Modeling Individual and Small Group Health Insurance

Andrew D. Wei, FSA, MAAA

Modeling Individual and Small Group Health Insurance

Abstract

In this paper, we develop a formal framework to model the behavior of individual and small group health insurance policies in a competitive market. In this framework, anti-selective lapse is not assumed, but arises as a result of rational consumers seeking the best value in a competitive market where the premium rates are restricted by law. We obtain a model that addresses a number of fundamental questions which have not been fully answered previously, including the effect of price-induced anti-selective lapse on the mix of insureds in a block and the optimal renewal price to maximize the aggregate profit of a block. We apply the model to examine the diverse aggregate behavior (e.g., aggregate lapse, loss ratio, and profit) of different blocks in various scenarios, and obtain an illustration of "assessment spiral" – a large premium rate hike can cause the loss ratio of a block to increase. We also apply the model to demonstrate the impact of major market influences (competitor price level, trend as well as underwriting cycles) and the impact of an insurer's internal drivers (risk selection, firm-specific cost, and differentiation). One significant finding, with possible implications for product design and regulations, is that for a large class of blocks, the insurer can maximize aggregate profit while keeping the rate increase moderate and the lapse rate low, and as a result, these blocks are inherently stable and sustainable.

JEL Classification codes: D11, G22, I11

Keywords: Anti-selection, Lapse, Model, Health, Insurance

1. Introduction

Individual and small group health insurance (ISH), two major forms of health insurance in the market, share several distinguishing characteristics. Both are characterized by annual term contracts as well as small policy size[1]. The ISH contract also contains two key provisions by law: 1) A policy must be renewable at the insured's discretion, and 2) strict limits are placed on how much the premium rates can vary within the same class[2]. These provisions are intended to mitigate, due to the small size, the renewal risk or the inability of the insured (or the group) to renew the contract for another term (at an affordable rate) when some members are expected to incur large claims. But this regulatory remedy is not, with major implications on ISH behavior, a full protection because its renewal guarantees are dependent on the particular experience of an insurer[3]. See overviews of ISH insurance and rating laws in O'Grady(1988) and Bluhm(2003), and major issues of the markets in Blumberg (1996).

1.1 A**ggregate Behavior**

In managing ISH insurance, we are often interested in the aggregate behavior of policies in a block. In fact, a major actuarial responsibility is to determine the aggregate behavior (e.g., lapse, loss ratio, and profit) of an ISH block in different scenarios. There are several reasons for our interest in the aggregate behavior of a block: First, a block of policies constitute a practical base unit in which the insurance risks are pooled across individual policies and evaluated. In addition, major decisions concerning product features, underwriting, and pricing are often made

[1] The small group is usually defined as an employer group with less than 50 employees.
[2] The premium rates are subject to rating band for small group insurance and cannot be increased for medical reasons for individual medical insurance.
[3] If a block is poorly managed, the renewal and rate limit guarantee may not be adequate. The insurer can choose to cancel the entire block.

at the block level. Important macro factors such as medical cost trend and competitive price level in the market also tend to affect all policies in a block. Finally, perhaps most significant, the ISH laws and regulations, especially the way in which premium rate restrictions are formulated, create an intrinsic interdependence among the insureds[4] at the block level. A good quantitative framework to understand the aggregate behavior of ISH blocks is essential to managing ISH successfully. Lacking an accurate understanding of those behavior could lead to failures of the ISH programs as described by Bolnick (1983).

Unfortunately, modeling the aggregate behavior of a block of policies can be difficult, and traditional actuarial models are not adequate for such a task. Part of the difficulty lies in the fact that an ISH block consists of insureds of heterogeneous types. An important phenomenon, known as antiselective lapse (ASL), is that the healthy lives tend to lapse at a higher rate than the impaired lives. As a result, the mix of insureds of different health status in a block does not remain static overtime. Furthermore, the change in the mix of insureds from ASL is accelerated in response to changes such as premium rate increases. One key element missing in the traditional actuarial models is an adequate mechanism for modeling the change in the mix of insureds in a block under the influences of different factors.

1.2 Goals

Our main goal is to model the aggregate behavior of a block of ISH policies and the impact of various factors. In doing so, we wish to address several specific problems that do not yet have a satisfactory solution: 1) describe in quantitative terms how ASL arises in ISH insurance, and establish the effect of rate restrictions on ASL; 2) determine the change in the mix of insureds in

---

[4] The term "insured" will be used to refer to one of several entities: 1) an individual who is a sole member of a policy, 2) a family covered under a single policy, or 3) a small group covered under a single policy.

a block in response to influences of factors such as premium rate increase; 3) incorporate the impact of competitors' prices in the market; and 4) find the optimal pricing solution that maximizes the aggregate profit of a block in different scenarios.

1.3 Approach

We take a "micro" approach. We begin by establishing a model of individual ISH insured behavior. A basic assumption is that the individual insured is a rational seeker of the best value in the market. We define "market price" as a weighted average of competitor prices. We show that "market price" can also be determined as the price at which the individual insured's lapse probability is exactly one-half. We formalize how an insurer sets premium rate based on the expected cost as well as competitive strategy, subject to rate restrictions. A key notion essential to understand ASL behavior is "excess risk", which represents the portion of the insurance risk not reflected in the premium rate due to rate restrictions. We obtain formal relationships for individual insured behavior in the competitive market. We show that ASL arises naturally as a result of the insured's rational choice when the premium rates are restricted.

Next we derive the model for the aggregate behavior of ISH policies in a block. A key step is to obtain a realistic yet tractable formal characterization for the mix of insured types in a block. We accomplish this by constructing a one-parameter type distribution based on insureds' "excess risk". In fact, for a wide range of applications, we assume a Pareto distribution of excess risk. We derive the model for the aggregate behavior of a block by summing the individual behavior established earlier over all insureds in the block. Under further simplifying assumptions, we can actually obtain explicit analytic formulas for the model of the aggregate behavior of a block.

We apply the model to the behavior of different ISH blocks and to answer several fundamental questions. First, we examine the change in the mix of insureds from ASL and the diverse aggregate behavior (e.g., lapse, loss ratio, and profit) of different ISH blocks. In particular, we obtain an illustration of "assessment spiral" – a large rate increase actually results in a higher loss ratio in a block As a key application, we obtain the optimal pricing solutions for various ISH blocks. We find that for a large class of blocks, maximizing profit can be achieved with minimal lapses, and these blocks are stable and sustainable when managed properly. This could have implications for both product design and rate regulations. We also examine the impact of market price level and market cost level. Finally, we address the impact of underwriting cycle, and the impact of several internal drivers related to an insurer's risk selection, firm-specific cost, and differentiation. We demonstrate the fact that in an underwriting down cycle, the insurer has no choice but to accept the lower profit.

1.4 Existing Models

A major difference between this model and the existing actuarial models lies in how ASL is represented within the model. In traditional actuarial models, ASL is not explicitly modeled, but the effect of ASL is a model input represented by fixed durational factors[5]. See overviews of the main actuarial models in O'Grady (1988) and Bluhm (2003). With explicit assumptions for differential lapse by health status, Bluhm (1983) established how ASL can lead to durational increases in loss ratio in a block. In this model, we take a step further by establishing how ASL arises in an ISH block, and incorporating the effect of premium rate, market factors, and rate restrictions. ASL is not assumed, but emerges endogeneously in the model.

---

[5] To be accurate, duration factors represent both the effect of ASL and that of underwriting wear-off.

It is instructive to contrast ASL with adverse selection, which is often cited to explain ISH behavior in the competitive market. The central assumption of the theory is asymmetric information, i.e., the insured knows more about the future claims than the insurer, and as a result, could select against the insurer. See Belli (2001) for an overview of adverse selection in health insurance markets. In this model, the assumption of asymmetric information is not required, and, as a matter of fact, ASL can arise in a market with perfect information.

The aggregate behavior of ISH policies presents, to my best knowledge, a distinct type of behavior that has not been considered previously within a formal framework. In particular, the notion of excess risk with its central role in this model has no real counterpart in the current models from related micro-economics, financial, and health economics fields. See Mas-Colell (1995), Campbell (1996), Panjer (1998), and Folland (2004) for standard texts in those fields. Furthermore, optimal pricing of ISH, due to excess risk, differs from that considered by the existing pricing models. See Wilson (1996), Pashigian (1995), Nagle (1995), and Doland (1996) for current pricing theory and methodology for various applications.

Organization of this paper is as follows: Section 2 introduces the basic assumptions and establishes the model of individual insured behavior in the competitive market. Section 3 derives the model of the aggregate behavior of an ISH block. In Section 4, we apply the model to the aggregate behavior of different ISH blocks and the impact of various factors, and obtain the optimal pricing solutions. Section 5 concludes the paper with discussion and considerations for future research.

2. The Framework

In this section, we establish the model of the individual insured behavior in a competitive market. We first consider the individual lapse behavior and insured choice in a competitive market, and define “market price” as an index of competitors’ prices. We then formalize how the insurer sets premium rates based on both cost and competitive strategy, subject to rate restrictions. We introduce the notion of “excess risk” which represents the impact of rate restrictions on the premium rate. We obtain key relationship about the individual insured behavior, including a formula for ASL at the individual level.

2.1 Lapse Response to Price

We distinguish two main types of lapses: base lapse and price-induced lapse. In the former case, the cause for the insured’s lapses is unrelated to price: examples include change in employment or enrollment in Medicare. In the later case, the insured, in response to a price change, may switch to another insurer, or leave the ISH insurance category all together. In this model, we consider, for simplicity, only lapses that are results of insureds switching insurers in response to a price change.

Let $L(p)$ denote the probability of lapse for the insured x when the log of price [6] is p. We assume that $L(p)$ is an increasing function of p of the form

$$L(p) = S_a(p - p^*(x)) \tag{2.1}$$

where $S_a(z)$ is an *s-function*[7] satisfying $S_a(z) \to 1$ as $z \to \infty$, $S_a(z) \to 0$ as $z \to -\infty$, and $S_a(0) = 1/2$, and $p^*(x)$ is a parameter we shall refer as *half lapse price* because it corresponds to the price at which the probability of lapse for x is exactly one-half.

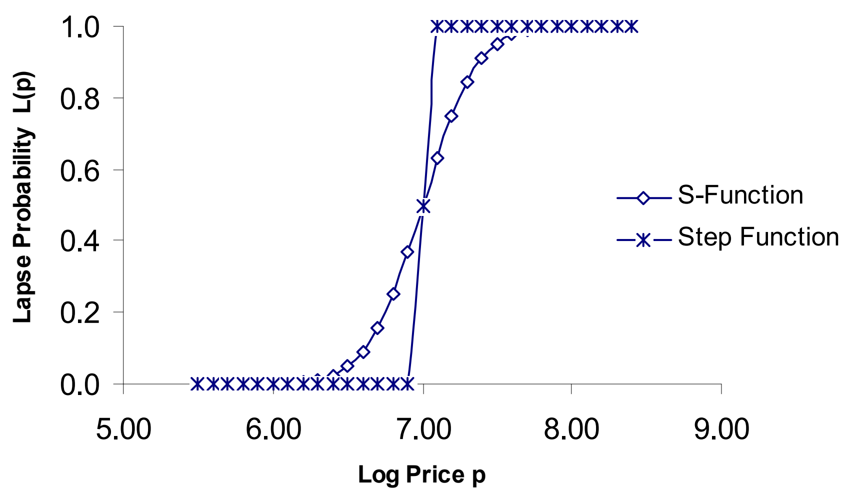
An important special case of the s-function is a *step function* given by

[6] The log form is preferred here because it converts multiplicative relationships to additive ones; its range (-∞, ∞) is better suited for linear function.
[7] Examples of S function include: cumulative normal and logistic function.

$$S_0(z) = \begin{cases} 0 & if\ z < 0 \\ \frac{1}{2} & if\ z = 0 \\ 1 & if\ z > 0 \end{cases} \tag{2.2}$$

Figure 1 shows an s-function $L(p) = S_a(p - p^*)$ and a step function $L(p) = S_0(p - p^*)$ for $p^* = 7$.

Figure 1
Lapse Functions



2.2 Insured's Choice in the Market

Consider a market that consists of insurer $j = 1, 2, ..., N$, where N is an integer. We assume that each insurer j sells only a single product, also denoted by j [8], for price p_j. Due to the differences in perceived quality and associated transaction cost with switching, the insured x is willing to pay an extra price or a *differentiation premium* α_j for product j (α_j maybe negative). The adjusted price $p'_j = p_j - \alpha_j$ is called the *equivalent price* for product j.

---

[8] For notational simplicity, we use i to denote either the insurer or the product that the insurer sells.

An insured x chooses between a pair of products i and j based on considerations of the differences in price, as well as quality and transaction cost. Formally, the probability of the insured choosing j over i, denoted by $Z_{i,j}$, is assumed to be a function of the form:

$$Z_{i,j} = S_b(p'_i - p'_j) \tag{2.3}$$

where S_b is also an s-function satisfying $S_b(z) \to 1$ as $z \to \infty$, $S_b(z) \to 0$ as $z \to -\infty$, and $S_b(0) = 1/2$. There is a natural interpretation for (2.3): If $p'_i < p'_j, Z < 1/2$, the insured is less likely to choose j over i. If $p'_i > p'_j, Z > 1/2$, the insured is more likely to choose j over i. If $p'_i = p'_j$, $Z = 1/2$, the insured is indifferent between j and i. Let i denote the current insurer for insured x. One of the N differentiation premiums can be arbitrarily chosen. This means that for any weights w_j with $\sum_{j \neq i} w_j = 1$, we choose α_i in such a way that the condition $\sum_{j \neq i} w_j \cdot \alpha_j = 0$ is satisfied. We shall define weights w_j later.

2.3 Lapse Response and Choice

Now we can express $L(p_i)$, the lapse probability for insured x from insurer i, as the weighted sum of the probabilities of insured x choosing insurer j over current insurer i:

$$L(p_i) = \sum_{j \neq i} w_j \cdot S_b(p'_i - p'_j) \tag{2.4}$$

where

w_j = the probability of the insured x would consider insurer $j \neq i$ as an alternative to the current insurer i[9]. (2.5)

By definition, we have $\sum_{j \neq i} w_j = 1$.

Remark: The above Equation (2.4) relates the lapse behavior of an insured to the choice behavior of the insured in the market.

2.4 Market Price

We define *market price* $m_i(x)$ of a policy for insured x from insurer i as an index of competitive prices given by:

$$m_i(x) = \sum_{j \neq i} w_j \cdot p_j(x) \tag{2.6}$$

where w_j is the weight defined in (2.5).

A distribution $f(z)$ is said to be *balanced* if the mean is equal to the median or $z^* = \bar{z}$. Examples of balanced function include symmetric functions. From now on we shall assume that the first derivatives of the s-functions are balanced and continuously differentiable.

Now we relate lapse response to market price in the following proposition:

Proposition 1:

Assume that first derivatives of the s-functions S'_a in (2.1) and S'_b in (2.3) are balanced and continuously differentiable, then we can show that the market price is equal to the half lapse price minus the differentiation premium

$$m_i(x) = p_i^*(x) - \alpha_i . \tag{2.7}$$

It follows that we can express the lapse response function as

$$L(p_i) = S_a(p_i - m_i(x) - \alpha_i) \tag{2.8}$$

[9] Weight w_j is generally dependent on p_j for $j \neq i$, but not on p_i.

In the following the subscript i is sometimes dropped from the notations. We will use the convention of not spelling out subscripts when there is no confusion.

Proof:

The proof is given in Appendix A1.

Proposition 1 implies that market price *m* can be estimated by observing the insured's lapse responses as the price increases.

2.5 Decomposition of Cost

The cost of the policy for insured x is the sum of claim cost and expenses associated with the insured.

Consider a block consisted of multiple *rating classes*[10]. Let *g* denote a rating class and $x_0{}^g$ denote a standard risk or a healthy insured in the rating class *g*. One of the rating classes in the block may be designated as the *base rating class,* or a *reference rating class,* denoted by g_0.

The log cost *c(x)* associated with a policy for insured x in one period can be decomposed as the sum of three distinct components

$$c_i(x) = \hat{c}_i + c_i^g + c_i^h \tag{2.9}$$

where

$\hat{c}_i = c_i(x_0{}^{g_0})$ is referred to as *base cost* or *cost level,*

$c_i^g = c_i(x_0{}^g) - c_i(x_0{}^{g_0})$ is referred to as *rating class cost factor,*

$c_i^h = c_i(x) - c_i(x_0{}^g)$ is referred to as *relative risk cost factor.*

Define *market cost level* $\tilde{c}$ as an index of the base costs of all competing insurers in the market given by:

$$\tilde{c} = \sum_{j \neq i} w_j \cdot \hat{c}_j \tag{2.10}$$

where w_j is the weight defined in (2.5).

Then $\hat{c}_i$ can be further decomposed into two components:

$$\hat{c}_i = \tilde{c} + \beta_i \tag{2.11}$$

where β_i represents the firm-specific cost component referred as *firm cost factor*. We have $\sum_{j \neq i} w_j \cdot \beta_j = 0$.

2.6 Components of Premium

Now we decompose the premium rate. Using the previous notations, the log premium rate *p(x)* of a policy x in the rating class *g* over a single period can be decomposed in a similar fashion:

$$\begin{aligned} p_i(x) &= \hat{p}_i + p_i^{\,g} + p_i^{\,u} \\ &= \hat{p}_i + p_i^{\,g} + p_i^{\,h} - p_i^{\,d} \end{aligned} \tag{2.12}$$

where

$\hat{p}_i = p_i^{x_0^{g0}}$ is referred to as *base price, base rate, or price level*,

$p_i^{\,g} = p_i^{x_0^{g}} - p_i^{x_0^{g0}}$ is referred to as *rating class premium factor*,

$p_i^{\,u} = p_i(x) - p_i^{x_0^{g}}$ is referred to as *relative risk load,*

$p_i^{\,d} \geq 0$ is referred to as *rate reduction,*

$p_i^{\,h} = p_i^{\,u} + p_i^{\,d}$ is referred to as *relative risk premium factor*.

---

[10] A rate class consists of policies that have the same demographic rating variables such as age, gender, and benefit variables but may differ in health status.

Remarks

Rate reduction p_i^d in general should represent both the effect of rate restrictions and the effect of current insurer i's underwriting precision relative to those of the competitors in the market[11]. In an ISH block, the effect of rate restrictions is generally far greater than that of underwriting precision. For simplicity, in the following we shall ignore the latter effect and assume that p_i^d represents only the former effect. But keep in mind that it is not difficult to generalize the model so that p_i^d can represent both effects of rate restrictions and underwriting precision.

Examples of Rate Restrictions

For small group, the rating corridor with 25% upper and lower limits can be written as $\log(75\%) < p^u < \log(125\%)$. For individual medical, the fact that the initial rating class can not be changed for health status reasons can be expressed as $p^u = p^{u_0} \geq 0$, where p^{u_0} is the initial risk load at time of issue.

2.7 Setting Premium Rate

We want to formalize the setting of the premium rate by the insurer. The components of premium rate of a policy x with insurer i are determined based on the equations below:

$$p_i^g(x) = c_i^g(x) \tag{2.13}$$

$$p_i^h(x) = c_i^h(x) \tag{2.14}$$

$$p_i^u(x) = p_i^h(x) - p_i^d(x) \tag{2.15}$$

[11] It is possible to show that inferior underwriting relative to the market raises the value of p_i^d and superior underwriting relative to the market lowers the value of p_i^d.

$$\hat{p}_i = \hat{c}_i + \phi_i \tag{2.16}$$

where ϕ_i represents deviation of the base price from the base cost.

Remarks

Intuitively, the first three equations state that the *rating class premium factor* and the *relative risk load* are set to match the corresponding cost components, with further adjustment for the effect of rate restrictions or the underwriting accuracy.

The last equation states that the base price can deviate from the base cost. This deviation is by definition unrelated to rate restriction. The sources of this deviation ϕ_i are two-fold: 1) the insurer needs to set the premium rate in advance, based on forecasted base cost for the future which can not be determined precisely, and 2) the insurer may wish to adopt a pricing strategy that deviates from the cost-plus pricing method.

Substituting equations(2.13), (2.14), (2.15) and (2.16) into (2.12), we obtain a single equation that relates premium rate to cost:

$$p_i(x) = c_i(x) - p_i^d(x) + \phi_i \,. \tag{2.17}$$

Substituting $\phi_i = \hat{p}_i - \hat{c}_i$ from (2.16) into (2.17) and rearranging the equation, we have another expression for the rate reduction

$$p_i^d(x) = (c_i(x) - \hat{c}_i) - (p_i(x) - \hat{p}_i) = \Delta c_i(x) - \Delta p_i(x) \tag{2.18}$$

where $\Delta c_i(x) = c_i(x) - \hat{c}_i$ and $\Delta p_i(x) = p_i(x) - \hat{p}_i$.

2.8 Market Rate Reduction

Consider a market with guaranteed issue laws. In such a market, the insurer cannot fully rate many high risk insureds to whom the insurer must issue a policy. This essentially imposes a rate restriction on the new business rate. Let $p_j^d(x)$ denote the insurer j 's rate reduction for insured x who is switching to insurer j from the current insurer i. Then the *market rate reduction* $\tilde{p}^d(x)$ is defined as an index of the rate reductions of all competitors given by

$$\tilde{p}^d(x) = \sum_{j \neq i} w_j \cdot p_j^d(x) \tag{2.19}$$

where w_j is the weight defined in (2.5).

2.9 Market Price and Cost

We define $\tilde{\phi}$ as an index of firm-specific deviations with

$$\tilde{\phi} = \sum_{j \neq i} w_j \cdot \phi_j \tag{2.20}$$

where w_j is weight defined in (2.5), and ϕ_j is the deviation factor for insurer j.

Simplifying Assumptions

For the remainder of this paper, we make further assumptions: 1) weight w_j is constant for all insured x from the current insurer i and for each insurer $j \neq i$; 2) α_j is constant for each insurer j with normalization condition $\sum_{j \neq i} w_j \alpha_j = 0$; and 3) β_j is constant for each insurer j with normalization condition $\sum_{j \neq i} w_j \beta_j = 0$, where w_j is weight defined in (2.5).

Now we can relate market price to cost by the equation

$$m_i(x) = c_i(x) + \tilde{\phi} - \beta_i - \tilde{p}^d(x) \tag{2.21}$$

The proof of equation (2.21) is given in Appendix A2.

2.10 Market Price Level, Market Cost Level, and Underwriting Cycle Index

Define *market price level* $\tilde{m}$ as an index of base prices adjusted for differentiation

$$\tilde{m} = \sum_{j \neq i} w_j \cdot (\hat{p}_j - \alpha_j) \tag{2.22}$$

where $w_{i,j}$ is the weight defined in (2.5) with $\sum_{j \neq i} w_j \alpha_j = 0$.

For standard risk x in the base rating class, we have $\tilde{p}^d(x) = 0$. It follows from (2.21) that the market price level is related to the market cost level as follows:

$$\tilde{m} = \tilde{c} + \tilde{\phi} \tag{2.23}$$

We see that $\tilde{\phi} = \tilde{m} - \tilde{c}$ can provide a measure for the relativity of the market price level to the market cost level which changes with an underwriting cycle. We shall refer $\tilde{\phi}$ as the *underwriting cycle index*

2.11 Excess Risk

Let $\hat{m}_i$ denote *base market price* given by $\hat{m}_i = \sum_{j \neq i} w_j \cdot \hat{p}_j$ where w_j is the weight defined in (2.5) and $\hat{p}_i$ is the base price.

Define the *excess risk* $v_i(x)$ of a policy x as the portion of the relative market price not reflected in the relative price

$$v_i(x) = (m_i(x) - \hat{m}_i) - (p_i(x) - \hat{p}_i) = \Delta m_i(x) - \Delta p_i(x) \tag{2.24}$$

where $\Delta m_i(x) = m_i(x) - \hat{m}_i$ is referred to as *relative market price* and $\Delta p_i(x) = p_i(x) - \hat{p}_i$ is referred to as *relative price.*

Substituting $\Delta p_i = \Delta c_i - p_i^d$ from (2.18) and $\Delta m_i = \Delta c_i - \tilde{p}^d$ from (2.21) into (2.24), we obtain another expression for excess risk

$$v_i(x) = \Delta m_i(x) - \Delta p_i(x) = p_i^d(x) - \tilde{p}^d(x) \tag{2.25}$$

Equation (2.25) states that excess risk is the difference between the rate reduction from the current insurer and the market rate reduction for new business[12]. When there are no guaranteed issue laws, we have $\tilde{p}^d(x) = 0$, and the excess risk is equal to the rate reduction from the current insurer.

2.12 Anti-selective Lapse Formula

Re-arranging the first equality of (2.24) yields the relationship:

$$p_i(x) - m_i(x) = \hat{p}_i - (\hat{m}_i + v_i(x)) \tag{2.26}$$

Now we express the lapse response function L in terms of the base price, base market price, and excess risk:

$$L(p_i) = S_a(p_i - (m_i(x) + \alpha_i)) = S_a(\hat{p}_i - (\hat{m}_i + \alpha_i + v_i(x))) \tag{2.27}$$

Equation (2.27) is the ASL formula for the individual insured. Note that L is decreasing with respect to excess risk $v(x)$. Insureds with lower excess risk (presumably the healthier ones) would have higher lapse probabilities than those with higher excess risk (presumably the sicker ones).

---

[12] Similar to rate reduction, the excess risk can be generalized to include both the effect of rate restrictions and the effect of underwriting precision relative to the competitors in the market.

3. Aggregate Behavior

In this section, we derive the aggregate behavior of a block of ISH policies. We first define the excess risk distribution for an ISH block, and use it as a mathematical characterization for the mix of insured types of the block. Then we derive the model of the aggregate behavior of a block by summing the behavior of individual insureds over the entire block. The resulting model consists of explicit analytic formulas for the aggregate behavior (e.g., lapse, loss ratio, and profit) of a block.

A note on notation: In describing the aggregate behavior of a block, we prefer to use variables in the standard form as opposed to the log form, and shall use the upper case letters for the standard form and the lower case letters for the log form. For example, we shall prefer to use price, denoted by P, as opposed to the log price p.

3.1 Excess Risk Distributions

Consider all insureds in a rating class g. Let $H = P^h$ denote the relative risk, $U = P^u$ the risk load, and $V = P^v$ the excess risk. The mix of insured in rating class g is given by the density distribution:

$$k_g(H,U) = \text{ policy density function with relative risk H \& risk load U} \qquad (3.1)$$

From (2.25) and (2.15), we have $H = U\,V\widetilde{P}^d$, where $\widetilde{P}^d$ is market rate reduction. Substitute $H = U\,V\widetilde{P}^d$ into $\kappa_g(H,U)$ and calculate the distribution

$$f(V) = \int k(H(UV\tilde{P}^d),U)\,U\,dU \qquad (3.2)$$

The normalized $f(V)$ is called the *excess risk distribution* of rating class g. Note that $f(V)$ is a risk load U weighted distribution.

3.2 Pareto Distribution

We shall make extensive use of the generalized Pareto distribution which has the form

$$F_{d,e}(V) = 1 - \left(1 + (V-1)/d\right)^{-e}, \text{ for } V \geq 1, \tag{3.3}$$

where $e > 1$ and $d > 0$ are parameters. The corresponding Pareto density function is

$$f_{d,e}(V) = \frac{e}{d}\left(1 + (V-1)/d\right)^{-(e+1)} \tag{3.4}$$

The Pareto distribution has been widely used in modeling insurance claims (see Klugman (2005)). It is used in a similar capacity for modeling excess risk distributions.

3.3 The Aggregate Model

We begin by providing notation and definitions. Let R denote the rate increase. Define *market rate increase* R_0 *as*

$$R_0 = \hat{M}/(\hat{P}_0/A) - 1 \tag{3.5}$$

where $\hat{M}$ is the market price level, A is the differentiation premium factor, $\hat{P}_0$ is the initial base rate, and we assume $\hat{P}_0/A < \hat{M}$.

Define premium rate function $P(R)$ as

$$P(R) = \begin{cases} P_0 \cdot (1+R) & \text{if } R < R_0 \\ P_0 \cdot (1+R_0)(1+R-R_0) & \text{if } R \geq R_0 \end{cases}$$

where P_0 is the initial premium rate.

Denote adjusted premium rate $P'(R) = P(R)/A$.

Let

$Persistency(R)$ = Persistency at rate increase R

$Inforce(R)$ = Number of policies inforce at rate increase R

$Lapse(R)$ = Proportion of premium lapsed at rate increase R

$Premium(R)$ =Aggregate premium at rate increase R

$Cost(R)$ = Aggregate cost at rate increase R

$Lr(R)$ =Loss ratio at rate increase R

$Profit(R)$=Aggregate profit at rate increase R

$\overline{V}$ (R) = Average excess risk at rate increase R.

For simplicity, we shall derive the aggregate model for an ISH block that consists of a single rating class. We obtain the explicit analytic formulas for the aggregate model by making further assumptions about lapse response function and excess risk distribution.

Proposition 2:

Consider an ISH block that consists of a single rating class. Let us assume that the excess risk distribution is a Pareto density function and the lapse response function $L(p)$ is a step function as defined in (2.2). Then we have the following analytic formulas for the aggregate behavior of the block:

F1) $Persistency(R) = 1 - F_{d,e}\left(1 + R - R_0\right)$

F2) $Lapse(R) = F_{d,e}\left(1 + R - R_0\right)$

F3) $Inforce(R) = I_0 \cdot (1 - F_{d,e}(1 + R - R_0))$

F4) $Premium(R) = I_0 \cdot (1 - F_{d,e}(1 + R - R_0)) \cdot \hat{P}(R)$

F5) $Cost(R) = I_0 \cdot (1 - F_{d,e}(1 + R - R_0)) \cdot C_0 \cdot \overline{V}(R)$

F6) $Lr(R) = C_0 \cdot (\overline{V}(R) / \hat{P}(R))$

F7) $Profit(R) = I_0 \cdot (1 - F_{d,e}(1 + R - R_0)) \cdot (\hat{P}(R) - C_0 \cdot \overline{V}(R))$

where

$$\hat{P}(R) = \begin{cases} \hat{P}_0 \cdot (1+R) & if\, R < R_0 \\ \hat{P}_0 \cdot (1+R_0)(1+R-R_0) & if\, R \geq R_0 \end{cases} \tag{3.6}$$

is the base premium rate,

$$F_{d,e}(1+R-R_0) = \begin{cases} 0 & if \quad R < R_0 \\ 1-\left(1+(R-R_0)/d\right)^{-e} & if \quad \mathrm{R} \geq R_0 \end{cases} \tag{3.7}$$

is the Pareto distribution with $V = 1 + R - R_0$,

$$\overline{V}(R) = \begin{cases} \left(\dfrac{ed}{e-1}\right) \times \left(\dfrac{(e-1)+d}{ed}\right) & \text{if } R < R_0 \\ \left(\dfrac{ed}{e-1}\right) \times \left(\dfrac{(e-1)+d}{ed} + \dfrac{R-R_0}{d}\right) & \text{if } \mathrm{R} \geq \mathrm{R}_0 \end{cases} \tag{3.8}$$

is the average excess risk at rate increase R, and

C_0 = Initial cost for the standard risk in the rating class.

$I_0 = Inforce(0)$ = Initial number of policies in in-force

$\overline{V}_0 = V(0)$ = Initial average excess risk.

Remarks

We shall relax the step lapse response assumption and demonstrate that F1-F7 still hold in the general sense in Appendix A3. The general forms of F1-F7 without the step lapse and Pareto assumptions can be found in the proof below.

Proof of Proposition 2:

We only need to prove F1 and F5. The rest of the formulas follow from the definitions. We prove in steps. We first obtain the general forms of the formulas F1 & F5, then apply the step lapse response assumption, and the Pareto excess risk distribution assumption to obtain F1 and F5.

By (2.24) and (3.5) we have $M(x) = (P_0(x)/A)\cdot V(x)(1+R_0)$. Substituting the last equation and the expression of P(R) in (3.6), the lapse response function (2.1) can be written as

$$L(R) = S_a(\log P(R) - (\log M + \log A))$$
$$= \begin{cases} S_a(\log((1+R)/(1+R_0)) - \log V) & \text{if } R < R_0 \\ S_a(\log(1+R-R_0) - \log V) & \text{if } R \geq R_0 \end{cases} \tag{3.9}$$

In particular, when S is the step lapse function (2.2) , we have

$$L(R) = \begin{cases} 0 & \text{if } R < R_0 \\ S_0(\log(1+R-R_0) - \log V) & \text{if } R \geq R_0 \end{cases} \tag{3.10}$$

where

$$S_0(\log(1+R-R_0) - \log V) = \begin{cases} 1 & \text{if } V < 1+R-R_0 \\ 1/2 & \text{if } \mathrm{V=1+R\text{-}R_0} \\ 0 & \text{if } \mathrm{V>1+R\text{-}R_0} \end{cases} \quad \text{if} \quad R \geq R_0 \tag{3.11}$$

We first obtain the general form of the persistency formula:

$$Persistency(R) = \int_1^\infty f(V)[1 - S_a(\log P(R) - \log V)]dV$$
$$= \begin{cases} \int_1^\infty f(V)[1 - S_a(\log((1+R)/(1+R_0)) - \log V)]dV & \text{if } \mathrm{R<R_0} \\ \int_1^\infty f(V)[1 - S_a(\log(1+R-R_0) - \log V)]dV & \text{if } \mathrm{R \geq R_0} \end{cases} \tag{3.12}$$

Then we apply the step lapse assumption (3.11) to reduce the above formula to

$$Persistency(R) = \begin{cases} 1 & if \;\; R < R_0 \\ 1 - \int_{1+R-R_0}^{\infty} f(V)dV & if \;\; R \geq R_0 \end{cases} \quad (3.13)$$
$$= \begin{cases} 1 & if \;\; R < R_0 \\ 1 - F(1+R-R_0) & if \;\; R \geq R_0 \end{cases}$$

Finally, we substitute the Pareto distribution $F(1+R-R_0) = 1-\left(1+(R-R_0)/d\right)^{-e}$ into (3.13) to arrive at formula F1.

Similarly, we first derive the general forms of the aggregate cost function:

$$Cost(R) = I_0 \int_1^{\infty} f(V)C_0 \cdot V[1 - S_a(\log P(R) - \log V)]dV$$
$$= \begin{cases} I_0 \int_1^{\infty} f(V)C_0 \cdot V[1 - S_a(\log((1+R)/(1+R_0)) - \log V)]dV & \text{if} \;\; R < R_0 \\ I_0 \int_1^{\infty} f(V)C_0 \cdot V[1 - S_a(\log(1+R-R_0) - \log V)]dV & \text{if} \;\; R \geq R_0 \end{cases} \quad (3.14)$$

Then we apply the step lapse assumption (3.11) to reduce the last term into

$$Cost(R) = \begin{cases} I_0 C_0 \int_1^{\infty} f(V) \cdot V dV & if \;\; R < R_0 \\ I_0 C_0 \int_{1+R-R_0}^{\infty} f(V) \cdot V dV & if \;\; R \geq R_0 \end{cases} \quad (3.15)$$

We may write the above as

$$Cost(R) = Inforce(R) \cdot C_0 \cdot \overline{V}(R), \quad (3.16)$$

where

$$\overline{V}(R) = \begin{cases} \int_1^{\infty} f(V)VdV \Big/ \int_1^{\infty} f(V)dV & if \;\; R < R_0 \\ \int_{1+R-R_0}^{\infty} f(V)VdV \Big/ \int_{1+R-R_0}^{\infty} f(V)dV & if \;\; R \geq R_0 \end{cases} \quad (3.17)$$

Finally, we substitute the Pareto distribution $f_{d,e}(V) = e/d\left(1+(V-1)/d\right)^{-(e+1)}$ into (3.17) and perform basic calculus to obtain the formula F5. This completes the proof.

4. Application

In this section, we apply the aggregate model to the behavior of different ISH blocks and the impact of various factors. We first examine the mix of insured and the aggregate behavior of lapse, loss ratio, and profit of these ISH blocks. We obtain the optimal pricing solutions for these ISH blocks. We also examine the impact of the market price level and market cost level. Finally, we determine the impacts of underwriting cycle and the effect of risk selection, firm cost, and differentiations.

4.1 ISH Blocks and Mix of Insureds

We shall consider several ISH blocks in Table 1. The Pareto parameter e, d, the base cost C_0, and the average excess risk V_0 are provided for each block. In addition, the base market price level $\hat{M}$ and base premium rate $\hat{P}_0$ are provided. For simplicity, $\hat{M}$, A, and $\hat{P}_0$, are assumed to be the same for all blocks, and as a result, so are the market rate increase $R_0 = \hat{M}/(\hat{P}_0/A) - 1$.

Figure 2 graphs the excess risk distributions in these blocks. A higher excess risk is generally associated with an impaired life[13]. In all blocks, the insured density decreases with excess risk V, with the highest insured density at $V = 1$. But the slopes of the decrease in these blocks are different. The slope is the steepest in Block 3, and the flattest in Block1. Intuitively,

---

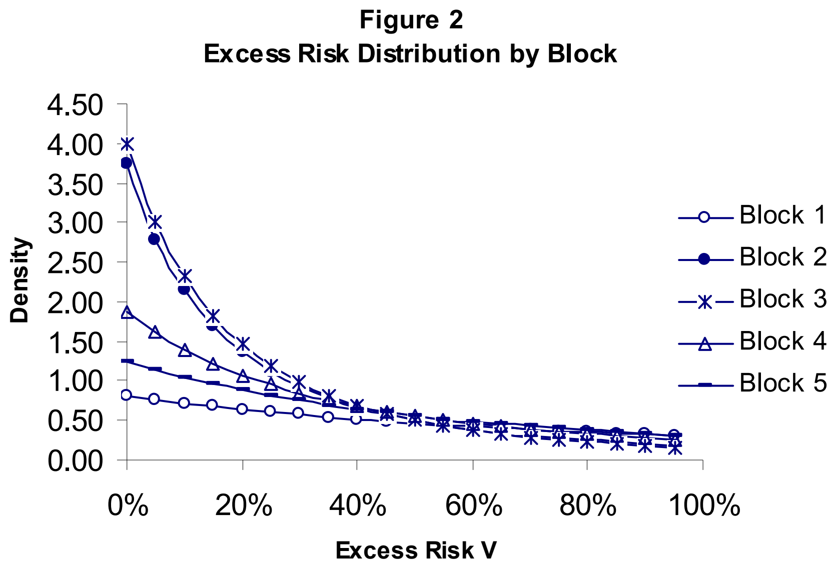
[13] To be precise, low excess risk only means that the relative risk load H is close to the relative cost C^h. A healthy live has an excess risk that is close to zero. However, an impaired live can have a low excess risk when the premium rate can be set to the cost without restrictions as in the case of a newly underwritten policy.

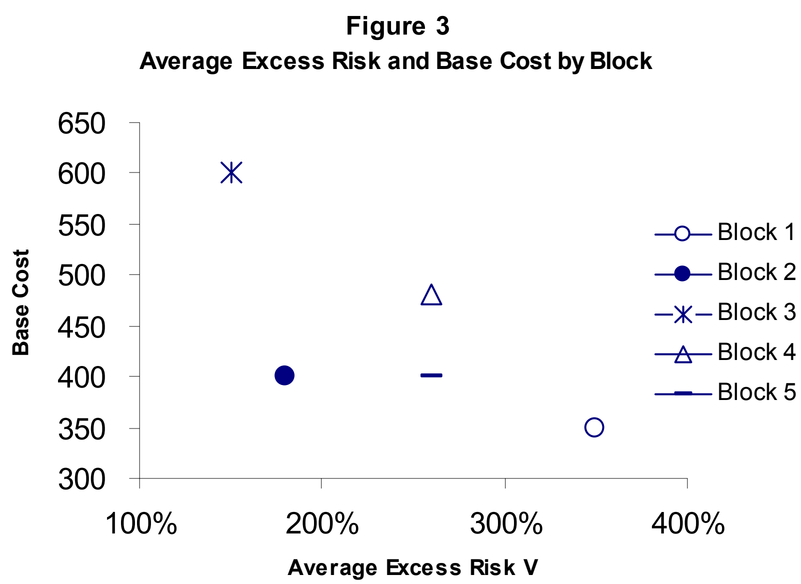
Block 3 has the least proportion of impaired lives, and Block 1 has the greatest proportion of impaired lives.

Figure 3 plots these blocks along the two dimensions: average excess risk $\bar{V}_0$ vs. base cost C_0. Block 1 has the highest average excess risk and the lowest base cost. Block 2 has a low average excess risk and a low base cost. Block 3 has the lowest average excess risk and the highest base cost. Blocks 4 and 5 are intermediate in both average excess risk and base cost, with the base cost in Block 4 slightly higher than that in Block 5.

Table 1
Parameters of Block 1 - Block 5

Parameter	Block 1	Block 2	Block 3	Block 4	Block 5
e	2	1.5	2	1.5	2
d	2.5	0.4	0.5	0.8	1.6
$\bar{V}_0$	3.5	1.8	1.5	2.6	2.6
C_0	350	400	600	480	400
M_0	1200	1200	1200	1200	1200
A	1.0	1.0	1.0	1.0	1.0
P_0	1000	1000	1000	1000	1000
R_0	20%	20%	20%	20%	20%





4.2 ASL and Rate Increase

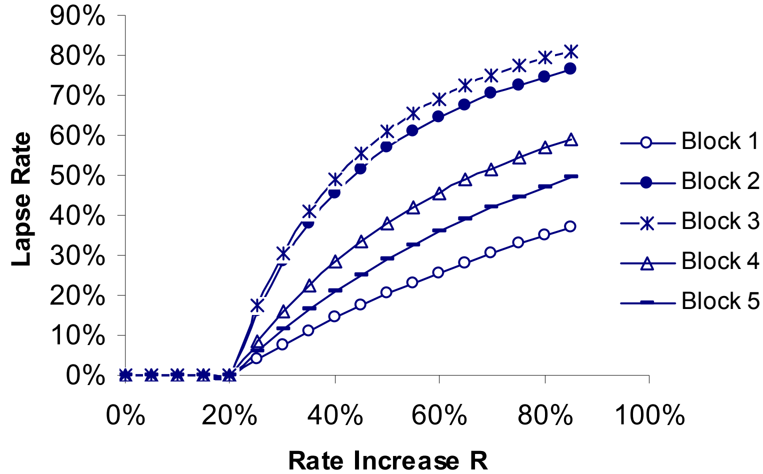
We begin by considering how ASL occurs in a block as R increases. When $R < R_0$, by $R_0 = \hat{M} / (\hat{P}_0 / A) - 1$, we have $\hat{P}(R) / A < \hat{P}(R_0) / A = \hat{M}$. This implies that for all insured x in the block, $P(x) / A < M(x)$, and $L(x) = 0$ (under the step lapse assumption). Thus ASL does not occur. When $R \geq R_0$, we have $\hat{P}(R) / A \geq \hat{M}$. For all insured x with $V(x) < (P(R) / A) / \hat{M}$, we have $P(R)/A > V(x)\cdot\hat{M} = M(x)$, and $L(x)=1$. Similarly, for all insured x with $V(x) > (P(R)/A)/\hat{M}$, we have $P(R)/A < M(x)$, and $L(x)=0$. Thus, as R increases, more and more insureds with increasingly higher excess risk would start to lapse.

4.3 Aggregate Lapse vs. Rate Increase

Figure 4 shows how the aggregate lapse rate changes with R in these blocks. For $R < R_0$, the aggregate lapse rate is zero in all blocks. This is expected because $R < R_0$ implies $P(x)/A < M(x)$ for all x. For $R \geq R_0$, the aggregate lapse rates increase with R in all blocks. But the slopes of the lapse rate increases are different. The slope of lapse rate increase is closely linked to the excess risk distribution of the block.

In Block 1, the lapse rate increase is the slowest, because the excess risk distribution has the least steep slope. In Block 3, the lapse rate increase is fastest because the excess risk distribution has the steepest slope. The lapse rate increase is faster in Block 3 than in Block 4, which in turn is faster than in Block5, which corresponds to the slopes in their excess risk distributions.

Figure 4
Aggregate Lapse by Block



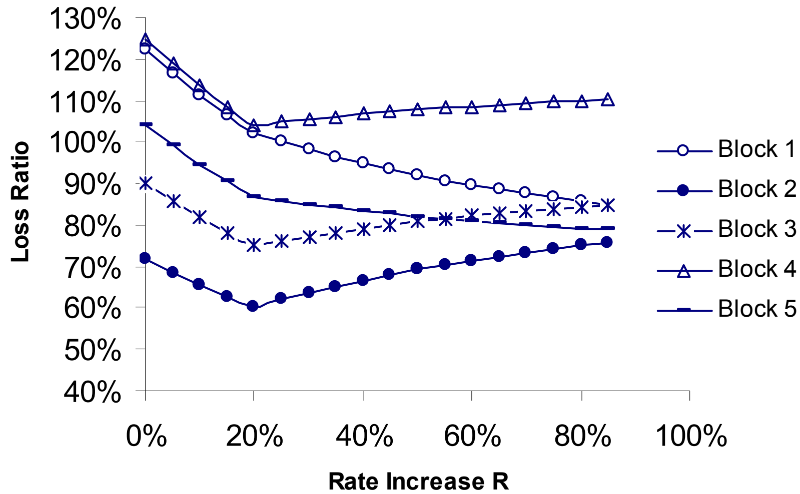
4.4 Loss Ratio vs. Rate Increase

Figure 5 shows how the loss ratio changes with R by block. When $R < R_0$, the loss ratio decreases with R in all blocks. This is because when the lapse rate is zero, the aggregate cost of the block remains constant but the aggregate premium keeps increasing with R.

When $R \geq R_0$, the loss ratio exhibits different behavior in these blocks. To understand this, recall that $Lr(R) = C_0 \cdot (\overline{V}(R) / \hat{P}(R))$, and note that the change in $\overline{V}(R)$ reflects the impact of ASL on the mix of insured.

In Block 1, $Lr(R)$ continues to decrease with R for $R \geq R_0$, but at a slower rate than before because of a moderate increase in $\overline{V}(R)$ due to a moderate impact of ASL. In Block 2 and Block 3, the loss ratio reaches a minimum at $R = R_0$ and reverses direction to rise with R for $R \geq R_0$. This is because the increase in $\overline{V}(R)$ exceeds the increase in *P(R)* due to a strong impact of ASL.

Figure 5
Loss Ratio by Block



Assessment Spiral

Significantly, Block 2 provides an illustration of the "assessment spiral" phenomenon: The loss ratio $Lr(R)$ after the rate increase of R=85% is actually higher than the loss ratio at R=0 prior to the rate increase.

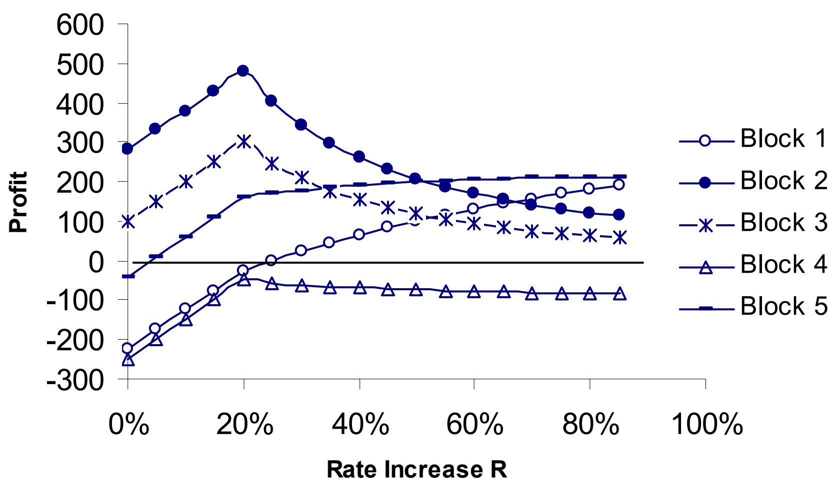
In Block 4, the loss ratio rises slowly with R. The increase in $\overline{V}(R)$ is slightly higher than the increase in *P(R).* In Block 5, the loss ratio decreases slowly as R increases, where the increase in $\overline{V}(R)$ is slightly lower than the increase in *P(R)*.

4.5 Aggregate Profit vs. Rate Increase

Figure 6 shows how the aggregate profit changes with R. For $R < R_0$, the aggregate profit increases with R in all blocks. This is consistent with the earlier observations that the loss ratio decreases with R in all blocks.

For $R \geq R_0$, the behavior of aggregate profit varies significantly by block. In Block 1, the aggregate profit continues to increase with R at a rapid rate, going from a net loss to a net profit. In Block 2 and Block 3, the aggregate profit reaches the maximum at $R = R_0$ and starts to decrease for $R \geq R_0$. In Block 4 and Block 5, the aggregate profit attains a local maximum at $R = R_0$, and then remains relatively constant as R increases. The aggregate profit of block 4 remains in negative territory for $R > R_0$, while the aggregate profit of block 5 stays in positive territory. The difference in behavior can be similarly attributed to the impact of ASL on the mix of insureds and ultimately, the excess risk distributions.

Figure 6
Aggregate Profit by Block



4.6 Optimal Pricing

We now determine the optimal price increase that maximizes the aggregate profit for each block. In Block 1, we see that for R inside the range of the chart, the aggregate profit continues to increase with R[14]. Thus the optimal price increase is to maximize the premium rate increase R to the extent possible. In practice, insurer's ability to increase the premium rate would be severely limited by rating laws.

In Block 2 and Block 3, the optimal rate increase is $R = R_0$, where $R_0 = \hat{M} / (\hat{P}_0 / A) - 1$. At the optimal price level, we have $\hat{P}_0 / A = \hat{M}$, or the adjusted price level matches the market price level. As a result, the lapse rate at the optimal price increase is minimal. This motivates the definition below.

Sustainable Blocks

Definition: An ISH block is called *sustainable* if the block attains the maximum profit at $R = R_0$.

Comments: Blocks 2 and 3 are examples of sustainable ISH blocks. Sustainable blocks have a decent proportion of low excess risk which tends to be associated with healthy lives. Sustainable blocks are inherently stable because the insurer should have a strong profit incentive to keep the rate increase at a moderate level and the lapse rate low[15].

In Block 4, the profit remains in negative territory at all premium rate levels. Rate increase $R = R_0$ is a local optimum. Note that as $R \to \infty$, $Profit(R) \to 0$ so that $R \to \infty$ is the optimal rate increase for Block 4. But in the real world, due to regulatory limits on rate increase, $R = R_0$ is likely to be the practical optimal choice.

In Block 5, the optimal rate increase is to maximize the rate increase. But given the small gain in profit as R becomes large, we may prefer, in the real world, rate increase $R = R_0$, which would allow us retain all insureds at the cost of a slightly lower profit, a trade-off that would be considered as "optimal" in an uncertain market.

---

[14] The exception is the extreme case of $R \to \infty, Profit(R) \to 0$.

[15] A potential implication is that some forms of rate increase limits indexed to the medical cost trend can promote stability in the ISH market at no cost to ISH insurers.

Note that for all blocks except Block 2, the least optimal rate increase is $R = 0$ or no rate increase at all.

4.7 Estimating Market Price Level

For optimal pricing, we need to know the market price level. But estimating $\hat{m}_i = \sum_{j \neq i} w_j \cdot \hat{p}_j$ directly can be difficult in practice as $\hat{p}_j$ is not readily available.

Fortunately, we can estimate $\hat{m}_i$ using the relationship $L(\hat{p}) = S_a(\hat{p} - (\hat{m} + \alpha))$ in (2.7), where S_a is a monotonic function. This equation suggests that we can determine $\hat{p} - (\hat{m} + \alpha)$ by measuring $L(\hat{p})$ as a function of $\hat{p}$. As $L(\hat{p}) \to 1/2$, we have $\hat{p} \to \hat{m} + \alpha$, which means that when the lapse rate of standard risks approaches ½, we have $\hat{p} = \hat{m} + \alpha$.

To keep $\hat{m}(t)$ up-to-date on a more frequent basis, we can use an index of select competitors' prices, denoted by $idx(t)$, to track the movement of $\hat{m}(t)$ between time t_1 and time t_2:[16]

$$\hat{m}(t_2) \cong \hat{m}(t_1) + \lambda \cdot [idx(t_2) - idx(t_1)] \tag{4.1}$$

where λ is a coefficient that can be determined from historical data.

4.8 Effect of Market Price Level and Market Cost Level

Now we shift our focus to the impact of market price level (MPL) and market cost level (MCL), which can change due to competitors' price actions, general cost inflation, and medical inflation.

[16] The index has a smaller number of competitors compared with a full market index.

Figure 7 shows the aggregate profit of a sustainable block at different market price levels. The parameters of the block are given in Table 2. We see that a higher (or lower) market price level translates into a higher (or lower) aggregate profit of the block for all R. Moreover, the optimal price increase is to match the new market price level, with correspondingly higher (or lower) maximum profit.

Table 2
Block Parameters and Market Price Level (MPL)

Parameter	MPL=1200	MPL=1250	MPL=1150
e	1.5	1.5	1.5
d	0.4	0.4	0.4
$\overline{V}_0$	1.8	1.8	1.8
C_0	400	400	400
M_0	1200	1250	1150
A	1.0	1.0	1.0
P_0	1000	1000	1000
R_0	20%	25%	15%

Figure 7
Effect of Market Price Level on Aggregate Profit

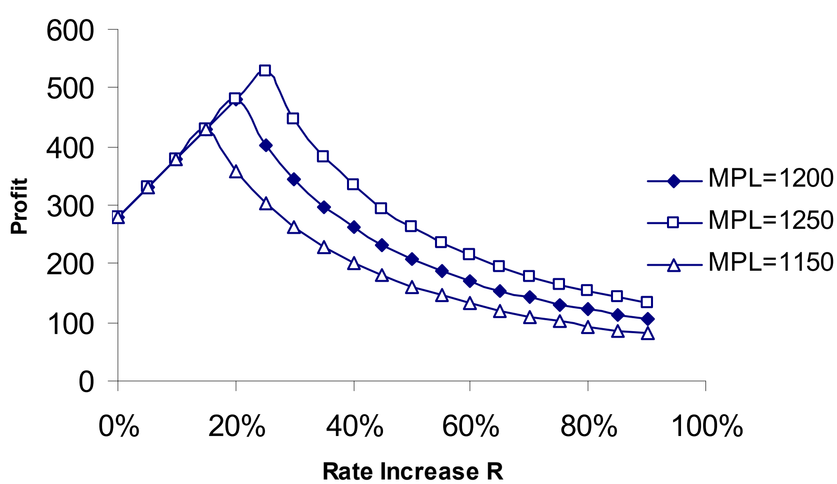
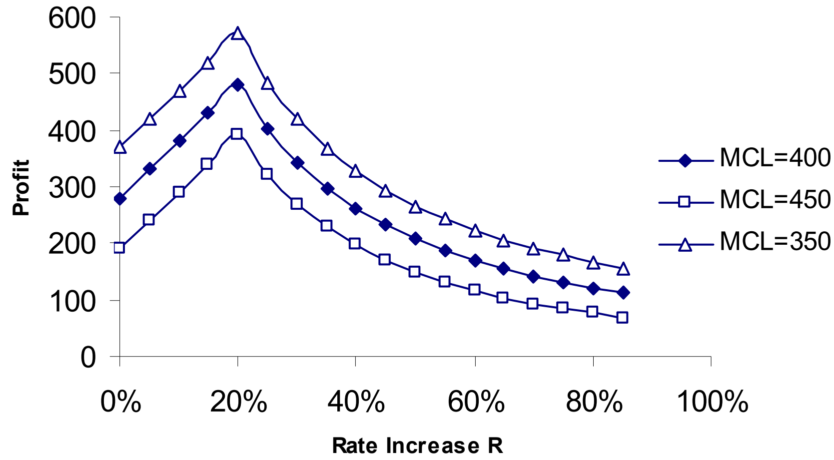


Figure 8 shows the aggregate profit of the same block at different market cost levels. The parameters of the block are given in Table 3. We see that higher (or lower) cost level means a lower (or higher) aggregate profit of the block for all R. However, the optimal price increase remains the same regardless of the market cost level, but the corresponding maximum profit is lower (or higher).

Table 3
Block Parameters and Market Cost Level (MCL)

Parameter	MCL=400	MCL=450	MCL=350
E	1.5	1.5	1.5
D	0.4	0.4	0.4
$\overline{V}_0$	1.8	1.8	1.8
C_0	400	450	350
M_0	1200	1200	1200
A	1.0	1.0	1.0
P_0	1000	1000	1000
R_0	20%	20%	20%

Figure 8
Effect of Market Cost level on Aggregate Profit



4.9 Effect of Underwriting Cycle and Internal Drivers

Finally, we apply the model to understanding the impact of underwriting cycle and some internal drivers.

Define *profit capacity* of a block as the maximum aggregate profit attained when the price increase is optimal. In other words, profit capacity is the potential profit to be realized by optimal pricing.

For simplicity, we consider the case of a sustainable ISH block for which the maximum profit is attained when $\hat{P}/A = \hat{M}$. It follows easily from formula F5 that the profit capacity of a sustainable block can be expressed as

$$\text{Pr}\,\mathit{ofit\ Capacity} = I_0 \cdot \tilde{C} \cdot [\mathrm{A} \cdot \tilde{\Phi} - \mathrm{B} \cdot \overline{V}_0] \qquad (4.2)$$

where I_0 = initial in-force, $\tilde{C}$ = market cost level, A = differentiation premium factor, B = firm cost factor, $\tilde{\Phi}$ = underwriting cycle index, and $\overline{V}_0$ = initial average excess risk.

The last expression in (4.2) formalizes the impact of underwriting cycle and internal drivers on the profit capacity of a block.

Effect of Underwriting Cycle

Equation (4.2) states that the profit capacity of a block is dependent on the underwriting cycle $\tilde{\Phi}$ which the insurer has no control over. In a down cycle, the lower value of $\tilde{\Phi}$ means a lower profit capacity of the block, and the insurer has no choice but to accept the lower profits. In an up cycle with a higher value of $\tilde{\Phi}$, the insurer can achieve higher potential profit.

Effect of Internal Drivers

Equation (4.2) also states how the profit capacity of a block is related to several key internal drivers. To increase the profit capacity of the block, the insurer has three basic strategies: 1) raise the differentiation premium A by raising perceived quality; 2) reduce the firm-specific cost B, which usually means lowering expense and obtaining better provider discounts; 3) lower the average excess risk $\overline{V}_0$, which means gaining an advantage in health risk assessment accuracy[17].

The quantity $\mathrm{A}\cdot\tilde{\Phi}-\mathrm{B}\cdot\overline{V}_0$ behaves like a profit margin. When $\mathrm{A}\cdot\tilde{\Phi}-\mathrm{B}\cdot\overline{V}_0>0$, the insurer has a positive profit capacity, and when $\mathrm{A}\cdot\tilde{\Phi}-\mathrm{B}\cdot\overline{V}_0<0$, a negative profit capacity. An implication is that for a given value of $\tilde{\Phi}$, insurers with different values of A, B, and $\overline{V}_0$, would fare differently in terms of ability to stay profitable.

5. Conclusion

We have developed a model for the behavior of ISH policies in a competitive market based on a formal framework. In this model, ASL arises as a result of consumers' rational choices in a

[17] As mentioned earlier, excess risk may be generalized to include both effect of rate restrictions and underwriting precision relative to the competitors in the market.

competitive market with rate restrictions. We applied the model to address several fundamental problems and major applications. We examined a wide range of aggregate behavior of ISH blocks in various scenarios. We obtained the optimal pricing solutions for different types of ISH blocks. We also examined the impact of changes in the market, underwriting cycle, and several internal drivers. Along the way we obtained some significant findings and new insights. Key model parameters, such as market price, can be estimated from empirical data.

A significant finding was that for a large class of so-called "sustainable" ISH blocks, the insurer can achieve maximum profit with a minimal lapse rate, and the optimal rate increase is consistent with the general medical cost trend. These blocks can be managed to produce long term stability and profitability. A potential implication is that some forms of renewal rate cap indexed to a suitable medical cost index, could help the stability of the ISH market, benefiting both the insureds and the insurers. This can be accomplished via either product design or rate regulations.

We can generalize the model in several directions. First, we can, as mentioned earlier, relax the main assumptions by considering non-step lapse response, non-Pareto excess risk distribution, and a block with multiple rating classes. We can also generalize the model by considering multiple local markets with local price and cost. Finally, we can incorporate benefit dampening, non-price-induced lapses, and non-switching price-induced lapses which were omitted in the model but can provide added realism for practical application.

In this paper, we did not consider future new business explicitly. Part of the reason, other than paper length, is that for the short timeframe we are concerned with, future business does not substantially alter the aggregate behavior of a block in the model. Incorporating future business can be done in close parallel to the existing policies in the current model. The sales behavior can be modeled analogously to the lapse behavior. Note that all future business has an excess risk value of zero, because the renewal rate restrictions do not apply and the guaranteed issue laws would have the same effect on all insurers (Strictly speaking we also need to assume no difference in underwriting effectiveness among insurers. See Section 2).

We may want to consider extending the model to the multi-period case. We did not consider the multi-period case in this paper because to do so would add a great deal of complexity without providing genuine new insights. For aggregate behavior of ISH blocks, the single-period model we developed is sufficient in practice with exceptions of relatively rare situations. Fortunately, many key ideas and techniques described in this paper, including excess risk and market price, can be readily generalized to the multi-period case.

Appendices

A1. Proof of Proposition 1

Note that $S_a^{'}(p_i)$ is balanced means $p_i{*} = \overline{p}_i$. Note that under the differentiability assumption, we have $dL/dp_i = \sum_{j \neq i} w_j \times dS_b / dp_i$, where w_j is independent of p_i. Thus the half lapse price for insured x can be expressed as

$$
\begin{aligned}
p_i^* &= \overline{p}_i \\
&= \int \frac{dL_i(p_i)}{dp_i} p_i \, dp_i \\
&= \sum_{j \neq i} w_j \times \int \frac{dS_b(p'_i - p'_j)}{dp_i} p_i \, dp_i \\
&= \sum_{j \neq i} w_j \times \int \frac{dS_b(p'_i - p'_j)}{dp_i} [(p'_i - p'_j) + (p'_j + \alpha_i)] \, dp_i \\
&= \sum_{j \neq i} w_j \times [\int \frac{dS_b(p'_i - p'_j)}{dp_i} (p'_i - p'_j) dp_i + (p'_j + \alpha_i) \times \int \frac{dS_b(p'_i - p'_j)}{dp_i} dp_i]
\end{aligned}
$$

Note that S_b is balanced implies

$$\int \frac{dS_b(p'_i - p'_j)}{dp_i} \times (p'_i - p'_j)\, dp_i = 0$$

and that S_b is a s-function implies

$$\int \frac{dS_b(p'_i - p'_j)}{dp_i}\, dp_i = 1$$

where $p_i{}'$ and p_i are linearly related by a constant term. Substituting these two integral values into the last expression for the half-lapse price, we have

$$\begin{aligned} p_i^* &= \sum_{j \neq i} w_j \times [0 + (p'_j + \alpha_i) \times 1] \\ &= \sum_{j \neq i} w_j \times (p_j - \alpha_j + \alpha_i) \\ &= \sum_{j \neq i} w_j \times p_j + \alpha_i = m_i + \alpha_i \end{aligned}$$

Note the normalization condition $\sum_{j \neq i} w_j \cdot \alpha_j = 0$ in the last step. This completes the proof.

A2. Proof of Market Price –Cost Relationship

By (2.17) we have

$$\begin{aligned} m_i &= \sum_{j \neq i} w_j \cdot p_j \\ &= \sum_{j \neq i} w_j \cdot (c_j - p_j^d + \phi_j) \\ &= \sum_{j \neq i} w_j \cdot c_j - \tilde{p}^d + \tilde{\phi} \end{aligned}$$

By (2.9), (2.10), and (2.11), we have

$$\begin{aligned} \sum_{j \neq i} w_j \cdot c_j &= \sum_{j \neq i} w_j \cdot (\hat{c}_j + c^g + c^h) \\ &= \tilde{c} + c^g + c^h \\ &= (\hat{c}_i - \beta_i) + c^g + c^h \\ &= c_i - \beta_i \end{aligned}$$

Thus

$$\begin{aligned} m_i &= (c_i - \beta_i) + \tilde{\phi} - \tilde{p}^d \\ &= c_i + \tilde{\phi} - \beta_i - \tilde{p}^d \end{aligned}$$

This completes the proof.

Remark: Note that if new business is fully underwritten, we have $\tilde{p}^d = 0$. Then the relationship is reduced to $m_i = c_i + \tilde{\phi} - \beta_i$.

A3. Non-step Lapse Assumption

We want to demonstrate that F1 – F7 still hold in the general sense if we relax the step lapse assumption. We shall do so by using numerical examples. Let us consider Block 2 from Table 1 and a non-step lapse function given by a cumulative normal distribution with a scaling factor of 0.02

$$S(z) = \int_{-\infty}^{z} 1/\sqrt{2\pi}\exp(-1/2\cdot(z/0.02)^2)ds\,.$$

Using the general form of the formulas from the proof of Proposition 2 in Section 4, we compute numerically the aggregate profit and the aggregate lapse of the block using both the above non-step lapse function and the step lapse function.

The results are shown in Figure 9 and Figure 10. The aggregate profit and lapse of the block remain the essentially same but are somewhat smoother after we switch from the step lapse to non step lapse function. The key points of interest such as the optimal rate increase have shifted only slightly. The similar observations can be made for the other formulas. Thus the general relationships represented by F1-F7 still hold with the non-step lapse assumption.

Figure 9
Aggregate Profit With Non-step Lapse Response

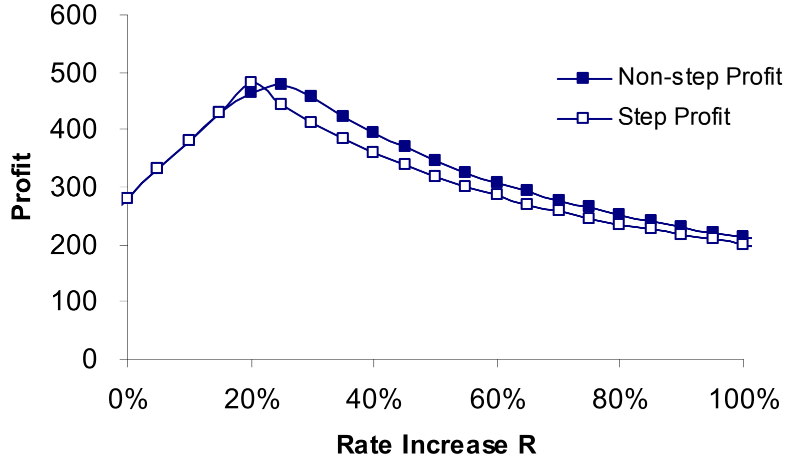
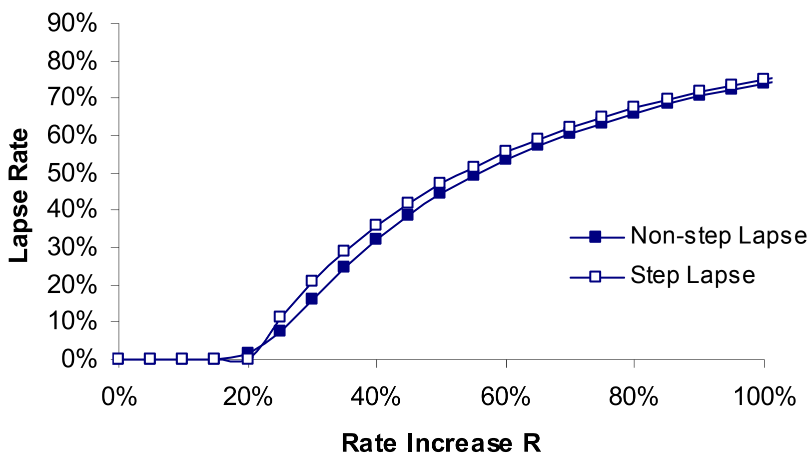


Figure10
Aggregate Lapse Rate With Non-step Lapse Response



References

Belli, Paolo. 2001. How Adverse Selection Affects the Health Insurance Market. World Bank Policy Research Working Paper No. 2574.

Bolnick, Howard J.. 1983. Why Small Group Programs Fail. Best's Review, Vol. 84, No. 6, A. M. Best Company.

Bluhm, William P.. 1983. Cumulative Anti-selection Theory. TSA XXXIV, pp. 215-246, Society of Actuaries.

Bluhm, William P., eds. 2003. Group Insurance 4th ed., ACTEX Publications

Blumberg, Linda J. and Nichols, Len. 1996. Health Insurance Market Reforms. Urban Institute Research Report.

Campbell, John Y., Lo, Andrew W., MacKinlay, A. Craig. 1996. The Econometrics of Financial Markets. Princeton University Press.

Dolan, Robert J. and Simon, Hermann. 1996. Power Pricing. The Free Press.

Folland, Sherman, Goodman, Allen C., and Stano, Miron. 2004. The Economics of Health and Health Care, 4th ed. Prentice Hall.

Greene, William H.. 2000. Econometric Analysis, 4th ed. Prentice Hall.

Klugman, Stuart A., Panjer, Harry H., and Willmot, Gordon E.. 2005. Loss Models: From Data to Decisions, 2nd ed. John Willey &Sons.

Mas-Colell, Andreu, Whinston, Michael D., and Green, Jerry R. 1995. Microeconomic Theory. Oxford University Press.

Nagle, Thomas T. and Holden, Reed K.. 1995. The Strategy and Tactics of Pricing, 2nd ed. Prentice Hall.

O'Grady, Francis T. 1988. Individual Health Insurance. Society of Actuaries.

Panjer, Harvey H., eds. 1998. Financial Economics: With Applications to Investments, Insurance, and Pensions. The Actuarial Foundation.

Pashigian, Peter B.. 1999. Price Theory and Applications. McGraw-Hill.

Wilson, Robert B.. 1993. Nonlinear Pricing. Oxford University Press.