DEMAND ELASTICITY, RISK CLASSIFICATION AND LOSS COVERAGE: WHEN CAN COMMUNITY RATING WORK?

BY
R. GUY THOMAS

ABSTRACT

This paper investigates the effects of high or low fair-premium demand elasticity in an insurance market where risk classification is restricted. The effects are represented by the equilibrium premium, and the risk-weighted insurance demand or “loss coverage”. High fair-premium demand elasticity leads to a collapse in loss coverage, with an equilibrium premium close to the risk of the higher-risk population. Low fair-premium demand elasticity leads to an equilibrium premium close to the risk of the lower-risk population, and high loss coverage – possibly higher than under more complete risk classification. The demand elasticity parameters which are required to generate a collapse in coverage in the model in this paper appear higher than the values for demand elasticity which have been estimated in several empirical studies of various insurance markets. This offers a possible explanation of why some insurance markets appear to operate reasonably well under community rating, without the collapse in coverage which insurance folklore suggests.

KEYWORDS

Adverse selection, loss coverage, risk classification, demand elasticity, community rating.

1. INTRODUCTION

Conventional wisdom or “insurance folklore” suggests that the absence of risk classification in an insurance market is likely to lead to an adverse selection spiral, which may progress until the market largely disappears. A succinct articulation of this concept is provided by the policy document Insurance & Superannuation Risk Classification Policy published by the Institute of Actuaries of Australia (IAA, 1994), which explains:

“In the absence of a system that allows for distinguishing by price between individuals with different risk profiles, insurers would provide an insurance or annuity product at a subsidy to some while overcharging others. In an open

market, basic economics dictates that individuals with low risk relative to price would conclude that the product is overpriced and thus reduce or possibly forgo their insurance. Those individuals with a high level of risk relative to price would view the price as attractive and therefore retain or increase their insurance. As a result the average cost of the insurance would increase, thus pushing prices up. Then, individuals with lower loss potential would continue to leave the marketplace, contributing to a further price spiral. Eventually the majority of consumers, or the majority of providers of insurance, would withdraw from the marketplace and the remaining products would become financially unsound”.

However, the concept that restrictions on risk classification lead inevitably to market collapse is difficult to reconcile with the operation of many extant insurance markets. In particular, there are various markets in which regulators impose some restrictions on risk classification, or even mandate “community rating”, whereby little or no classification of risk is permitted. Examples of restrictions which stop short of pure community rating include the prohibition of rating by gender, race or genetic test results in a number of insurance markets. Examples of community rating in voluntary health insurance include schemes in Ireland, Australia, Switzerland and South Africa, and US states including New York, New Jersey, and Vermont. These schemes generally include some provision for risk equalization payments between insurers, or stop-loss state reinsurance (but not always, and the relevant provisions are not always actually used). However the prevalence and persistence of various community rating schemes or partial restrictions on rating factors does not seem consistent with the notion that regulatory limitations on risk classification lead inevitably to market collapse.

This paper uses a simple model of an insurance market with two risk groups, one lower-risk and one higher-risk. Insurance market outcomes in the absence of risk classification are characterized by the pooled premium charged to all policyholders, and the risk-weighted insurance demand or “loss coverage”. It is shown that insurance market outcome in the absence of risk classification depends on a parameter for the elasticity of demand for insurance at an actuarially fair premium, that is the “fair-premium demand elasticity”. High fair-premium demand elasticity leads to an equilibrium premium close to the risk of the higher-risk population, and a low loss coverage. But for sufficiently low fair-premium demand elasticity, this market collapse does not occur; instead, the market stabilizes with a relatively low premium, and relatively high loss coverage – possibly higher than under more complete risk classification. The ranges for the demand elasticity parameter characterized as “high” and “low” are separated by a threshold range for the parameter which leads to an unstable market outcome – either multiple equilibria (that

---

1 A number of other US states apply varying degrees of restrictions on risk classification in health insurance, such as maximum premium relativities, or allowing some rating factors but not others. Gale (2007) gives further details.
is, more than one pooled premium is capable of equilibrating insurers’ revenue and claims), or near-equilibria for an extended range of premium values. Generally, the threshold range of values for the demand elasticity parameter above which a collapse in coverage occurs in the model in this paper appears higher than the values for demand elasticity which have been estimated in empirical studies of various insurance markets. The high demand elasticity conditions which correspond to the threshold for a large fall in coverage in the model in this paper may explain why some insurance markets appear to operate reasonably well under community rating, without the collapse in coverage which insurance folklore suggests.

A number of previous authors have suggested that insurance folklore accounts of adverse selection spirals might sometimes be overblown. Siegelman (2004) surveys the use of rhetoric about adverse selection in legal judgments and public policy advocacy, drawing a contrast with the limited and sometimes contradictory evidence from empirical studies. Buchmueller & DiNardio (2002) note that “Whilst the notion that community rating leads to adverse selection death spirals appears to have passed into the “conventional wisdom”, at least amongst industry analysts and policy experts, this is not a result which arises naturally from the simplest economic models of insurance”. These authors found no evidence for an adverse selection spiral after the State of New York introduced pure community rating for health insurance in 1993. From a practical perspective, some actuaries may feel that their experience of community rating is more negative. One careful study did report evidence of an adverse selection spiral in health insurance (Cutler & Reber, 1998)\(^2\); but their example was essentially a case of selection against one insurer amongst many in a scenario where different health insurance plans offered differing benefit structures and premiums. However, this is not the same as the selection against the whole market which it is often said will lead to collapse of the market under mandatory community rating. Intuitively, in the scenario of multiple health plans, the various choices are reasonably close substitutes, and so demand elasticity for any one plan may be high; but in the scenario of mandatory community rating, remaining uninsured is often not a close substitute for being insured, and so demand elasticity for insurance from all providers could be lower. The model in this paper allows the effects of different demand elasticities to be explored in greater depth.

There are a number of approaches to modeling adverse selection in recent actuarial literature. One approach uses Markov models with an assumed high degree of adverse selection, in the sense that a small proportion of the population

\(^2\) Cutler & Reber (1998) report that Harvard University offered employees a choice between different health insurance plans with different benefits, originally with an employer contribution as a fixed percentage of the premiums. The employer contribution was then changed to the same flat contribution irrespective of which plan the employee chose. This led to a rapid migration of younger, healthier employees from the more expensive plan with better benefits to the cheaper plan with lower benefits. The more expensive plan suffered reducing enrolments and progressively increasing per capita costs, leading to its withdrawal three years after the change in employer contributions.
acquires private information (e.g., a genetic test result) indicating much higher risk, and this leads to much higher transition intensity into the insured state, or a tendency to buy much larger amounts of insurance; the effect of this adverse selection is measured by the increase in the pooled insurance price, compared with the price if the private information did not exist. In this approach, it is sufficient to use exogenous and very high assumptions for the transition intensities, because the rarity of the private information considered means that even extreme assumptions lead to negligible increases in the pooled premium (Macdonald, 1997, 1999, 2003). However, if the private information considered is more common and indicative of only moderately higher risk, extreme assumptions may be neither plausible nor sufficient. In these circumstances a second approach is to postulate utility functions for insureds, and that adverse selection arises only if lower risks achieve lower expected utility by insuring at the pooled price than by not insuring (Macdonald & Tapadar, 2007). A third approach is to model insurance demand from lower and higher-risk groups as a function of the pooled price, with a demand elasticity parameter, and investigate insurance market outcome under different demand elasticities (De Jong & Ferris, 2006). The present paper follows this third approach.

Some previous economics literature has drawn attention to the possibility of multiple equilibria in markets with adverse selection, in particular Wilson (1979, 1980). However, these papers focus on the markets for goods of varying quality – where the quality is known to sellers, but unobservable by buyers – rather than on insurance markets. Rose (1993) showed that although multiple equilibria in markets for goods of varying quality are theoretically possible, they are extremely unlikely, provided that the quality distribution of the goods follows any of a range of plausible probability distributions. The present paper provides some insight into why multiple equilibria may also be unlikely, although not impossible, in models of insurance markets.

The rest of this paper is structured as follows. Section 2 outlines the insurance market model, and characterizes the possible equilibria according to whether loss coverage is higher or lower than under full risk classification. Section 3 investigates conditions for multiple equilibria in the model, considering the relative sizes of the higher and lower-risk populations, their relative risks, and demand elasticity. Conclusions are given in Section 4.

2. THE MODEL

We assume that potential insureds can in principle be divided into two groups, which we refer to as populations. Members of population 1 are lower-risk, and members of population 2 are higher-risk. We assume that regulation requires all insurers to use community rating, that is a single common insurance price must apply irrespective of whether an individual risk belongs to the higher or lower-risk population; competition in risk classification of a “cream-skimming”
nature is not possible. We also assume that the probability of a loss is independent of insurance purchase, that is moral hazard is ignored. Equilibrium in the insurance market occurs when total premiums equal total claims, that is insurers make zero profits in equilibrium.

The demand for insurance from population $i$ at a premium of $\pi$ is specified along the lines suggested by de Jong & Ferris (2006):

$$d_i(\pi) = P_i e^{1 - (\pi/\mu_i)\lambda_i} \quad i = 1, 2$$  \hspace{1cm} (1)

where

- $P_i$ is the number of members of the population of risk class $i$ who buy insurance at an actuarially fair premium, that is when $\pi = \mu_i$
- $\mu_i$ is the risk (expectation of claim) for population $i$
- $\lambda_i$ is an elasticity parameter for insurance demand of population $i$.

The specification of $P_i$ as the number of members of population $i$ who buy insurance at an actuarially fair premium assumes that all insurance is for unit sum assured, that is every agent either buys one unit of insurance or none. This simplification is convenient for exposition, but it is not necessary: $P_i$ could alternatively be regarded as the fair-premium money demand for insurance from population $i$ (as in de Jong & Ferris, 2006).

The formula can be interpreted as follows: when $\pi$ is very small relative to $\mu_i$ demand from population $i$ will be high. As the “premium loading” $\pi/\mu_i$ increases, demand from population $i$ declines along an inverse exponential curve towards zero; this reflects the fact that if the premium loading $\pi/\mu_i$ is high enough, almost no-one would buy insurance. This specification allows a range of plausible demand curves to be specified for the relevant range of $\mu_1 < \pi \leq \mu_2$. This is illustrated in Figure 1.
The parameter $\lambda_i$ specifies the responsiveness of demand from population $i$ to changes in the premium $\pi$. The price elasticity of demand $d_i$ with respect to $\pi$ is defined as

$$\left| \frac{\pi}{d_i} \cdot \frac{\partial d_i(\pi)}{\partial \pi} \right|$$

where we have taken absolute values to obviate the usual negative sign of the elasticity. This is equivalent to

$$\left| \frac{\partial \ln[d_i(\pi)]}{\partial \ln \pi} \right| = \lambda_i(\pi / \mu_i)^{\lambda_i} \quad (2)$$

**Fair-premium demand elasticity**

Note that the parameter $\lambda_i$ represents the price elasticity of demand for insurance from population $i$ when $\pi = \mu_i$, that is when the premium is equal to the true risk for population $i$. Hence we refer to $\lambda_i$ as the “fair-premium demand elasticity”. The actual elasticity of demand from population $i$ at any other premium $\pi$ is given by Equation (2) above. It can be seen that demand is more elastic at higher premiums, and less elastic at lower premiums; this accords with the usual economic intuition.

**Specifying an equilibrium**

The total premium income from the two populations when a single rate of pooled premium $\pi$ is charged will be

$$\pi (d_1(\pi) + d_2(\pi)) \quad (3)$$

The total claims cost (total insured losses) from these policies will be

$$d_1(\pi)\mu_1 + d_2(\pi)\mu_2 \quad (4)$$

and the insurers’ expected profit (loss, if negative) when charging this pooled premium is total income less total claims, that is (3) – (4).

An equilibrium pooled premium $\pi^*$ is a value of $\pi$ for which the expected profit is zero. The existence of a premium for which expected profit is zero can be demonstrated as follows. Clearly setting $\pi = \mu_1$ will lead to negative expected profits, because at least some higher risks will buy insurance at this cheap price. Setting $\pi = \mu_2$ leads to either zero expected profits, or (provided at least some lower risks participate at this high price) strictly positive expected profits. The expected profit in our model is a continuous function of $\pi$. Thus there is at least one solution $\pi = \pi^*$ in the interval ($\mu_1$, $\mu_2$] such that expected profit is zero.
The above exposition is subject to the implicit constraint that the increased demand from higher risks at the pooled premium, $d_2(\pi)$, cannot exceed the total number of higher risks in the population. This constraint is non-binding provided the fair-premium demand $P_2$ is a sufficiently small fraction, say $\tau_2$, of the total (insured and uninsured) higher-risk population. It is easy to set up the model to include explicit “fair-premium demand fractions” $\tau_i < 1$, representing the proportions of each population who buy insurance at an actuarially fair premium. In practice we might expect $\tau_1 > \tau_2$, because for lower risks insurance is cheaper relative to the prices of other goods and services. However, changing the relative values of the $\tau_i$ is mathematically equivalent to changing the relative values of the fair-premium demands $P_i$, and hence for parsimony we omit the $\tau_i$ in this paper.

**Example**

To illustrate the use of this model, we set $P_1 = 9,000, P_2 = 1,000; \text{ and } \mu_1 = 0.01, \mu_2 = 0.04$. It seems plausible that $\lambda_1 < \lambda_2$, because for lower risks insurance is cheaper relative to the prices of other goods and services. We shall investigate the equilibrium for various pairs of values of the elasticity parameters $(\lambda_1, \lambda_2)$ for the two populations.

![Graph showing total income and claims](image_url)

**Figure 2:** Low elasticity: $(\lambda_1, \lambda_2) = (0.5, 0.9)$ giving increased loss coverage under pooling.
Figure 2 shows the equilibrium in this model for relatively inelastic demand, \((\lambda_1, \lambda_2) = (0.5, 0.9)\). Figure 2 can be interpreted as follows. The horizontal dashed line, labeled \((I = C)\), is a reference level representing total premiums (and, by assumption, total claims) if risk-differentiated premiums are charged to the two populations. All other curves represent premiums or claims when a single rate of premium \(\pi\) is charged to both populations. On the left hand side of the graph, where the single rate of premium \(\pi\) is low, demand for insurance at this price is high, and so total claims paid are high. Because of the low premium, total premiums collected are low; the market is far from equilibrium, and insurers make large losses. Insurers will therefore increase the single premium, and some customers will leave the market. As customers leave the market, total claims decrease monotonically (the downward sloping curve); but total premiums collected increase, because the increase in premium rate outweighs the number of customers leaving the market. So the curve of total premium income slopes upwards, at least initially. The intersection (shown by the arrow) of the darker curves for total premiums and total claims represents a pooling equilibrium.

More examples

Figure 3 shows the result for more elastic demand \((\lambda_1, \lambda_2) = (0.8, 1.2)\), with all other parameters as in Figure 2. Note that in this case the equilibrium is at a lower level of total income and total claims than in the risk-differentiated population. A public policymaker might regard this as a worse outcome than the result under full risk classification.

Figure 4 illustrates the result for very elastic demand, \((\lambda_1, \lambda_2) = (1.5, 2.0)\), with all other parameters as in Figure 2. The total premiums and total claims curves intersect very close to the terminal value for \(\pi\), with virtually all the lower risks out of the market. A public policymaker would probably regard this as a bad outcome from restricted risk classification.

Loss coverage

The pooling equilibrium in Figure 2 is at a higher level of total premiums and total claims than in the risk-differentiated market; that is, a 7% higher number of losses is now compensated by insurance, despite a 14% lower number of policies sold. This happens because under the assumed demand elasticities, the shift in coverage towards higher risks and away from lower risks when risk classification is restricted more than outweighs the reduction in number of policies sold. The lower number of policies sold corresponds to lower insurance demand, that is \(\sum_i d_1(\pi)\); the higher number of losses compensated by insurance corresponds to higher insurance demand weighted by risk, that is \(\sum_i d_2(\pi) \mu_i\). We refer to this ‘risk-weighted’ insurance demand as the loss coverage:

\[
\text{Loss coverage} = \sum_{i=1}^{2} d_i(\pi) \cdot \mu_i
\]
FIGURE 3: High elasticity: \((\lambda_1, \lambda_2) = (0.8, 1.2)\) giving reduced loss coverage under pooling.

FIGURE 4: Very high elasticity: \((\lambda_1, \lambda_2) = (1.5, 2.0)\), greatly reduced loss coverage under pooling.
From a public policy perspective, loss coverage may be a better metric than number of policies sold for comparing the effects of alternative risk classification schemes. This is because loss coverage focuses on the expected losses compensated by insurance (risk-weighted insurance demand), which seems a better indicator of the social efficacy or benefit of insurance to the whole population than number of policies sold (un-weighted insurance demand). The loss coverage metric implies that the public policymaker places higher priority on higher-risk individuals being insured: for example, insurance of one higher-risk individual is worth the same to the policymaker as insurance of two lower-risk individuals, if the probability of loss for each higher risk is twice that of each lower risk. In other words, insurances held by higher and lower-risk individuals are regarded as equally desirable by the policymaker ex post, when all uncertainty about who will suffer a loss has been resolved; but insurances held by higher-risk individuals are regarded as (risk-proportionately) more desirable by the policymaker ex ante.

For a public policymaker who uses the loss coverage metric, adverse selection is not necessarily adverse. Sufficiently low demand elasticities are consistent with a moderate level of adverse selection, which may reduce insurance demand but increase loss coverage (as shown in Figure 2). However, higher demand elasticities are consistent with “too much” adverse selection, which reduces both insurance demand and loss coverage (as shown in Figure 3, or in a more extreme manner in Figure 4). Under the loss coverage criterion, public policy on risk classification can be seen as a question of degree: given the demand elasticities in a particular market, what restrictions on risk classification (if any) are required to induce the optimal degree of adverse selection, which maximises the loss coverage? Applications, extensions and limitations of the loss coverage concept are discussed in greater detail in Thomas (2008).

When comparing alternative risk classification schemes, it is often convenient to define loss coverage in some normalized form. For example, later in this paper we will use the ratio of expected losses covered under community rating to expected losses covered under risk-differentiated premiums, so that loss coverage is normalized to be 1 under risk-differentiated premiums. Alternatively, loss coverage could be normalized to be 1 under compulsory insurance of the whole population.

**Generalizing for more than two risk groups**

We can generalize the above to any distribution of risks in the population, rather than just higher and lower-risk groups, as follows. Let \( g \) be a risk parameter, let \( \mu_g \) be the expected loss for a risk with parameter \( g \), let \( f_g \) be the density of the risks in the whole population, and let \( r_g \) be the demand from risks with risk parameter \( g \). The expected demand for insurance from the whole population is

\[
E[r_g] = \int r_g f_g \, dg.
\]  

(6)
The loss coverage is

$$E[r_g \mu_g] = \int r_g \mu_g f_g dg.$$  \hspace{1cm} (7)

The expected claim per contract is

$$\frac{E[r_g \mu_g]}{E[r_g]} = \frac{\int r_g \mu_g f_g dg}{\int r_g f_g dg}.$$  \hspace{1cm} (8)

and for an equilibrium in the presence of adverse selection, the pooled premium needs to be equal to this.

Graphically, an equilibrium premium is determined where the expected claims per contract crosses the 45-degree line representing the premium, as shown at the arrow in Figure 5. Note that if the actual premium increases above the equilibrium level, insurers make progressively increasing profits; if the premium decreases, insurers make progressively increasing losses. The monotonic nature of the profit function implies that the equilibrium shown is stable and well-defined.

![Figure 5: Determination of pooled premium: a well-defined single equilibrium.](image)

3. POSSIBILITY OF MULTIPLE EQUILIBRIA

To generate multiple equilibria, the expected claims per contract must cross the 45-degree single premium line more than once. Whether this is possible under suitable elasticity conditions depends on the relative sizes of the higher and lower-risk populations, and their relative risks. Figure 6 shows the multiple
equilibria resulting from the following parameters: $P_2 = 5\%$ of total population; $\mu_1 = 0.01$, $\mu_2 = 0.04$; $\lambda_1 = \lambda_2 = 1.35$. Equilibria are shown by the arrows at $\pi = 0.0166$, 0.0261 and 0.0355, and the market is very close to equilibrium for $\pi$ anywhere in the range $[0.0166, 0.0355]$. This near-equilibrium for a wide range of premiums suggests an unstable market, in two senses: (i) increases or reductions in the premium anywhere in between the upper and lower equilibria lead to only small losses or profits for insurers (that is, the profit signal from an incorrect premium remains weak even when the premium deviates considerably from an equilibrium); and (ii) at the middle equilibrium, increases in the premium from this level initially lead to small losses for insurers, and reductions in the premium initially lead to small profits for insurers (that is, the profit signal has the wrong sign). In both these senses, the market is unstable.

The plot of total income and total claims corresponding to Figure 6 is shown as Figure 7. The total income and total claims curves intersect in three places shown by the arrows along their downward slope. As the premium increases between the first and third intersections, profit changes very little, but loss coverage drops dramatically.

True multiple equilibria arise only from a limited critical range of elasticity parameter values, which may almost never apply in practice, and so do not correspond to any regularly observed real-world phenomena. However, any elasticity values higher than this critical range also lead to an undesirable equilibrium, in the sense that the premium is much higher and the loss coverage is much lower than the result under full risk classification. In this sense, the critical range of elasticity parameters associated with multiple equilibria can be thought of as a “threshold” range at or above which unsatisfactory equilibria arise. We shall see later that for some relative populations and relative
risks, true multiple equilibria can never arise; but in such cases there may still be a critical range of elasticity values which lead to “near-equilibrium” over an extended range of values for the premium. Again, any elasticity values higher than this threshold range lead to an undesirable equilibrium, with much higher premium and much lower loss coverage than under full risk classification.

Specifying the conditions for multiple equilibria

The conditions for multiple equilibria which were represented graphically in Figure 6 can be specified as follows. First note that when the premium is zero, average cost per claim exceeds the premium. When the premium is \(\mu_2\), average cost per claim must be less than or equal to the premium. Hence if the average cost per claim curve is to cross the premium line more than once, it must at some point cross the premium line from below. This gives us two necessary
and sufficient conditions which must be satisfied simultaneously to give a “middle” equilibrium, which is one of at least three possible equilibria\(^3\):

\[ \frac{\partial}{\partial \pi} \left( \frac{E[r_g \mu_g]}{E[r_g]} \right) > 1 \]  \hspace{1cm} (9)

and

\[ \frac{E[r_g \mu_g]}{E[r_g]} = \pi \]  \hspace{1cm} (10)

In words, we can interpret these conditions as follows. As the premium \( \pi \) increases, higher risks are attracted and lower risks drop out of the market; the average cost per claim increases, but generally at a lower rate than the (unit) increase in premium. For multiple equilibria to arise, the average cost per claim of the risks attracted by the pooled premium needs to increase faster than the unit increase in premium, over some interval of the feasible range for \( \pi \); and simultaneously, the average cost per claim needs to be equal to the premium \( \pi \).

Alternatively, we can think of the rate of increase of the average cost per claim as the price elasticity of the average risk attracted by the pooled premium. For multiple equilibria, this price elasticity needs to exceed 1, over some interval of the feasible range for \( \pi \); and simultaneously, the average cost per claim needs to be equal to the premium \( \pi \).

We can investigate this further using the two-populations model specified earlier. \( \lambda_1 \) and \( \lambda_2 \) were previously both regarded as free parameters, but a flexible alternative (with obvious extensions for more than two risk groups) is to specify a “base” elasticity, say \( \lambda \), for the larger, lower-risk population 1 and then set

\[ \lambda_2 = \left( \frac{\mu_2}{\mu_1} \right)^{\alpha} \lambda \]  \hspace{1cm} (11)

where \( \alpha \) is an index for the variation of the fair-premium demand elasticity as the fair premium itself changes. This specification is convenient for expository purposes because it allows us to plot equilibrium premiums and loss coverages against the single elasticity parameter \( \lambda \). There are no absolute theoretical limits on the value of \( \alpha \), but it can be seen that \( 0 \leq \alpha \leq 1 \) may be a reasonable range –

- if \( \alpha = 0 \), fair-premium demand elasticity is the same for both risk groups;
- if \( \alpha < 0 \), fair-premium demand elasticity is inversely related to the fair premium; this seems unlikely, because it implies a negative income effect from the higher-risk fair premium representing a larger part of the consumer’s total budget constraint;

\(^3\) With the form of insurance demand used in this paper, it does not appear to be possible to generate an average cost per claim curve which crosses the premium line more than three times; but more idiosyncratic demand specifications may be capable of doing so.
if \( \alpha > 1 \), fair-premium demand elasticity increases more than proportionately with the fair premium; this is possible, if the fair premium for higher risks represents a large part of their total budget constraint; but for the typical case where insurance is a small part of the consumer’s total budget constraint, it seems unlikely.

We will show figures in tables for \( \alpha = 0, \frac{1}{3} \) and \( \frac{2}{3} \). It turns out that the qualitative pattern of results is similar for all these values, and so for brevity all graphs are based on \( \alpha = 1/3 \) only. This seems a reasonable value, in that it allows for fair-premium demand elasticity in the higher-risk population to be moderately higher than in the lower-risk population, as is expected from the income effect (i.e., the higher-risk fair premium represents a larger part of the consumer’s total budget constraint). There is no specific empirical evidence or theoretical argument to support \( \alpha = \frac{1}{3} \), but the qualitatively similar pattern of results obtained from other values in the range \( 0 \leq \alpha \leq 1 \) suggests that this is not critical to our results.

We set \( P_1 : P_2 = 95:5 \), rather than 90:10 as in section 2 above\(^4\), and hereafter denote this by the phrase “5% higher-risk population fraction”. Recall that for multiple equilibria, the average cost per claim needs to cross the 45-degree premium line from below; that is, the average cost per claim for the risks attracted by the pooled premium needs to increase faster than the unit increase in the premium, over some part of the feasible range for the premium. Intuitively, this could occur either because the risk difference between the two populations is high, or because fair-premium demand elasticity is high; and the higher the risk difference, the smaller the fair-premium demand elasticity required to produce multiple equilibria. This intuition is confirmed in Table 1, which shows the critical ranges of fair-premium demand elasticities \( \lambda \) which generate multiple equilibria, for various relative risks.

What happens if the fair-premium demand elasticity is near, but not within, the critical range which generates multiple equilibria? If the elasticity is slightly higher, the equilibrium premium will be slightly higher than the highest of the multiple equilibria; and if the elasticity is slightly lower, the equilibrium premium will be slightly lower than the lowest of the multiple equilibria. In effect, the multiple equilibria correspond to a “jump” in the plot of equilibrium premium or loss coverage against the fair-premium demand elasticity \( \lambda \). This “jump” effect is illustrated in Figure 8, which shows equilibrium premium and corresponding loss coverage as a function of fair-premium demand elasticity \( \lambda \) for the case \( \mu_2 / \mu_1 = 4 \) and \( \alpha = 1/3 \) (i.e., the central cell of Table 1). The dashed horizontal line in the left panel in Figure 8 represents the population-weighted average of the risk-differentiated premiums. The dashed horizontal line in the right panel represents the loss coverage if risk-differentiated premiums are

---

\(^4\) The reason for starting our investigations with a population ratio of 95:5 rather than 90:10 as before is that we shall see later that given the reasonable constraint \( \lambda_2 \geq \lambda_1 \), a 90:10 population ratio never produces multiple equilibria (Table 4 later in the paper gives details).
charged; this is normalized to a value of 1. The gaps left in the premium and loss coverage curves for $1.33 < \lambda < 1.40$ correspond to the region where any value of $\lambda$ generates three equilibria, which are located between the upper and lower limits indicated by the ends of the curves on either side of the gaps. The three crosses show the multiple solutions generated for the specimen value $\lambda = 1.35$; an analogous triad of solutions arises for any $\lambda$ in the range $1.33 < \lambda < 1.40$.

The characteristic sigmoid pattern of the graphs shown in Figure 8, with a jump in the premium and loss coverage around the region of multiple solutions, provides a basis for distinguishing between an archetypal adverse selection spiral, with a large increase in premium and reduction in loss coverage compared to the result under risk-differentiated premiums, and other scenarios where the insurance system stabilizes after only a modest rise in premiums. If fair-premium demand elasticity is at or above the threshold range which generates multiple equilibria, we have an archetypal adverse selection spiral. But if fair-premium demand elasticity is below the threshold range, the equilibrium premium is only slightly above the population-weighted average of the risk-differentiated premiums.

From a public policy viewpoint, any equilibrium for $\lambda > 1.33$ in Figure 8 might be regarded as a bad outcome from restricted risk classification, because loss coverage is drastically reduced as compared with the result if risk-differentiated premiums are charged. For lower values of $\lambda$, say $\lambda < 1$, the reduction in loss coverage is much smaller; a public policymaker might in some cases regard this as a “price worth paying” to satisfy other policy objectives such as social solidarity. For $\lambda < 0.71$, loss coverage actually increases slightly under restricted risk classification as compared with risk-differentiated premiums. In such cases, the adverse selection resulting from restricted risk classification

<table>
<thead>
<tr>
<th>Relative risk $\frac{\mu_2}{\mu_1}$</th>
<th>Ranges for fair-premium demand elasticity $\lambda$ which generate multiple equilibria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0$</td>
</tr>
<tr>
<td>2</td>
<td>$2.76 - 3.26$</td>
</tr>
<tr>
<td>3</td>
<td>$1.70 - 1.87$</td>
</tr>
<tr>
<td>4</td>
<td>$1.33 - 1.40$</td>
</tr>
<tr>
<td>5</td>
<td>$1.13 - 1.16$</td>
</tr>
<tr>
<td>6</td>
<td>$1.002 - 1.012$</td>
</tr>
</tbody>
</table>

$^\dagger$ There are no true multiple equilibria for these combinations of $\alpha$ and relative risks. The values shown are those corresponding to “near multiple equilibria”.

TABLE 1
FAIR-PREMIUM DEMAND ELASTICITIES $\lambda$ WHICH GENERATE MULTIPLE EQUILIBRIA, FOR VARIOUS RELATIVE RISKS (5% HIGHER-RISK POPULATION FRACTION)
FIGURE 8: Equilibrium premium and loss coverage as a function of $\lambda$ for 5% higher-risk population fraction; crosses show specimen multiple solutions for $\lambda = 1.35$.

is arguably not “adverse” at all; a public policymaker might regard the increased loss coverage as a good outcome.\(^5\)

What if the higher-risk fraction of the population differs from the 5% assumed in Table 1? If the higher-risk fraction is smaller, for example a higher-risk fraction of $2\frac{1}{2}\%$ of the total population, the demand elasticity parameters which generate multiple equilibria increase, as shown in Table 2. In one sense, this makes multiple equilibria less plausible: some of the required demand

| Relative risk $\mu_2/\mu_1$ | Ranges for fair-premium demand elasticity $\lambda$ which generate multiple equilibria
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0$</td>
</tr>
<tr>
<td>2</td>
<td>$2.98 - 4.94$</td>
</tr>
<tr>
<td>3</td>
<td>$1.83 - 2.71$</td>
</tr>
<tr>
<td>4</td>
<td>$1.44 - 1.96$</td>
</tr>
<tr>
<td>5</td>
<td>$1.23 - 1.58$</td>
</tr>
<tr>
<td>6</td>
<td>$1.09 - 1.35$</td>
</tr>
</tbody>
</table>

\(^5\) The increase in loss coverage for $\lambda < 0.71$ is very small in Figure 8, but larger increases can be generated if the higher-risk population is a larger fraction of the total population, or if the difference in fair-premium demand elasticities between the two populations is larger; for example, see Figure 10 in this paper.
elasticity parameters are now very high. But in a different sense, the smaller higher-risk fraction makes it easier to generate multiple equilibria: as shown in Table 2, multiple equilibria are generated by $\lambda$ in wider ranges than in Table 1.

Figure 9 shows equilibrium premium and corresponding loss coverage as a function of fair-premium demand elasticity for the case $\mu_2/\mu_1 = 4$ and $\alpha = 1/3$ (ie the central cell of Table 2). The heavy dashed horizontal lines represent the population-weighted average of the risk-differentiated premiums (in the left panel), and the loss coverage if risk-differentiated premiums are charged (in the right panel). The gaps left in the plots for $1.42 < \lambda < 1.85$ correspond to the region where any value of $\lambda$ generates three equilibria, which are located between the between the upper and lower limits indicated by the ends of the curves on either side of the gap. The three crosses show the multiple solutions generated by $\lambda = 1.70$; an analogous triad of solutions arises for any $\lambda$ in the range $1.42 < \lambda < 1.85$. It can be seen that for any $\lambda \leq 1.42$, the equilibrium pooled premium is only very slightly higher than the population-weighted average of the risk-differentiated premiums (the horizontal dashed line at $p = 0.01075$). Loss coverage is generally slightly lower than under risk-differentiated premiums, unless $\lambda < 0.73$, where it becomes very slightly higher. For any $\lambda \geq 1.85$, the equilibrium pooled premium is far above the population-weighted average premium, and loss coverage is drastically reduced.

What happens if the higher-risk fraction of the population is more than 5%? If the higher-risk fraction is say 10%, then to generate multiple equilibria, $\lambda_2$ substantially less than $\lambda_1$ is required, as shown in Table 3. However $\lambda_2$ substantially less than $\lambda_1$ is often rather implausible. The higher price of insurance for higher risks, relative to the price of other goods and services, lead us to expect that $\lambda_2$ would generally be higher than $\lambda_1$. Given the reasonable constraint $\lambda_2 \geq \lambda_1$, we can specify a critical fraction for the higher-risk population fraction; crosses show specimen multiple solutions for $\lambda = 1.70$. 

![Figure 9: Equilibrium premium and loss coverage as a function of $\lambda$ for 2.4% higher-risk population fraction](image-url)
which ensures that no multiple equilibria arise. This critical fraction is shown for various population relative risks in Table 4.

When the higher-risk population fraction only modestly exceeds the fraction required to ensure no multiple equilibria, there is still a region of “near multiple equilibria”, where the equilibrium premium and the corresponding loss coverage change rapidly with changes in $\lambda$. This effect is illustrated in the upper panels of Figure 10, which show the plots for a 10% higher-risk population fraction, with $\mu_2/\mu_1 = 4$ and $\alpha = 1/3$. The region of “near multiple equilibria” is centred around $\lambda \approx 1.23$; it can be seen that the equilibrium premium and loss coverage plots against $\lambda$ display sigmoid and reverse sigmoid patterns respectively, with their highest rates of change centred around $\lambda = 1.23$. As the higher-risk population fraction is increased further, the sigmoid and reverse sigmoid patterns gradually flatten out. This effect is illustrated in the lower panels in Figure 10, which show the plots for a 20% higher-risk population fraction, still with $\mu_2/\mu_1 = 4$ and $\alpha = 1/3$. The less skewed population (80:20 instead of 90:10) leads to a less steep “jump” between the low-premium and high-premium regions.

**TABLE 3**

APPROXIMATE FAIR-PREMIUM DEMAND ELASTICITIES $\lambda_1$ AND $\lambda_2$ REQUIRED TO PRODUCE MULTIPLE EQUILIBRIA, FOR VARIOUS RELATIVE RISKS (10% HIGHER-RISK POPULATION FRACTION)

<table>
<thead>
<tr>
<th>Relative risk $\mu_2/\mu_1$</th>
<th>Approximate fair-premium demand elasticity $\lambda_1$ required to produce multiple equilibria</th>
<th>$\lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\approx 2.60$</td>
<td>&lt;1.44</td>
</tr>
<tr>
<td>3</td>
<td>$\approx 1.60$</td>
<td>&lt;0.52</td>
</tr>
<tr>
<td>4</td>
<td>$\approx 1.23$</td>
<td>&lt;0.28</td>
</tr>
<tr>
<td>5</td>
<td>$\approx 1.05$</td>
<td>&lt;0.11</td>
</tr>
<tr>
<td>$\geq 6$</td>
<td>Cannot generate multiple equilibria</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4**

THRESHOLD HIGHER RISK POPULATION (AS FRACTION OF TOTAL POPULATION) REQUIRED TO ENSURE NO MULTIPLE EQUILIBRIA, GIVEN $\lambda_2 \geq \lambda_1$

<table>
<thead>
<tr>
<th>Relative risk $\mu_2/\mu_1$</th>
<th>Threshold higher risk population $P_2$ (as fraction of total population) required to ensure no multiple equilibria, given $\lambda_2 \geq \lambda_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$&gt; 9.0%$</td>
</tr>
<tr>
<td>3</td>
<td>$&gt; 7.3%$</td>
</tr>
<tr>
<td>4</td>
<td>$&gt; 6.2%$</td>
</tr>
<tr>
<td>5</td>
<td>$&gt; 5.6%$</td>
</tr>
<tr>
<td>6</td>
<td>$&gt; 5.1%$</td>
</tr>
</tbody>
</table>
To summarize the numerical results in Tables 1 to 4 and Figure 10, we make the following observations.

- Variations in the parameter $\alpha$ have only a small effect on the demand elasticity parameter required to generate multiple equilibria. This can be understood as follows. The parameter $\alpha$ indexes the demand elasticity “premium” or “margin” for the higher-risk population over that of the lower-risk population. The much larger size of the lower-risk population means that at most equilibria, most of the income and claims relate to the lower-risk population. Hence the results are driven mainly by the “base” demand elasticity parameter, with only a small effect from the demand elasticity “premium” or “margin” for the higher-risk population.
As relative risk \( \frac{\mu_2}{\mu_1} \) increases, the demand elasticity parameter required to generate multiple equilibria decreases. This can be understood as follows. Recall that multiple equilibria require the average cost per claim for risks attracted by the pooled premium to be increasing faster over some range than the unit increase in the premium. This can happen either because the demand response to the premium increase is high, or because the difference in the risks is high.

As relative risk \( \frac{\mu_2}{\mu_1} \) increases, the range of demand elasticity parameters which generate multiple equilibria becomes narrower. This can be understood by reference to Figure 6, as follows. As the demand elasticity parameter increases through the critical range, the scenario of three intersections between average claim costs and premium shown in Figure 6 moves towards a single intersection at a premium above the highest of the three intersections. For the multiple equilibria to disappear, the average cost per claim curve in Figure 6 must move upwards sufficiently so as to eliminate its intersections with the 45-degree premium line at lower premiums. If the relative risk \( \frac{\mu_2}{\mu_1} \) is low, the average cost per claim changes only slowly as the elasticity parameter (and hence insurance demand) changes; and hence the multiple intersections in Figure 6 persist for a wider range of demand elasticity parameters.

As the higher-risk fraction of the population increases, the demand elasticity parameter required to generate multiple equilibria decreases. However, if the higher-risk fraction of the population exceeds a threshold, multiple equilibria cannot arise, given the reasonable constraint \( \lambda_2 \geq \lambda_1 \). This can be understood as follows. With a larger higher-risk fraction, the higher-risk population makes a material contribution to the average cost per claim over the full range \( [\mu_1, \mu_2] \) of feasible premium levels. As \( \pi \) is increased, lower risks have a higher propensity to leave the market than higher risks; but the material weighting of the higher risks, at all premium levels, means that the higher rate of exit of lower risks never raises the rate of change of the average cost per claim above 1, with the average cost per claim also simultaneously equal to the premium. In effect, the material weighting of higher risks in the average cost per claim over the full range \( [\mu_1, \mu_2] \) of feasible premium levels acts as a “drag” on the rate of change of the average cost per claim. This “drag” on the rate of change of the average cost per claim from the weighting of higher risks also increases with the relative risk \( \frac{\mu_2}{\mu_1} \). The latter point explains the last line in Table 3: for high relative risk, \( \frac{\mu_2}{\mu_1} \geq 6 \), even very low \( \lambda_2 \), much less than \( \lambda_1 \), cannot generate multiple equilibria.\(^6\)

---

\(^6\) For further insight into this “drag” from a larger higher-risk population fraction, consider a case where the skew is reversed, so that the higher-risk population forms the larger part of the total population, say \( P_2 = 90\% \). In this case, it is obvious that over the full range \( [\mu_1, \mu_2] \) of feasible premium levels, the lower risks will always contribute very little to the average cost per claim. The average cost per claim will always be close to the risk level for the higher-risk population, and will change only slightly as \( \pi \) changes.
Conditions which preclude multiple equilibria irrespective of relative populations and risks

It is also possible to specify broad regions of the parameter space for \((\lambda_1, \lambda_2)\) where multiple equilibria can never arise, irrespective of the higher-risk population fraction and the relative risk. First note that the total claims curve in Figures 2, 3, 4 and 7 must always slope downwards over the full feasible range \((\mu_1, \mu_2)\). Multiple equilibria arise if the total income curve intersects the total claims curve more than once in the range \((\mu_1, \mu_2)\). This cannot happen if the total income curve slopes upwards over the full range \((\mu_1, \mu_2)\). The total income curve is the sum of the income from populations 1 and 2. So if the total income curve for each population slopes upwards over the full range – that is, there is no maximum in the premium income curve below \(\mu_2\) – then multiple equilibria cannot occur.

The condition for a maximum in total premium income from population \(i\), say \(\Phi_i\), is

\[
\frac{\partial (\pi \cdot d_i(\pi))}{\partial \pi} = 0 \Rightarrow \lambda_i (\pi / \mu_i)^{\lambda_i} = 1
\]  

(12)

For population 1, note that \(\pi / \mu_1 < \mu_2 / \mu_1\) and hence

\[
\lambda_1 (\mu_2 / \mu_1)^{\lambda_1} \leq 1
\]

(13)

is a sufficient condition for the absence of a maximum in the premium income curve \(\Phi_1\) below \(\mu_2\). But \(\mu_2 / \mu_1 > 1\), and hence the condition (13) for absence of a maximum in \(\Phi_1\) below \(\mu_2\) requires \(\lambda_1 \leq 1\). For any given relative risk \((\mu_2 / \mu_1)\), the maximum value of \(\lambda_1\) which admits a maximum in \(\Phi_1\) below \(\mu_2\) can then be found as the unique solution for \(\lambda_1\) of

\[
\left( \frac{\mu_1}{\mu_2} \right) = \lambda_1 (\mu_2 / \mu_1)^{\lambda_1}
\]

(14)

Moving on to population 2, and applying equation (12) above, note that \(\pi / \mu_2 \leq 1\). Hence \(\lambda_2 < 1\) is a sufficient condition for the absence of a maximum in \(\Phi_2\) below \(\mu_2\).

The above conditions lead to Table 5, which shows for a range of relative risks \((\mu_2 / \mu_1)\) the regions of the \((\lambda_1, \lambda_2)\) parameter space in which multiple equilibria can never occur, irrespective of the relative sizes and fair-premium take-ups of the two populations.

The conditions in Table 5 are sufficient to ensure that multiple equilibria cannot occur, irrespective of the higher-risk population fraction; but they are not necessary conditions. They can be thought of as conditions to give a graph of total premiums and total claims of the type characterized by Figure 2: the total income curve slopes upwards over the entire range \((\mu_1, \mu_2)\). But even if the total income curve slopes downwards towards the upper end of the range
(\(\mu_1, \mu_2\)], it may intersect the total claims curve only once: that is, a graph of the type characterized by Figure 3 or Figure 4. Therefore the \(\lambda_i\) can be higher than the values in Table 5 and still be “safe” in the sense of not produce multiple equilibria. Multiple equilibria arise only when the graph is of the type characterized by Figure 7.

Comparison with empirical demand elasticities

The fair-premium demand elasticity parameters in Tables 1 to 5 can be compared with empirical estimates of demand elasticity for various classes of insurance. We defined demand elasticity using absolute values for convenience in this paper, but the estimates in empirical papers are generally given with the negative sign, and so we quote them in that form. For example, for yearly renewable term insurance in the US, an estimate of –0.4 to –0.5 has been reported (Pauly et al, 2003). A questionnaire survey about life insurance purchasing decisions produced an estimate of –0.66 (Viswanathan et al, 2007). For private health insurance in the US, several studies estimate demand elasticities in the range of 0 to –0.2 (Chernew et al., 1997; Blumberg et al., 2001; Buchmueller and Ohri, 2006). For private health insurance in Australia, Butler (1999) estimates demand elasticities in the range –0.36 to –0.50 (higher than in the US, perhaps because Australia’s universal Medicare is a better substitute for private insurance than is available to most people in the US). These magnitudes are significantly lower than the threshold values of fair-premium demand elasticity required to generate multiple equilibria (or above the threshold, low-coverage equilibria) in our model. Demand elasticity for some types of insurance might be higher, particularly where there is a good substitute for the insurance. However, actual demand elasticity magnitudes significantly lower than the threshold values which lead to multiple equilibria (or above the threshold, low-coverage equilibria) in our model may explain why some insurance markets appear to
operate reasonably well under community rating, without the collapse in coverage which insurance folklore suggests.

The remarks in the preceding paragraph are made in the context that all insurers are required to operate the same risk classification regime, as was assumed when setting up the model in section 2 of the paper. The remarks will probably not apply where different insurers are permitted to compete by offering different risk classification regimes. In this scenario, where different insurers’ products are close substitutes offered on different underwriting terms, it seems plausible that the demand elasticity for insurance from a particular insurer might be much higher than demand elasticity for insurance from any insurer. The “adverse selection spiral” may then be a good description of the fate of one insurer which does not classify risks adequately in a setting where others insurers do. But this does not necessarily mean that it is a good description of the fate of a market where no insurers classify risk.

4. CONCLUSION

This paper has used a simple model with two populations, one higher-risk and one lower-risk, to investigate insurance market outcomes when insurers are not permitted to differentiate premiums by risk level. Market outcomes were characterized by the equilibrium pooled premium when risk classification was restricted, and the corresponding risk-weighted insurance demand or “loss coverage.” It was suggested that from a public policy perspective, loss coverage (risk-weighted insurance demand) might be a better metric from a public policy perspective than number of policies sold (un-weighted insurance demand), because loss coverage focuses on the expected losses actually compensated by insurance.

For the model in this paper, the main results are as follows. Insurance market outcome in the absence of risk classification can be related to a parameter for the elasticity of demand for insurance at an actuarially fair premium, that is the “fair-premium demand elasticity.” High fair-premium demand elasticity leads to an equilibrium premium close to the risk of the higher-risk population, and a much lower loss coverage than under risk-differentiated premiums. This can be thought of as the conclusion of an archetypal adverse selection spiral. From a public policy perspective, it might be considered a bad outcome from community rating. But for sufficiently low fair-premium demand elasticity, this collapse in coverage does not occur; instead, the market stabilizes with a premium only slightly higher than the population-weighted average of risk-differentiated premiums, and relatively high loss coverage – possibly higher than under more complete risk classification. From a public policy perspective, this might be considered a good outcome from community rating.

Footnote 2 summarises an example, as reported in Cutler & Reber (1998).
The ranges for the demand elasticity parameter characterized above as “high” and “low” are separated by a threshold range for the parameter which leads to an unstable market outcome – either multiple equilibria (that is, more than one pooled premium is capable of equilibrating insurers’ revenue and claims), or near-equilibria for an extended range of premium values. Generally, the threshold range for the fair-premium demand elasticity parameter at or above which an unsatisfactory outcome arises in the model in this paper appears higher than the demand elasticity values which have been estimated in empirical studies of a number of insurance markets. The high demand elasticity which is required to generate an unsatisfactory outcome offers a possible explanation of why some insurance markets appear to operate reasonably well under community rating, without the collapse in coverage which insurance folklore suggests. Finally, it is emphasized that these conclusions are based on a specific model, which may require some generalization before the results are widely applied.

REFERENCES


R. GUY THOMAS

*School of Mathematics, Statistics & Actuarial Science*
*University of Kent*
*Canterbury CT2 7NF*
*United Kingdom*
*E-mail: r.g.thomas@kent.ac.uk*