi)$ is depicted as black dots in Figure 10. The empty circles represent the average costs of insuring the corresponding pools, and so $\Gamma(\pi_H)$ is the cost of only insuring the high risk type H , $\Gamma(\pi_M)$ is the average cost of insuring both the medium M and high risk H types, and $\Gamma(\pi_L) = \mathbb{E}[\Pi]l$ is the average cost of insuring all three risk types.

We have chosen these values such that there are two competitive equilibria: one good equilibrium in which everyone is insured at premium $p_1 = \mathbb{E}[\Pi]l$, and another bad equilibrium in which only the high risk type is insured at a higher premium $p_2 = \pi_H l > p_1$.²⁷

²⁷With discrete types, competitive equilibria involve points with $\Omega(\pi) \geq \Gamma(\pi)$ rather than necessarily

There is no equilibrium where only the medium and high types π_M and π_H are insured, because the average cost $\Gamma(\pi_M)$ for that pool is higher than the willingness to pay of the medium type $\Omega(\pi_M)$, so the medium risk type would not buy insurance at premium $\Gamma(\pi_M)$ and the dynamics would unravel to the bad equilibrium.

Let us connect Figure 10 to the corresponding figures that we drew for the case of a continuum of types in Sections 2 and 3. Filling up the space between the three discrete types naturally leads to Figure 11. We see that, with continuous types and this pattern of curves, there are in fact three equilibria: a stable bad equilibrium, where only types $\pi \geq \bar{\pi}_2$ are insured (with $\pi_M < \bar{\pi}_2 < \pi_H$), an unstable interior equilibrium with critical type $\bar{\pi}_1$ between π_L and π_H , and a stable corner equilibrium where everyone with $\pi \geq \pi_L$ is insured. Notice that Figure 11 is very similar to Figure 3 shown in Section 3. In particular, for initial conditions to matter, the existence of an unstable equilibrium is still required, and the average cost and willingness to pay curves need to intersect at least twice in the interior. Moreover, the marginal buyer $\bar{\pi}_1$ in the unstable equilibrium represents the critical value of initial conditions that determines whether the good or bad equilibrium is reached eventually. The only difference between the discrete and the continuous cases is that the highest risk type in the discrete case may lay within the support shown for the continuous case.

B.2 Richer Forms of Consumer Heterogeneity and Multiple Loss Sizes

It is also straightforward to extend our analysis in Section 2 to allow for richer forms of consumer heterogeneity and multiple sizes of losses. Let the population be indexed by the continuous variable $\theta \in [0, 1]$ with distribution $F(\theta)$. Suppose there are S possible loss levels $l_s(\theta)$ indexed by s , which may differ across θ . The probability that type θ suffers a loss of size s is denoted by $\pi_s(\theta)$, where, of course, $\sum_{s=1}^S \pi_s(\theta) = 1 \forall \theta$. The expected loss for type θ is, therefore,

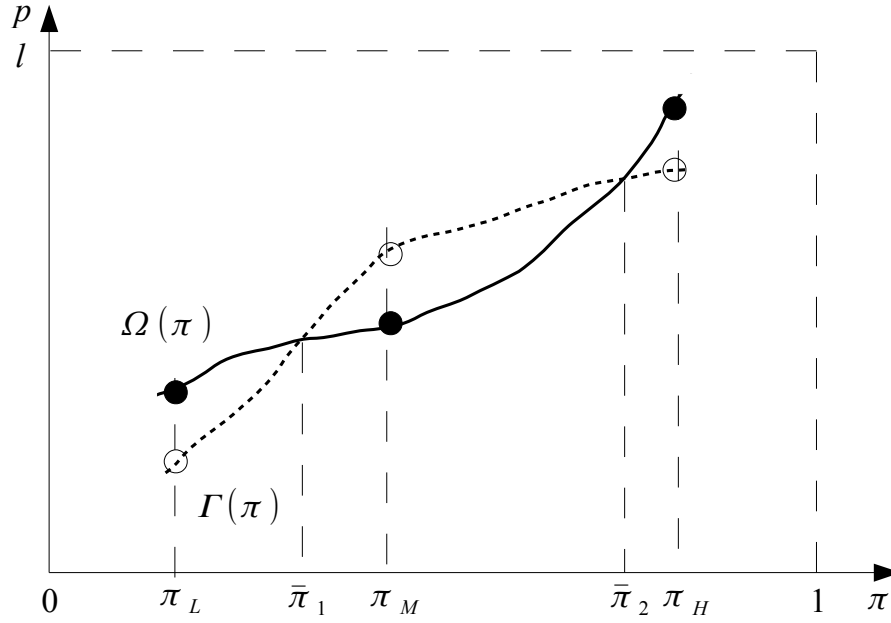
$$\sum_{s=1}^S \pi_s(\theta) l_s(\theta).$$

We can normalize the population type index θ so that the expected costs are increasing in θ . In particular, let us take θ as the *quantiles* of the average cost distribution, so that

$$\Gamma(\theta) = \int_{\theta}^1 \sum_{s=1}^S \pi_s(\theta') l_s(\theta') dF(\theta') \Big/ (1 - F(\theta))$$

$\Omega(\pi) = \Gamma(\pi)$. However, Definition 1 still applies. For instance, in the competitive equilibrium with premium p_2 , we have $\Omega(\pi_L) < \Omega(\pi_M) < p_2 < \Omega(\pi_H)$, so that only the highest type π_H demands insurance. Moreover, $p_2 = \Gamma(\pi_H) = \pi_H l$, and insurers make zero profits.

Figure 11: Connection between continuous and discrete type model



is the average cost of the most costly $1 - \theta$ share of the population, and $F(\theta) = \theta$. Clearly, $\Gamma(\theta)$ is still increasing in θ as before.

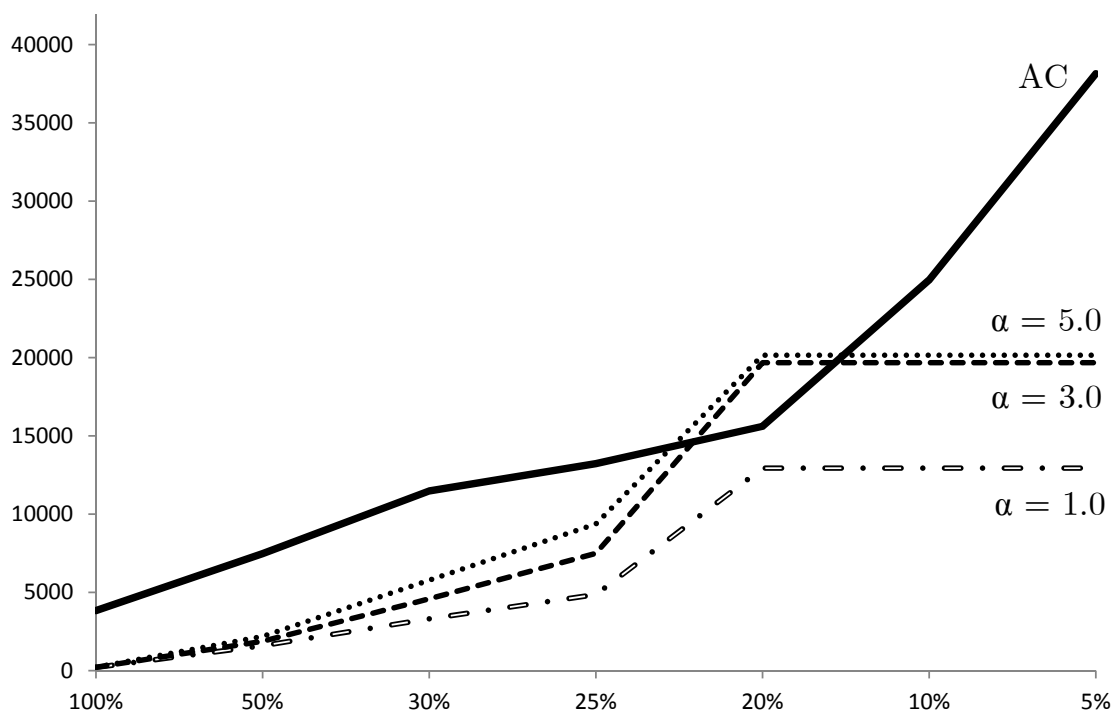
Correspondingly, we can capture the consumers' willingness to pay for insurance for those individuals who are located at the θ -quantile of the cost distribution. Formally, for each quantile θ , let $\Omega(\theta)$ be given by the highest value of Ω such that

$$u(w(\theta) - \Omega; \theta) = \sum_{s=1}^S \pi_s(\theta) u(w(\theta) - l_s(\theta); \theta).$$

Note that we can allow for both wealth levels $w(\theta)$ and preferences (notably risk-aversion) $u(c; \theta)$ to vary across quantiles of the cost distribution; for instance, higher expected cost individuals may on average be wealthier (since older) or more risk-averse (they see the doctor more often).

As long as $\Omega(\theta)$ remains increasing — and, hence, higher expected cost individuals on average have a higher willingness to pay for insurance — our entire analysis from before is maintained: a competitive equilibrium corresponds to a quantile θ where $\Gamma(\theta) = \Omega(\theta)$. We can also employ the same graphical approach as before, the only difference being that the π -axis turns into an θ -axis of quantiles of the cost distribution. At an equilibrium with critical quantile θ^* , the share of the population purchasing insurance is given by $1 - \theta^*$, and so $1 - \theta$ can also be interpreted as quantity of insurance as in Einav et al. (2010).

Figure 12: Willingness to Pay with Median Liquid Assets and Constant Loss Probability



Source: Medical Expenditure Panel Survey (Cohen and Uberoi 2013).

Explanation: Wealth equals net worth (assets less liabilities), including net housing wealth.

Figure 12 shows the empirical evidence from the Medical Expenditure Survey Panel and the Survey of Consumer finances where the probability of loss π is fixed (at 0.3) but the size of loss $l_s(\theta)$ is now allowed to vary across the types. As before, the horizontal axis corresponds to the top $X\%$ percent of spenders, where X (the “rank”) is the shown value. Now, however, the variation in spending comes from differences in loss amounts rather than probabilities. (Given the fixed value of π , a recursive algorithm parallel to that discussed in the text is used to impute the losses across the different values of $X\%$.) As before, the relatively small value of $\alpha = 1$ leads to unraveling. However, both the in-between and large values of α lead to multiple stable equilibria. The driving force is, again, limited liability. As $X\%$ gets small, the value of losses must grow in order to match spending levels in the MEPS. As a result, the willingness to pay is capped for a wider range of types at smaller values of $X\%$ corresponding to larger losses. Increasing the level of risk aversion, therefore, has very little impact on the demand for insurance in this range.