Pay What Your Dad Paid: Commitment and Price Rigidity in the Market for Life Insurance*

Radoslaw Paluszynski† Pei Cheng Yu‡

This version:
September 17, 2018

Abstract

Life insurance premiums display significant rigidity in the data, on average adjusting once every 3 years by more than 10%. This contrasts with the underlying marginal cost which exhibits considerable volatility due to the movements in interest and mortality rates. We build a model where policyholders are held-up by long-term insurance contracts, resulting in a time inconsistency problem for the firms. The optimal contract takes the form of a simple cutoff rule: premiums are rigid for cost realizations smaller than the threshold, while adjustments must be large and are only possible when cost realizations exceed it. We use a calibrated version of the model to show that it matches the data and captures several aspects of premium rigidity in the cross-section and over time.

Keywords: Life insurance, Time inconsistency, Hold-up problem, Commitment, Flexibility

JEL Classification Numbers: G22, L11, L14

*We are indebted to Motohiro Yogo for his encouragement and advice. We also thank Martin Byford, Igal Hendel, Kyle Herkenhoff, Anton Kolotilin, Edward Kung, Hongyi Li, Hodaka Morita, David Rahman, Breanne Richins, Jim Schmitz and Kei-Mu Yi for many helpful comments, as well as Witold Paluszyński for his support with automating the collection of data.

†University of Houston (e-mail: *rpaluszynski@uh.edu*)

‡University of New South Wales (e-mail: *pei-cheng.yu@unsw.edu.au*)

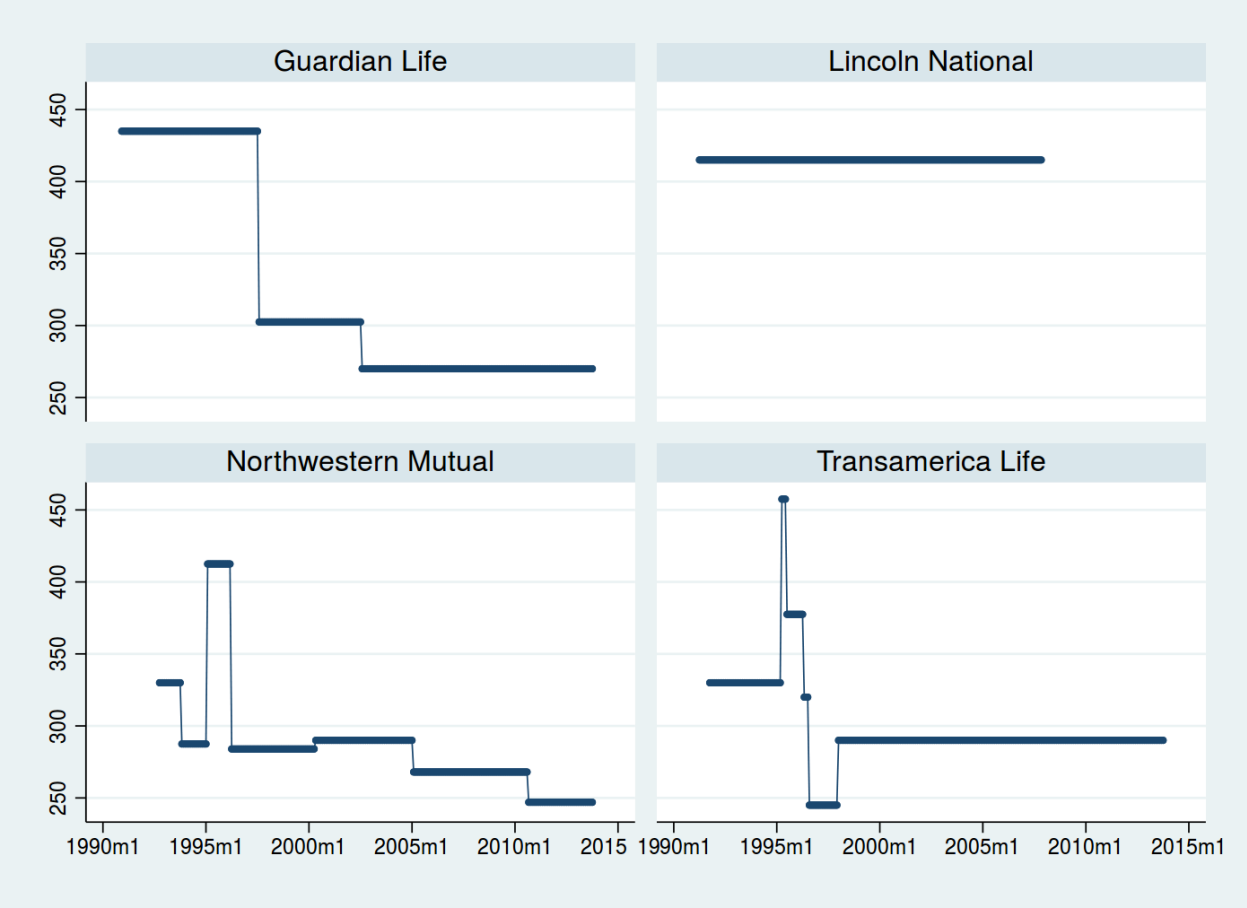
1 Introduction

Several studies have documented and modeled the stickiness of consumer goods prices, while price rigidity of long-term financial services remains less examined. This is important because such products constitute a non-negligible fraction of household spending. For example, the share of married household expenditures on personal insurance and pensions in 2016 was 11.1%, compared to 11.8% on food or 17.5% on transportation.[1] This paper documents the high degree of price rigidity for a specific long-term financial contract, life insurance. We propose a novel theory of endogenous price rigidity based on the trade-off between commitment and flexibility for the life insurance company, which is generated by the consumer hold-up problem in long-term contracting.

We show that life insurance premiums are characterized by long periods of rigidity with occasionally sizable adjustments. This is intriguing because the underlying marginal cost of life insurance is volatile over time, with a monthly coefficient of variation of 6.3% and the average absolute month-to-month change of 1.4% of the mean. In the data though, the overall probability of a monthly premium change amounts to just 2.6%. This implies an average premium duration of roughly 39 months, placing life insurance on the far-right tail of the price change frequency distribution documented by Bils and Klenow (2004). Figure 1 presents an illustrative plot of premiums over time for the most significant and longest-observed companies in our sample. Remarkably, some products have maintained a constant premium for over 20 years, rendering it possible that a son could pay the same amount for life insurance as his father used to, as the title of this paper goes.

To explain the empirical findings, we construct a model where the firm faces a commitment versus flexibility trade-off, which stems from the consumer hold-up problem. Consumers live up to three periods and form an overlapping generations structure. They decide between purchasing two types of policies from monopolistically competitive firms: renewable or non-renewable. To do so, they must incur a transaction cost, which represents the monetary expenses and opportunity cost of research and medical examination. In addition, they face adverse health shocks in the future, which could lead to significant premium hikes if they searched for a new policy. On the other hand, renewable policyholders are guaranteed coverage in the second period without incurring any additional cost. Therefore, renewable policyholders are locked into a long-term relationship with the company, limiting their future options. As a result, the firm has an incentive to increase renewal prices, but this lowers the consumers' willingness to sign. In essence, the firm is time-inconsistent and values commitment. At the same time, the firm also faces stochastic cost shocks in the second period, so it

[1] Bureau of Labor Statistics, Consumer Expenditures 2016.



Note: For illustration, we plot here the companies with: i. a share in the California life insurance market of at least 1% according to California Department of Insurance (2004), and ii. a continuous presence in our sample for at least 180 months (15 years). The total share in the life insurance market for these four companies was 9.3% in 2003.
Source: Compulife Software, 1990-2013.

Figure 1: ART premiums over time for selected companies

values flexibility. Policyholders do not observe the shocks and are unsure whether premium changes are due to being held-up or due to changes in the cost. Consequently, the firm designs premium schedules that leave consumers with no doubt that premium increases are due to sizable changes in the cost.

To balance the need for commitment and the desire for flexibility, the firm optimally commits to a constant pricing schedule for small cost variations, while price adjustments are large and only possible when cost shocks are substantial. This is because the premium adjustment is typically minor for small cost shocks, so it does not affect demand due to the policyholders being locked-in. This implies that any variation in prices for low cost shocks would not be credible because the firm could increase premiums slightly without

losing any consumers. As a result, the firm commits to a single low premium for all shocks sufficiently small to gain credibility. For premium adjustments to be credible, deviations from the low premium need to be significant so price increases induce enough reduction in demand. This way, adjustments can only be attributed to large cost shocks and not ascribed to opportunistic behavior.

We show that the optimal premium profile has a simple cutoff rule: premiums are low and rigid for marginal costs below an endogenously determined threshold, while above this threshold premiums are high and initially rigid before full flexibility is possible.[2] Our model explains why level-term insurance policies have a non-guaranteed premium schedule that affords them the room to be flexible, while the finalized premiums rarely deviate from it.

Having established the general properties of an optimal premium schedule, we proceed to solve the model numerically and calibrate it to match the quantitative features of ten-year renewable insurance. The model generates realistic premium amounts and predicts a jump in the premium of 12% when the cost shock realization switches between the low and high regions, in line with what we observe on average in the data. Interestingly, we find that in the numerical exercise the company does not find it optimal to allow for a fully flexible premium schedule, even for the largest cost shocks. Instead, the flexibility is achieved only by switching between the two rigid premiums. We then use the quantitative model to perform several comparative statics exercises, highlighting the subtle differences between the consumer's hold-up problem and the traditional monopoly power.

We finally test the model predictions directly on our dataset of insurance premiums. First, we find that between the 1990s and the 2000s, a period of time when the consumer's hold-up problem was likely weakened due to falling transaction costs and less adverse health shocks, the frequency of premium changes increased and the average size of such adjustments fell, bringing the pricing patterns of life insurance companies closer to those in typical consumer goods markets. Second, we show that as the level-term of a renewable policy increases, which relieves the quantitative power of our mechanism due to a higher probability of policy termination before the renewal date, premiums are also more likely to be adjusted and exhibit smaller jumps. Third, we demonstrate that life insurance companies tend to respond to marginal cost shocks predominantly on the *extensive margin*, by increasing the hazard of a premium change, while no apparent effect is detected on the *intensive margin*, by varying the size of a premium change. This observation is in line with our model where the pricing dynamics is based on a threshold rule. We conclude by testing common alternative models of price rigidity, such as time- and state-dependent sticky price models, and show that their predictions are not consistent with the facts about life insurance premiums.

[2]The upper rigidity result is similar to Melumad and Shibano (1991) or Alonso and Matouschek (2008).

To summarize, our paper makes three contributions. On the empirical side, we provide evidence for the frequency and size of price changes in the life insurance market, and propose an explanation for this phenomenon. On the theoretical side, we build a model where the optimal incentive compatible contract necessarily features price rigidity and a discrete jump. Finally, we calibrate our commitment versus flexibility model to the life insurance market and show that the predicted premium rigidity and jumps are quantitatively significant.

Our work builds upon several strands of literature in economics. In a seminal paper, Hendel and Lizzeri (2003) examine cross-sectional data on life insurance premiums to show that with one-sided commitment on the firm's side, and risk of health reclassification on the consumers' side, the optimal insurance contracts exhibit front-loading. Hendel and Lizzeri (2003) use the data from a single point in time (July 1997), making an implicit yet crucial assumption that companies never deviate from the current non-guaranteed premiums.[3] Our paper therefore provides both empirical and theoretical foundation for their assumption. Alternatively, Gottlieb and Smetters (2016) show that the front-loading of premiums can be explained by insurers' motive to profit from the policyholders who lapse early in their life-cycle upon experiencing adverse income shocks. The aforementioned papers develop theoretical models for the premium structure faced by a *fixed individual* over time. By comparison, our work tries to explain the premium evolution of a life insurance contract for a *fixed age group* over time.

More recently, Koijen and Yogo (2015) show that life insurance companies sold long-term policies at largely negative markups following the financial crisis of 2008. The authors highlight the role of financial frictions in that life insurers were able to improve their required capital holdings by selling discounted long-term policies. In a subsequent contribution, Koijen and Yogo (2016) demonstrate that the existence of "shadow insurance", i.e. the practice of ceding large amounts of liabilities to on- and offshore reinsurers allows companies to mitigate the impact of tightening capital regulations and limit their impact on equilibrium prices and quantity of underwritten insurance. In relation to these papers, we show that such strategies of life insurance companies may be additionally motivated by an explicit desire to keep prices constant in the face of changing costs and regulatory environment.

Our model is also related to the literature on optimal delegation and self-control. These papers typically analyze a principal-agent setting with no transfers in which a biased agent is better informed. The principal has full commitment and chooses a set of actions that the agent can take. Main references for the literature on delegation include Holmstrom (1984), Melumad and Shibano (1991), Alonso and Matouschek (2008) and Amador and Bagwell (2013). Similar frameworks have been applied to the commitment versus flexibility trade-off

[3]The authors support this by citing a July 1993 issue of Consumer Reports.

for time-inconsistent agents (Amador et al., 2006), and on the optimal level of discretion for policymaking (Athey et al., 2005; Halac and Yared, 2014, 2017, 2018). Our paper differs from the literature in three distinct ways. First, in our model, the time inconsistency of the firm is endogenous. The firm is able to decrease or even eliminate its intertemporal conflict, but we show quantitatively that it will not do so under empirically relevant parameters. Second, the optimal premium will always contain a discontinuous jump if the firm has discretion in adjusting the premiums in the future. This is in contrast to previous literature which has found conditions for interval delegation to be optimal. Third, our paper focuses on an industrial organization setting and emphasizes its quantitative importance in explaining a previously undocumented phenomenon.

The remainder of the paper is structured as follows. Section 2 describes the construction of our dataset and summarizes the main findings about price rigidity in the life insurance market. Section 3 develops the theoretical model. In Section 4 we present the main qualitative predictions of the model, calibrate it and perform a numerical analysis of the solution. Section 5 empirically tests the main predictions of the model and addresses common alternative theories. Section 6 concludes and discusses the broader implications of our theory. The Appendices contain the proofs to theorems and lemmas presented in the main text, a description of the numerical algorithm, and some more nuanced discussions of our data.

2 Life Insurance Prices

In this section we describe the empirical setting of our paper. We start by explaining how renewable level-term insurance works, introduce the dataset of historical premiums, and then discuss our findings on premium rigidity. We conclude by showing that marginal cost of life insurance is volatile over time, which presents a puzzle in light of the rigid premiums.

2.1 Contract Description

We focus our attention on the renewable level-term form of insurance. These contracts require a down payment of yearly premium at the moment of signing and stay in force for a pre-defined period, typically between one and twenty years. After the term expires, customers face a premium schedule that increases with age and are allowed to renew the policy without undergoing a medical reclassification. Table 1 presents the structure of a typical level-term insurance policy, commonly referred to as the Annual Renewable Term (ART), for the first 10 policy years. In order to help the consumers undertake this long-term commitment, the contract stipulates a projected path of premiums based on the rates

currently offered to older individuals in the same health category (the "Non-Guaranteed Current" column). This schedule is not binding though, and the company may change it at any point in the future. From a legal standpoint, the insurer only commits to an upper bound on future premiums (the "Guaranteed Maximum" column), which vastly exceeds the amounts that can be expected in a market equilibrium.

A natural question to ask is: how often do life insurance companies change their premium schedules? The next section answers this question by constructing a dataset of historical premiums to verify that companies indeed tend to honor these non-binding commitments.

Table 1: Structure of an Annual Renewable Term (ART) contract

Age	Guaranteed Maximum Contract Premium	Non-Guaranteed Current Contract Premium
30	270.00	270.00*
31	550.00	280.00*
32	565.00	285.00*
33	582.50	297.50*
34	605.00	302.50*
35	632.50	325.00*
36	670.00	330.00*
37	712.50	332.50*
38	757.50	350.00*
39	820.00	360.00*

Note: Sample contract offered by the Guardian Life Insurance Company of America (first ten years). Face value = $\$250,000$. The asterisk in the last column is a standard feature and indicates that premiums are non-guaranteed. Source: Compulife Software, December 2004.

It is important to highlight that Table 1 presents a type of contract referred to as Aggregate ART where the premiums vary exclusively by age. Similarly as in Hendel and Lizzeri (2003), our data also contains the Select and Ultimate (S&U) ARTs where prices are contingent on the time elapsed since the last medical exam. For example, a 31-year-old new S&U policyholder would be offered a lower premium than a renewing customer of the same age, unless the former undergoes an new medical examination. For completeness, our dataset pools together both types of level-term contracts.

2.2 Data Construction

We construct a sample of life insurance premiums from Compulife Software, a commercial quotation system used by insurance agents. The programs are released monthly, spanning the period from May 1990 until October 2013. For each of the 282 months collected, we

recover the premiums for 1-, 5-, 10- and 20-year renewable term policies offered by different companies.[4] Even though Compulife is not a complete dataset, it covers most of the major life insurers with an A.M. Best rating of at least A-. As the default customer profile we use a 30-year-old male, non-smoker, with the "regular" health category, purchasing a policy with face value of $250,000, in California. The choice of this particular state is by Compulife's recommendation, due to a relatively large population and wide representation of insurance companies. The obtained sample consists of 55,829 observations on annual premiums for 578 different policies offered by 234 insurance companies.[5] Naturally, over the course of 23 years these firms tend to disappear or merge, as well as discontinue their old products and launch new ones. For this reason, even though we keep track of such transformations whenever possible, each product is observed on average for just over 96 months (with a median of 84). We also eliminate the seemingly duplicate products offered by the same company, always keeping the one with lower price.

2.3 Historical Premiums

Table 2 provides a statistical description of price rigidity in our dataset. Among 578 distinct insurance products that appear for at least 12 continuous months in the sample, only 369 change their premium ever. The probability of a change in any month is 2.56%, resulting in an average premium duration of roughly 39 months. Table 2 includes a vast number of companies that do not adjust prices even once. This may be a deliberate business strategy, but it may also result from other, non-market-oriented factors.[6] Hence, we also calculate the statistics for the subsample of insurance policies that display at least one premium change. Among those products the probability of a monthly price adjustment increases slightly, but still remains low at 3.4%, resulting in an average duration of almost 28 months. It should also be noted that products whose price adjusts in the dataset, also tend to stay around for a longer time (114 months as opposed to an unconditional average of 96 months). This observation suggests that certain insurance policies tend to be discontinued (and replaced by new ones) rather than to deviate from the previously promised premium schedule.

We observe a total of 1432 instances of premium adjustment, consisting of 580 hikes and 852 drops. Whenever they occur, premium changes tend to be of large magnitude, over 10%

[4]To extract the data from Compulife programs, we obtain screenshots with premium listings and apply a dedicated optical character recognition (OCR) script to convert them into numeric data. This approach is particularly useful for the pre-1997 programs which can only be run under MS-DOS operating system.

[5]Because of occasional incompleteness of Compulife data (especially in the 1990s), we impute the prices whenever a discontinuity appears for up to at most 12 months. The imputed data constitutes roughly 1% of the final sample size. We also drop all the products that are observed for less than 12 continuous months.

[6]For instance, an insurance company that has no interest in selling certain types of policies may still offer them as a reference for tax authority.

on average, although we also observe many small changes which yields a median change of around 8%. Overall, the size distribution of price changes in the life insurance market looks similar to the broad class of consumer products, as documented by Klenow and Kryvtsov (2008).[7] In particular, the distribution is leptokurtic (kurtosis is around 5.5) meaning that more mass is concentrated around zero and in the tails relative to a normal distribution (kurtosis = 3). However, it is unclear how this distribution compares to other financial services, because very little evidence is available in the literature.

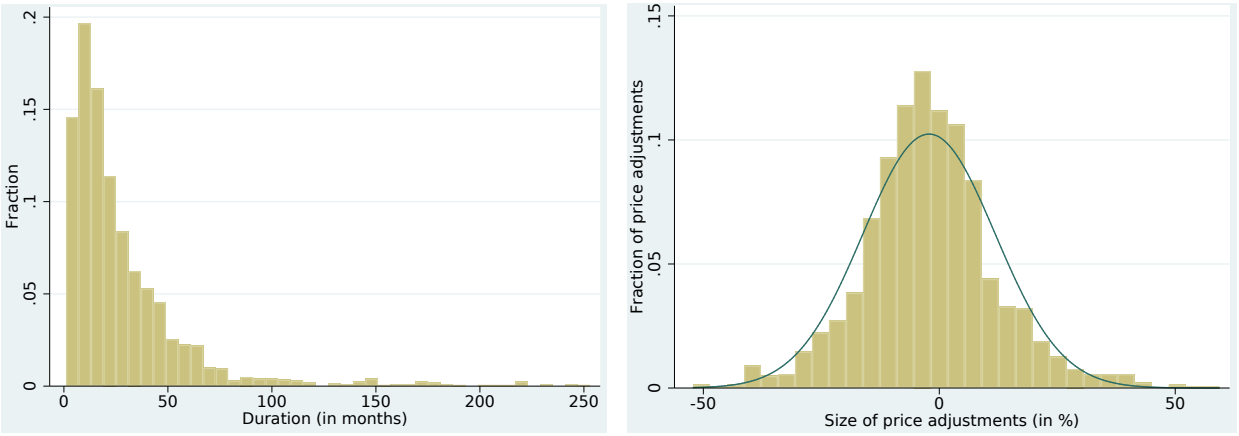
Table 2: Price rigidity in the sample

Total number of observations			55,829
Total number of insurance products observed			578
Total number of products that change price			369
Whole sample:			
Probability of a monthly price change (in %)			2.56
Median probability of a price change (in %)			1.73
Average number of observations per product			96.59
Excluding the companies that never adjust:			
Probability of a monthly price change (in %)			3.41
Median probability of a price change (in %)			3.23
Average number of observations per product			113.82
Total number of premium adjustments:			
All premium changes			1432
Premium hikes only			580
Premium drops only			852
Distribution of adjustment size (in %):	**Average**	**Median**	**Kurtosis**
All premium changes	10.74	7.94	5.56
Premium hikes only	10.58	7.41	5.99
Premium drops only	10.85	8.49	5.14

In order to visualize these findings, Figures 2(a) and 2(b) take a closer look at the distribution of premium durations and adjustment sizes. The former depicts a standard view of a distribution of durations with significant positive skewness and a long right tail reaching up to 20 years! Each bin in the histogram represents 6 months, which means that roughly

[7]Naturally, life insurance is different from typical CPI basket goods in that it provides a nominal face value rather than a real consumption value. Hence, inflation provides at best a second-order pressure on premium changes (via the company's strategic behavior), as we show in more detail in Appendix A. Hence, all else constant it should not be surprising that life insurance premiums are rigid even in the presence of positive inflation. On the other hand, as we demonstrate in Section 2.4, life insurance products exhibit volatile cost shocks which is not necessarily the case for many goods included in the CPI basket.

35% of premiums spell last up to 12 months, while the majority last longer than a year. The second chart presents the distribution of relative sizes of price adjustments, together with a fitted normal density plot. As it is clear from the summary statistics in Table 2, premium drops occur more often and are of slightly larger magnitude. The adjustments reach as much as 50% in both directions. The distribution also exhibits fatter tails and more concentration around zero than the normal one.



(a) Histogram of premium durations

(b) Histogram of adjustment sizes

Figure 2: Distribution of premium durations and adjustment sizes

In a final piece of data analysis, we explore the distribution in insurance premiums in our sample by examining the relative price dispersion. Figure 3 sketches a histogram of all premiums relative to the current monthly average (for a given renewable term), which is normalized to 100. The striking feature of the graph is the long right tail which implies that some life insurance policies are offered at a premium 2.5 times as high as the average in that category, at a given point in time. More generally, even though life insurance may seem to be a rather homogeneous financial product, we observe a significant dispersion across policies. This may be attributed to varying terms and conditions of different policies (we aggregate all products in the category "renewable level-term" by term duration), as well as the imperfectly competitive environment in which life insurance companies operate. These imperfections may include search frictions (Hortacsu and Syverson, 2004), information frictions or product differentiation (e.g. with respect to company reputation or brand loyalty).

2.4 Marginal Cost Estimation

The rigidity of premiums over time may not appear puzzling unless we have an understanding of the dynamics of the underlying marginal cost of life insurance. In this section we

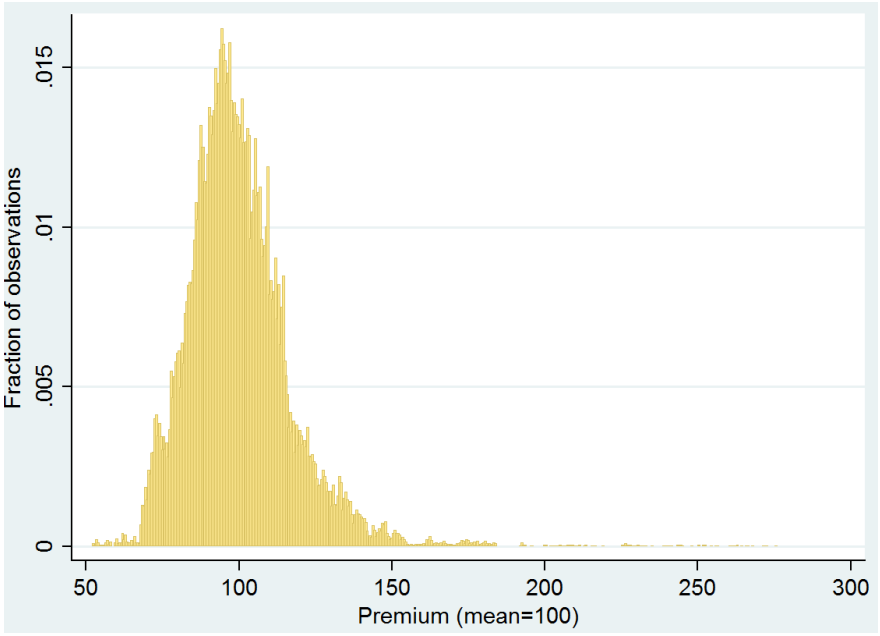


Figure 3: Distribution of insurance premiums, relative to the cross-sectional average

show how this cost can be constructed using publicly available data, specifically the interest and mortality rates. We approximate marginal cost by calculating the actuarially fair value of a renewable level-term policy (also referred to as the *net premium*). A precise description of the method we use is provided in Appendix D. Intuitively, a net premium can be thought of as a price that satisfies a zero-profit condition faced by the insurance company. Figure 4 presents a stylized illustration of an insurer's cash flow structure. A k-year renewable level-term insurance policy is effective from the moment the first premium is paid, period t, and stays in force for as long as the customer keeps renewing it. Premiums are increasing with age every k years and the benefit is paid out by the company at the moment of death of the insured, denoted $t+n$. In order to break-even, the firm must charge a stream of premiums $\{P_t\}$ such that the present expected value of cash flows between the company and the policyholder are equalized, i.e. $\sum_{s=0}^{N} \frac{E_t[P_{t+s}]}{R_t^s} = \sum_{s=0}^{N} \frac{E_t[B_{t+s}]}{R_t^s}$.

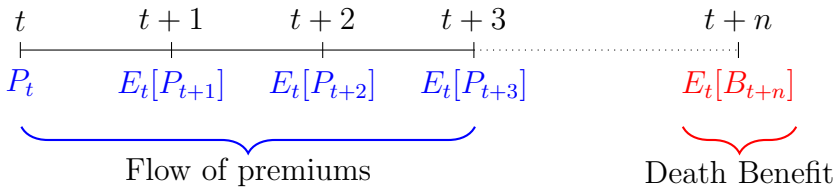


Figure 4: Stylized illustration of a life insurer's expected cash flows

Figure 5 plots the evolution of net premium for an ART policy, from May 1990 until October 2013.[8] It ranges from as low as \$196 (in November 1994) up to \$291 (in December 2008), with mean of \$216 and a standard deviation of \$13.6. The net premium exhibits considerable fluctuations over time that depend on the high frequency movements in the interest and low frequency movements in mortality rates. A slight upward trend can be noticed throughout the sample, which results from the two opposing empirical patterns - a decline in interest rates, and a decline in mortality of the insured. In particular, the net premium exhibits a sharp spike in December 2008 when interest rates plunged to record low, and a similarly high level in the post-2011 period of the zero lower bound.

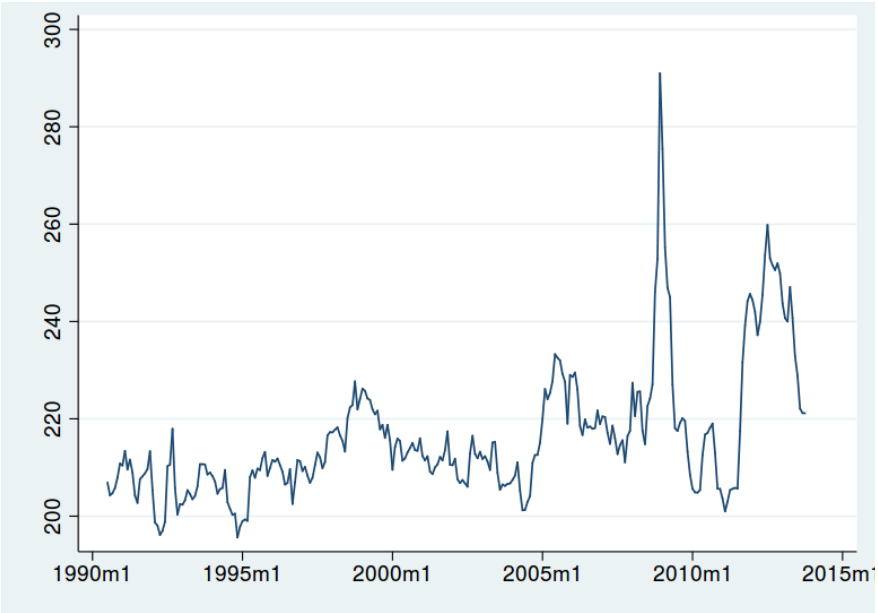


Figure 5: Net premium for an Annual Renewable Term policy over time

2.5 Data Interpretation

So far we have documented that life insurance premiums tend to be rigid over time and exhibit infrequent jumps, while the marginal cost of issuing policies is volatile. Here, we explain how these findings motivate the construction of our model in the next section.

Our model follows a generation of aging policyholders who face renewal decisions in the future. To see how the model ties into the data, denote the life insurance premium as function $P(t, a, x)$, where t is time, a is age, and x is the initial profile of a customer which we take as given. We focus on the profile of premiums for a fixed age $P(\cdot, 30, x)$ and we find that premiums $P(t, 30, x)$ are rigid with respect to t. By contrast, Hendel and Lizzeri (2003)

[8] We assume here that the customer does not voluntarily lapse before turning 60 years of age.

focus on the age profile of premiums at a fixed point in time, $P(t_0, \cdot, x)$. In reality, as time passes the policyholder also ages which makes arguments t and a move one-to-one. So for an Aggregate level-term policy, where premiums vary in age but not across cohorts, rigidity with respect to time implies that the current non-guaranteed rates in Table 1 are likely the premiums posted upon renewal. For example, if a premium $P(t, 31, x)$ is constant for all t, then a customer renewing at age 31 in 2011 pays the same premium as he anticipated at age 30 in 2010. In other words, the company keeps its promises towards the renewing customers. On the other hand, in a Select and Ultimate contract this does not have to be the case because customers of the same age may pay different premiums depending on when they had their last medical exam. While extracting the exact renewal premiums for such contracts from Compulife is technologically formidable, we verify that adjustments to the renewal schedules tend to coincide with changes to the newcomers' schedules.

A final remark is in order. The premium amounts we present here are nominal, while our model in the following section is formulated in real terms. As we explain in Appendix A, this is without loss of generality as long as inflation is constant (which is approximately true for the analyzed period of time in the United States). This is because while inflation erodes the value of premiums over time, it does so to the expected death benefit as well. Thus, nominally rigid premiums for a policy with fixed nominal face value translate into rigid real premiums *per dollar of real face value*.

3 The Model

In this section, we present a dynamic pricing model of renewable life insurance. After setting up the model, we characterize consumer demand, define incentive compatible premiums and the firm's optimization problem.

3.1 The Setup

3.1.1 Consumers and Preferences

We consider an economy consisting of overlapping generations of three-period-lived consumers. The economy operates in discrete time, $t = 0, 1, 2, \ldots$. At each date t, there is a continuum of consumers with demand for insurance, where a unit of them are young and the rest are old. We refer to the young born in t as consumers of generation t. For each generation t, the young decide whether to purchase insurance at t and whether to renew, lapse or search for a new policy when old at $t + 1$. The life insurance market does not exist for generation t consumers at $t + 2$, because they are dead in $t + 3$ and beyond.

We assume that all young consumers are of the same health category. Before becoming old, consumers face a population-average mortality risk $m_y \in (0,1)$. Additionally, before the final period of their life, a consumer's individual mortality rate is affected by a possible change in health status. We denote the population-average mortality risk of the old as $m_o \in (0,1)$.

Consumers are heterogeneous in their private reservation price to owning a policy, r_o, of a life insurance coverage when old, while it is assumed to be the same across all consumers when young: $r_y = r$.[9] Private valuation r_o is drawn from a continuous and differentiable distribution $h(r_o)$ and c.d.f. $H(r_o)$ with support $[\underline{R}, \overline{R}]$. We assume $\overline{R} > \underline{R} \geq 0$ and $\overline{R}$ is sufficiently large so there is positive demand for insurance coverage when old even if the firm is facing large cost shocks. We also assume that the hazard rate is non-decreasing for H. Only the distribution of valuations is common knowledge, so life insurance companies are unable to write individual-specific contracts. Consumers also have discount factor $\delta \in (0,1)$. We normalize the value of not owning life insurance to 0.[10]

3.1.2 Life Insurance Company

We model the pricing decision of a single life insurance company that faces exogenous competition in the form of stochastic outside options available to consumers. The market structure can be interpreted as monopolistic competition where the company faces downward-sloping demand due to the imperfect substitutability of insurance policies.[11]

The company faces a stochastic marginal cost shock c_o for insuring the old, which is randomly drawn from a continuous and differentiable c.d.f. G and p.d.f. g with support $[\underline{c}, \overline{c}]$.[12] The marginal cost for each date $c_{j,t}$, where $j \in \{y, o\}$ depends on the aggregate mortality rate and the interest rate. Formally, this relationship can be expressed as

$$c_{j,t} = \frac{m_{j,t}}{1 + i_t},$$

where i_t is the one-period risk-free interest rate. We do not take a stand on the distributions of $m_{j,t}$ and i_t, and instead we only model explicitly the univariate distribution of $c_{j,t}$. Also, since $m_{y,t}$ is small, we assume that $c_{y,t} = c_y$ and $m_{y,t} = m_y$ for all t.[13]

A key assumption is that the cost is privately observed by the insurance company. This

[9] The main results of the paper are unchanged if r_y is heterogeneous.

[10] It is important to note that since consumers have a positive valuation for having coverage, they are not risk neutral.

[11] This is due to the search and information frictions, as described by Hortacsu and Syverson (2004) and confirmed by the premium dispersion in our dataset. Section 3.1.3 introduces this assumption in more detail.

[12] The insured pool is large, so G does not vary with the size of the insured pool on the margin.

[13] The cost to insure the young is relatively stable in the data, for details see Figure 8 in Section 4.1.

assumption is reasonable given the complicated nature of the mapping from interest and mortality rates to marginal cost of renewable life insurance contracts, as well as limited availability of the data in real time.[14]

3.1.3 Renewability, Transaction Cost and Search Frictions

There are two types of policies offered in the market: renewable and non-renewable, each with the same face value which we take as given. The young may purchase one of these products, or neither of them. If they choose a renewable policy, then they have an option to renew when old regardless of the possible changes in their health status. On the other hand, a non-renewable policy expires after one period and consumers may purchase another non-renewable insurance bearing the risk of being reclassified to a different health group. In both cases, consumers can lapse after the first period (i.e. drop coverage altogether).

Prior to acquiring a new policy, consumers need to invest a transaction cost $\mu > 0$. It captures the cost of researching the market for available products, attending medical checkups, meeting with sales agents and answering detailed questionnaires, as well as being exposed to the contestability period.[15] If young consumers decide to purchase a non-renewable policy, then they must pay the transaction cost again to receive new coverage when old. On the other hand, this cost is avoided if consumers decide to extend their renewable policy.

Old consumers may choose to search for a new non-renewable policy, at the expense of paying the transaction cost μ again. There are two sources of risk associated with this action. First, as discussed above, consumers face a possibility of health deterioration which can result in much higher premium when purchasing non-renewable insurance.[16] The second source of risk comes from the search and information imperfections, resulting in ample dispersion of the premiums offered in the market. This means that a consumer who decides to search may end up finding a worse alternative, even if the health status is unchanged. To model these risks, the price of non-renewable insurance for the generation t old consumers is $P^{NR}_{o,t+1} = \epsilon c_{o,t+1}$, where ϵ is the uncertainty added to the marginal cost and it follows a right-skewed distribution with c.d.f. Z and p.d.f. z and support $(0, \infty)$. As a result, old consumers face a trade-off in deciding whether to search for a new policy or renewing with the current company. Their eventual choice depends on the relative size of the transaction cost, their private valuation of insurance, as well as the risk of searching for a new policy. Since c_y is

[14] For example, the mortality rates are limited to the insured pool and are privately collected by and shared among insurance companies. Furthermore, a report on life insurance cost disclosure to the Federal Trade Commission argued that most consumers are unable to evaluate the real cost of insurance (Bureau of Consumer Protection and Bureau of Economics, 1979).

[15] We refer the reader to the quantitative part in Section 4.1.1 for more details on the transaction cost.

[16] We will abstract from selection issues and assume that the non-renewable premium is independent of the policyholder's valuation.

constant, we assume that the non-renewable premiums for the generation t young consumers is constant across time, $P_{y,t}^{NR} = P_y^{NR}$.

3.1.4 Timing

We express the premium schedule for generation t old consumers as $\{P_{o,t+1}(c_{o,t+1})\}$ and the realization of the premium at cost $c_{o,t+1}$ as $P_{o,t+1}(c_{o,t+1})$. The renewable life insurance contract is defined as the premium schedule $\{P_{y,t}, \{P_{o,t+1}(c_{o,t+1})\}\}_{t=0}^{\infty}$.

At each t, the company announces $P_{y,t}$. Young consumers proceed to make their investment and purchasing decision. Prior to $t+1$, the company announces the schedule $\{P_{o,t+1}(c_{o,t+1})\}$. At $t+1$, $c_{o,t+1}$ is realized and the company sets $P_{o,t+1}$, which all surviving consumers observe. Then, the existing policyholders decide whether to renew, lapse or search for a new offer. Figure 6 summarizes the timing for generation t.

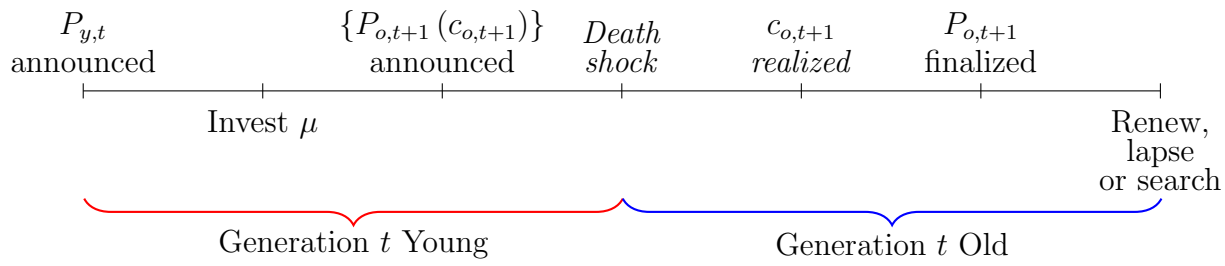


Figure 6: Timing of events

Even though $\{P_{o,t+1}(c_{o,t+1})\}$ is not announced at the beginning of t, it is common knowledge that it is selected optimally to balance commitment and discretion. In other words, consumers correctly anticipate $\{P_{o,t+1}(c_{o,t+1})\}$ upon signing. More details will be provided in the following sections.

We do not consider generation t new consumers or newcomers purchasing the renewable policy in $t+1$. In reality, consumers can sign up at any age. If the newcomers invest μ before the realization of $c_{o,t+1}$, then the hold-up problem persists. To be more concrete, young consumers who did not purchase renewables can choose to invest μ for an opportunity to have coverage when old. After μ is sunk, the cost is realized and the insurance company finalizes their premium. After observing the premium, the newcomers can choose to sign the contract or not. Under this timing, there exists a minimal valuation of the newcomers who invested μ. Since the firm has monopoly power, it has the temptation to increase premiums up to this minimal valuation once μ is sunk. Therefore, the insurance company is time-inconsistent and it can be shown that the optimal premium satisfies the same properties as the model without newcomers.

3.2 Characterizing the Demand

For the analysis in this section, we focus on consumers of generation t. Let $B^{ren}(r_o;t)$ denote the expected utility of a generation t renewable policyholder with private valuation r_o, and $B^{non}(r_o;t)$ denote the expected utility of individuals with the same profile who did not purchase a renewable policy. For any $\{P_{o,t+1}(c_{o,t+1})\}$, a generation t consumer with valuation r_o will purchase renewable life insurance in t if

$$(r - P_{y,t} - \mu) + (1 - m_y)\,\delta B^{ren}(r_o;t) \geq \max\left\{0, r - P_y^{NR} - \mu\right\} + (1 - m_y)\,\delta B^{non}(r_o;t), \quad (1)$$

where $B^{ren}(r_o;t)$ is

$$\int_{\underline{c}}^{\overline{c}} \max\left\{0, r_o - P_{o,t+1}(c_{o,t+1}), \int_{\epsilon} \max\left\{0, r_o - P_{o,t+1}(c_{o,t+1}), r_o - P_{o,t+1}^{NR}\right\} dZ(\epsilon) - \mu\right\} dG(c_{o,t+1})$$

and

$$B^{non}(r_o;t) = \int_{\underline{c}}^{\overline{c}} \max\left\{0, \int_{\epsilon} \max\left\{0, r_o - P_{o,t+1}^{NR}\right\} dZ(\epsilon) - \mu\right\} dG(c_{o,t+1}).$$

There are three benefits to purchasing a renewable policy, all reflected in the left-hand side of (1). First, renewable policyholders do not need to incur a transaction cost of μ if they choose to renew. Second, renewable life insurance contracts are similar to options. Due to one-sided commitment, policyholders are not obligated to renew if premiums are high. They could instead lapse or search for a new policy. Finally, policyholders can always renew if they are unable to find better deals after searching. The right-hand side of (1) captures the expected utility of not purchasing a renewable. In this case, individuals can choose to either not have coverage or be covered by a non-renewable when young. Also, when they are old, they have the same options as renewable policyholders except for the option of renewing. We will make the following assumption, so all consumers would have coverage when young.[17]

Assumption 1 $r \geq P_y^{NR} + \mu$.

By Assumption 1, a generation t young consumer purchases renewable if and only if

$$B^{ren}(r_o;t) - B^{non}(r_o;t) \geq \frac{P_{y,t} - P_y^{NR}}{(1 - m_y)\,\delta}. \quad (2)$$

Inequality (2) states that consumers purchase renewable insurance if and only if the expected

[17] This simplifies but does not change our analysis, because as long as μ is sufficiently large the hold-up problem persists.

difference in benefits between a renewable and a non-renewable is greater than the premium difference of the two types of insurance. The next lemma helps characterize the demand.

Lemma 1 *For any t, $B^{ren}(r_o;t)$ and $B^{non}(r_o;t)$ have the following properties: (i.) $B^{ren}(r_o;t) \geq B^{non}(r_o;t) \geq 0$ for all r_o, (ii.) $B^{ren}(r_o;t)$ and $B^{non}(r_o;t)$ are weakly increasing in r_o, (iii.) there exists sufficiently large $\tilde{r}_o$ such that $B^{ren}(r_o;t)$ and $B^{non}(r_o;t)$ are strictly increasing in r_o for any $r_o \geq \tilde{r}_o$. Also, if there exists $\hat{r}_o$ such that $B^{ren}(\hat{r}_o;t) > B^{non}(\hat{r}_o;t)$, then $B^{ren}(r_o;t) - B^{non}(r_o;t)$ is strictly increasing in $r_o \geq \hat{r}_o$.*

Lemma 1 states that if there are consumers who prefer renewables, then all consumers with higher r_o would also prefer renewables. If $P_{y,t} \leq P_y^{NR}$, then (2) is automatically satisfied. In particular, the demand for renewables is independent of r_o, because consumers purchase renewables as long as $r \geq P_{y,t} + \mu$ regardless of r_o. As a result, not all renewable policyholders expect to renew next period. Hence, there is no hold-up problem if $P_{y,t} \leq P_y^{NR}$, because the renewal demand is elastic with respect to premium changes.

The more interesting case is when $P_{y,t} > P_y^{NR}$, because all renewable policyholders expect to renew. This means only consumers with sufficiently large r_o would purchase renewables (by Lemma 1) and there is a lower bound in r_o for its pool of policyholders (by (2)). The firm would then be tempted to increase $P_{o,t+1}$ to this lower bound. This is because the renewal demand in $t+1$ is inelastic with respect to premium changes below this lower bound. The firm encounters a hold-up problem when $P_{y,t} > P_y^{NR}$. We will show numerically that despite this, the firm will set $P_{y,t} > P_y^{NR}$ at the optimum.

More formally, let $\bar{r}_o(P_{y,t}, \{P_{o,t+1}(c_{o,t+1})\})$ be defined as the threshold valuation such that (2) holds with equality. Notice that this threshold is increasing with respect to both arguments. To streamline notation, we write $\bar{r}_{o,t}$ with the implicit understanding that it depends on the premium schedules of both periods. Lemma 1 shows that all consumers with $r_o \geq \bar{r}_{o,t}$ purchase the renewable insurance. The demand function of generation t young consumers is

$$D_{y,t}(P_{y,t}, \{P_{o,t+1}(c_{o,t+1})\}) = 1 - H(\bar{r}_{o,t}). \tag{3}$$

The demand for renewing is

$$D_{o,t+1}(P_{y,t}, P_{o,t+1}(c_{o,t+1})) = (1-m_y)[1 - H(\max\{\bar{r}_{o,t}, P_{o,t+1}(c_{o,t+1})\})]. \tag{4}$$

The demand is weakly decreasing in premiums. Most importantly, renewal demand becomes perfectly inelastic for any $P_{o,t+1}(c_{o,t+1}) \leq \bar{r}_{o,t}$.

3.3 Incentive Compatibility

Before a young consumer buys a renewable policy, the renewal demand is downward sloping for all premiums. We refer to this pre-investment demand as the *ex-ante* demand and the demand following the signing of the contract as the *ex-post* demand. The *ex-ante* and *ex-post* demands are different which creates a time inconsistency problem for the insurance company. In particular, policyholders reveal their valuation is at least as large as $\bar{r}_{o,t}$ by purchasing. After consumers sign, the insurance company has an incentive to increase next period premiums up to $\bar{r}_{o,t}$, because the renewal demand is inelastic for prices below it.

The disparity between the *ex-ante* and the *ex-post* demand is the reason why the insurance company needs commitment. However, volatile cost shocks create an incentive to adjust premiums accordingly. To resolve this tension, the insurance company disciplines its pricing behavior by setting incentive compatible premium schedules.

For generation t old policyholders, we can define the following cost regions for $c_{o,t+1}$:

$$\overline{\mathcal{C}}_{o,t+1} = \{c_{o,t+1} \mid P_{o,t+1}(c_{o,t+1}) \geq \bar{r}_{o,t}\},$$
$$\underline{\mathcal{C}}_{o,t+1} = \{c_{o,t+1} \mid P_{o,t+1}(c_{o,t+1}) < \bar{r}_{o,t}\}.$$

It is worth pointing out that the dissonance between ex-ante and ex-post demands occurs for $c_{o,t+1} \in \underline{\mathcal{C}}_{o,t+1}$, while both demands are in agreement for $c_{o,t+1} \in \overline{\mathcal{C}}_{o,t+1}$.[18]

We require $\{P_{o,t+1}(c_{o,t+1})\}$ to be incentive compatible: for all t and all $c_{o,t+1}, c'_{o,t+1} \in [\underline{c}, \overline{c}]$,

$$[P_{o,t+1}(c_{o,t+1}) - c_{o,t+1}] D_{o,t+1}(P_{y,t}, P_{o,t+1}(c_{o,t+1})) \geq [P_{o,t+1}(c'_{o,t+1}) - c_{o,t+1}] D_{o,t+1}(P_{y,t}, P_{o,t+1}(c'_{o,t+1})). \quad (5)$$

Constraint (5) serves to reduce the possible set of renewal premiums $\{P_{o,t+1}(c_{o,t+1})\}$ the company can choose from. This is the same concept as *delegation sets* in Holmstrom (1984). The set is announced prior to $t+1$. Though the set of premiums is not legally binding, it is implicitly assumed that the firm has reputational concerns and believes that deviating from this set would instigate consumer retaliation which lowers its future sales.

In addition to incentive compatibility, the company needs to refrain from exploiting the consumers. The additional disciplinary measure requires the life insurance company to take $\bar{r}_{o,t}$ as given and choose $\{P_{o,t+1}(c_{o,t+1})\}$ such that (2) is satisfied for all $r_o \geq \bar{r}_{o,t}$. In essence, the company needs to deliver a minimal expected utility to the policyholders that corresponds to what the consumers expected when they were young.

[18] The analysis does not change if $\overline{\mathcal{C}}_{o,t+1} = \{c_{o,t+1} \mid P_{o,t+1}(c_{o,t+1}) > \bar{r}_{o,t}\}$ and $\underline{\mathcal{C}}_{o,t+1} = \{c_{o,t+1} \mid P_{o,t+1}(c_{o,t+1}) \leq \bar{r}_{o,t}\}$.

3.4 The Optimization Problem

From our model's setup, we can see that the insurance company maximizes the expected present value of profits by optimizing the profit for each generation, and that the optimal premium schedule is stationary. Hence, the company can maximize total expected present-valued profit by choosing the same premium schedule $(P_y, \{P_o(c_{o,t+1})\})$ for each generation t. This is because the environment is the same across generations, so the firm's problem does not change for each generation. The analysis will thus focus on the company maximizing its profit from a single generation. To simplify notation, we drop the date subscripts.

We consider a *sequentially optimal pricing rule*. The company chooses a premium schedule in each period that maximizes the present value of discounted profits taking into account that it will do the same in the future. Moreover, consumers purchase insurance taking the future behavior of the company into consideration. The endogeneity of consumer behavior differs from the concept of sequential optimality in Halac and Yared (2014).

Definition 1 *The sequentially optimal pricing rule is a contract* $\{P_y, \{P_o(c_o)\}\}$ *such that* $\{P_o(c_o)\}$ *solves the static optimization problem:*

$$\Pi_o(P_y) \equiv \max_{\{P_o(c_o)\}} \int_{\underline{c}}^{\overline{c}} (P_o(c_o) - c_o) D_o(P_y, P_o(c_o)) dG(c_o), \tag{6}$$

subject to (2) and (5) for any given P_y *and* $\bar{r}_o$*. The premium* P_y *solves the optimization problem for a given* c_y *by taking the optimal response* $\{P_o(c_o)\}$ *as given:*

$$\Pi = \max_{P_y} (P_y - c_y) D_y(P_y, \{P_o(c_o)\}) + \frac{1}{1+i} \Pi_o(P_y). \tag{7}$$

Given P_y *and the optimal response* $\{P_o(c_o)\}$*, consumers sign-up for renewable life insurance if (2) is satisfied.*

We choose to focus on the sequentially optimal pricing rule instead of the ex-ante optimal pricing rule, where the company chooses the current and all future premium schedules at the beginning of the generation, because companies may choose to renegotiate or announce possible changes to the premiums before the following period. For example, the premium schedule in Table 1 demonstrates how, in practice, life insurance companies do not tie themselves down to a fixed pricing rule.

The objective (6) is similar to finding a balance between discretion versus rules in delegation problems. The main difference is in how this trade-off is being created. The literature on optimal delegation and self-control has focused on situations where the disagreement between principal and agent or present and future-selves is exogenous. In (7), the insurance

company can completely eliminate its time inconsistency by setting $P_y \leq P_y^{NR}$. However, we will show numerically that the trade-off in life insurance contracts is endogenously generated by the company optimally setting $P_y > P_y^{NR}$. [19]

4 The Optimal Premium Schedule

In this section, we characterize the incentive compatible premium schedule and obtain its general properties. Note that all of the results here are for $P_y > P_y^{NR}$. Later in this section, we will calibrate the model and solve for P_y. The following lemma shows that the incentive compatible premium schedule is rigid for low cost shocks.

Lemma 2 *An incentive compatible premium schedule* $\{P_o(c_o)\}$ *satisfies the following:*

i. For $c_o \in \underline{\mathcal{C}}_o$, $P_o(c_o)$ *does not vary with* c_o *and* $\underline{\mathcal{C}}_o$ *has strictly positive measure.*

ii. $P_o(c_o)$ *is weakly increasing, and there exists* c^T *such that* $\underline{\mathcal{C}}_o = \left[\underline{c}, c^T\right)$ *and* $\overline{\mathcal{C}}_o = \left[c^T, \bar{c}\right]$.

Lemma 2 shows that the premium schedule is not sensitive to cost shocks within $\underline{\mathcal{C}}_o$. It also states that cost shocks within $\underline{\mathcal{C}}_o$ are non-negligible. In addition, Lemma 2 shows that there exits a cutoff c^T such that $\underline{\mathcal{C}}_o = \left[\underline{c}, c^T\right)$ and $\overline{\mathcal{C}}_o = \left[c^T, \bar{c}\right]$, which means that premiums do not respond to small cost shocks.[20]

The rigidity is caused by the inelastic renewal demand at low prices. Since $P_y > P_y^{NR}$, this means the firm needs to promise all policyholders a strictly positive expected utility at renewal, so $\underline{\mathcal{C}}_o$ is non-negligible. However, the firm knows the minimum valuation of the insured pool from (2) and is tempted to increase premiums up to it. As a result, any variation in premiums below $\bar{r}_o$ would not be credible, because the firm would always announce the highest premium below $\bar{r}_o$.

Denote $\bar{P}_o = P_o(c_o)$ and the demand $\bar{D}_o = D_o(P_y, P_o(c_o))$ for all $c_o \in \underline{\mathcal{C}}_o$. Lemma 2 allows us to rewrite the downward deviating incentive compatibility constraints into the following single constraint: $\forall c_o \in \overline{\mathcal{C}}_o, \forall c_o' \in \underline{\mathcal{C}}_o$, $[P_o(c_o) - c_o] D_o(P_y, P_o(c_o)) \geq \left(\bar{P}_o - c_o'\right) \bar{D}_o$. Similarly, we can also rewrite the upward deviating incentive compatibility constraints. In particular, we have the following binding incentive compatibility constraint at c^T :

$$\left(\bar{P}_o - c^T\right) \bar{D}_o = \left(P_o\left(c^T\right) - c^T\right)\left[1 - H\left(P_o\left(c^T\right)\right)\right]. \tag{8}$$

[19] An additional reason for $P_y > P_y^{NR}$ not modeled in this paper is the fact that firms may attempt to screen consumers with demand in the next period from those without. To see this, suppose some consumers have $r_o = 0$ while some have $r_o > 0$. The firm separates the two types by issuing renewable and non-renewable contracts. Consumers with $r_o = 0$ prefer the non-renewables $r - P_y^{NR} - \mu \geq r - P_y - \mu + \delta\eta$, where $\eta > 0$ is the expected benefit from selling the insurance in the secondary market. This implies $P_y - P_y^{NR} \geq \delta\eta$.

[20] Alternatively, defining $\underline{\mathcal{C}}_o = \left[\underline{c}, c^T\right]$ and $\overline{\mathcal{C}}_o = \left(c^T, \bar{c}\right]$ would not affect the analysis and results.

If (8) is non-binding, then it is not incentive compatible for c_o within a neighborhood of c^T.

The next lemma characterizes the incentive compatible premium for $\overline{\mathcal{C}}_o$, which shows that premiums are also rigid for intermediate cost shocks. Let $P_o^*(c_o)$ denote the optimal *frictionless premium* at c_o. We refer to it as the frictionless premium because the incentive compatibility constraints are not binding.

Lemma 3 *An incentive compatible premium schedule* $\{P_o(c_o)\}$ *satisfies the following:*

i. For $c_o \in \overline{\mathcal{C}}_o$, *if* $P_o(c_o)$ *is strictly increasing and continuous on an open interval* (c_o', c_o''), *then* $P_o(c_o) = P_o^*(c_o)$ *on* (c_o', c_o'').

ii. $P_o\left(c^T\right) > \bar{r}_o > \bar{P}_o$, *and there exists* $c^M > c^T$ *such that* $P_o(c_o)$ *does not vary with* $c_o \in \left[c^T, \min\left\{c^M, \overline{c}\right\}\right)$.

Lemma 3 shows that the company charges the optimal frictionless price if it has full flexibility, but it is not incentive compatible for the firm to have full flexibility for all costs in $\overline{\mathcal{C}}_o$. Most importantly, $\{P_o(c_o)\}$ has a jump discontinuity at c^T.

By Lemma 2 and the fact that $\bar{r}_o > \bar{P}_o$, the incentive compatible premium has to have a discrete jump if the company has flexibility to adjust premiums. What is interesting is the size of this jump. The incentive compatible premium has a discrete jump at c^T such that $P_o\left(c^T\right) > \bar{r}_o$, because the company has to induce enough lapsation to convince the consumers that an upward adjustment in premiums is due to adverse cost shocks and not an attempt to take advantage of the hold-up problem. The company can credibly respond to cost shocks only when it causes a simultaneous decrease in the demand for insurance.

The existence of a discontinuous jump is in contrast to the optimal delegation literature, where general conditions can be found to rule it out (Amador and Bagwell, 2013). The difference is the endogeneity of $\bar{r}_o$. To see why, suppose the admissible premium is within the range of $[A, B]$, then only consumers with $r_o > B$ would purchase renewables.[21] Therefore, $\bar{r}_o > B$ and the company would always charge a price of B. This implies that incentive compatible adjustments to premiums have to be significant and contain jumps.

Lemma 3 also states that incentive compatible premiums are rigid for $c_o \in \left[c^T, \min\left\{c^M, \overline{c}\right\}\right)$. This is because if the firm chooses $P_o^*(c_o)$ for all $c_o \in \left[c^T, \overline{c}\right]$, then the firm would deviate to the frictionless premium for cost shocks slightly below c^T. As a result, the firm chooses a rigid premium, which is the only other incentive compatible option for costs within $\left[c^T, \min\left\{c^M, \overline{c}\right\}\right)$.[22] This implies the firm is able to charge frictionless premiums only

[21] Consumers with $r_o = A$ would purchase non-renewables, because expected utility for buying renewables would be zero. Also, consumers with $r_o \in (A, B]$ would also not purchase, because the company would set $P_o = \bar{r}_o$ if they are part of the insured pool making $[A, B]$ not incentive compatible.

[22] This result is similar to the characterization of incentive compatible delegation rules with discontinuities in Melumad and Shibano (1991) and Alonso and Matouschek (2008).

for sufficiently large shocks, because the corresponding premiums would be large enough to cause a significant decrease in demand through lapsation to retain incentive compatibility. Let $\bar{\bar{P}}_o$ denote the rigid premium when $c_o \in \left[c^T, \min\left\{c^M, \bar{c}\right\}\right)$.

Theorem 1 *The incentive compatible premium has the following feature:*

$$P_o\left(c_o\right) = \begin{cases} \bar{P}_o & \underline{c} \leq c_o < c^T \\ \bar{\bar{P}}_o & c^T \leq c_o < \min\left\{c^M, \bar{c}\right\} \\ P_o^*\left(c_o\right) & \min\left\{c^M, \bar{c}\right\} \leq c_o \leq \bar{c} \end{cases}$$

with $\bar{P}_o < \bar{r}_o < \bar{\bar{P}}_o = P_o^*\left(c^M\right)$ *and*

$$\left(\bar{P}_o - c^T\right)\left[1 - H\left(\bar{r}_o\right)\right] = \left(\bar{\bar{P}}_o - c^T\right)\left[1 - H\left(\bar{\bar{P}}_o\right)\right]. \tag{9}$$

Theorem 1 shows that the incentive compatible premium is rigid for low and intermediate cost shocks and has an upward discrete jump at c^T. Figure 7 illustrates an optimal incentive compatible pricing rule. By Theorem 1, the life insurance company solves for the premiums $\left\{\bar{P}_o, \bar{\bar{P}}_o, P_y\right\}$ and the cost thresholds $\left\{c^T, c^M\right\}$ subject to (2) and (9). Since the premiums affect $\bar{r}_o$ and vice versa, the optimization problem is complex and it is difficult to obtain tractable analytical solutions. In the next subsection, we will solve the model numerically.

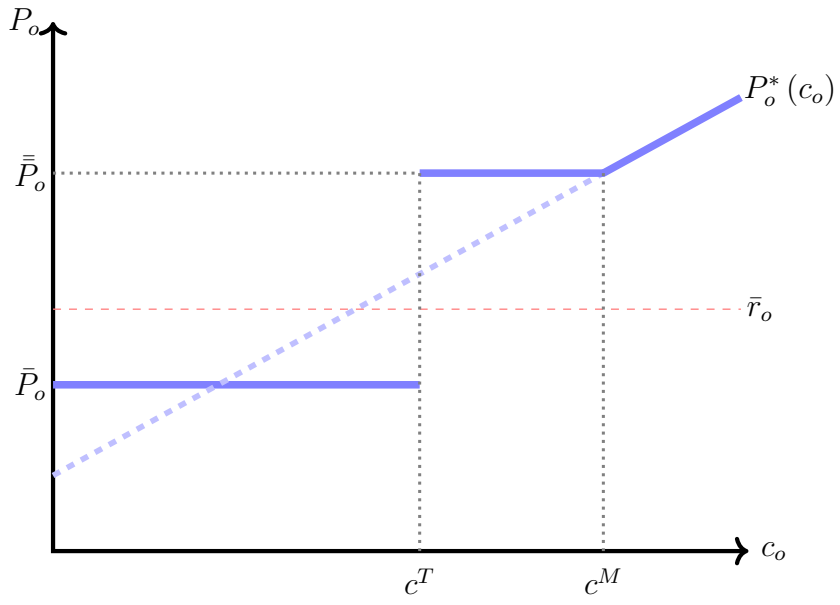


Figure 7: Incentive compatible premium profile

4.1 Numerical Analysis

4.1.1 Calibration

The model does not have an explicit closed-form solution.[23] We proceed by using the data and outside knowledge to assume reasonable parameter values and functional forms, and solve for an equilibrium numerically. Table 3 presents a summary of our calibration. Because we only focus on two periods, we assume that each period is equivalent to ten years and calibrate the remaining parameters to the features of ten-year level-term renewable insurance. While most of the model variables do not have a direct counterpart in the data, we attempt to make the calibration realistic while also keeping the numerical solution feasible. In what follows we describe our assumptions on the functional forms and parameter values.

Table 3: Parameter values in the model

Symbol	**Meaning**	**Value**
μ	Transaction cost	970
c_y	Cost shock for insuring young	1090
P_y^{NR}	Price of non-renewable insurance	1040
m_y	Mortality rate	0.007
i	Annual interest rate	0.04
δ	Discount factor	0.68
Cost shock distribution: uniform		
$\bar{c}$	Lower bound	2700
$\underline{c}$	Upper bound	3500
Valuation distribution: Generalized Pareto		
γ	Scale parameter	700
κ	Shape parameter	0
θ	Threshold parameter	3595
Search and health shock distribution: lognormal		
$\bar{\epsilon}$	Mean	-0.0104
σ_ϵ	Standard deviation	0.14425

The marginal cost to insure a consumer is computed directly from the data. Cost of covering the young c_y corresponds to the average expected death benefit payout for a 30-

[23] This is mainly because $\bar{r}_o$ affects the degree of rigidity and the size of the discrete jump, but it is also endogenously determined by the premiums.

year old male over the period of ten years.[24] This cost ranges from \$914 to \$1244 in the data, with a mean of \$1090. Because of the relatively small volatility of this variable evident in Figure 8 below, we simplify the model by taking c_y as given and setting it equal to the average. The cost of covering the old c_o is associated with the expected death benefit payout for a 40-year old male *who is renewing a policy purchased at age 30*. This involves using different (higher) mortality rates than for new 40-year-old customers who have just passed a medical exam. This cost ranges from \$2717 to \$3430 in the data, with an average of \$3064. The distribution of this shock is unlikely to be normal due to the presence of fat tails. This is confirmed by the Jarque-Bera test which returns a p-value of 0.075, providing grounds to reject the null hypothesis of normality. In order to make computation of some parts of the equilibrium analytically feasible, we assume that the distribution is uniform with bounds $[2700, 3500]$. Figure 8 illustrates our measure of the cost shocks over time.[25]

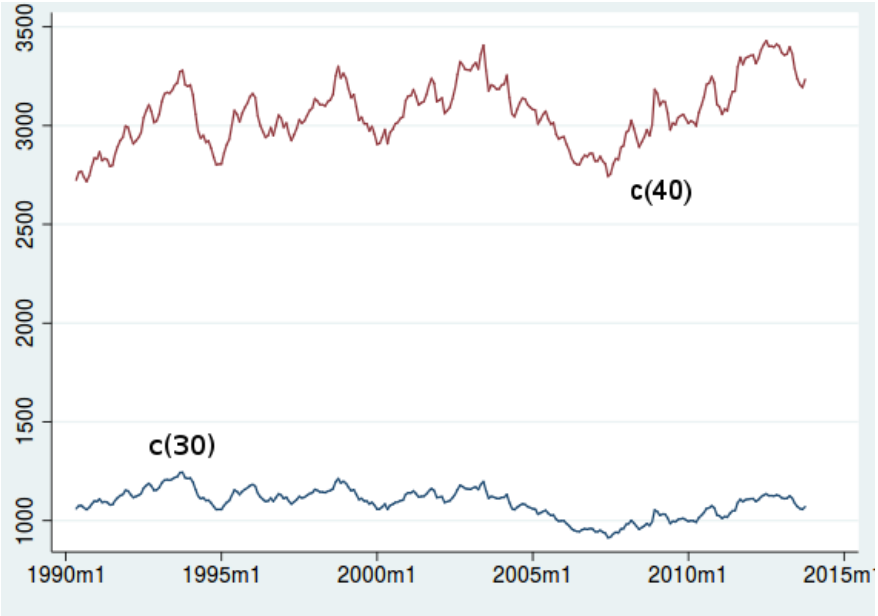


Figure 8: Cost shocks in the data

The distribution of consumers' private valuations does not have a clear counterpart in the data. To simplify the algorithm, we assume it to be a Generalized Pareto distribution, with the shape parameter of 0. This assumption makes it essentially a "shifted" exponential distribution, which features a constant inverse hazard rate $\frac{1-H(\cdot)}{h(\cdot)} = \gamma$, enabling us to obtain closed-form solutions for prices $\bar{P}_o$ and $\bar{\bar{P}}_o$ (details of the solution method are provided in

[24] As discussed in Appendix D, we abstract here from the issue of voluntary lapsation and assume that the consumer will continue to pay the premium for the entire period.

[25] In reference to the series presented in Figure 5, here we consider a fixed 10-year insurance term only and do not convert the cost shocks to annual values.

Appendix C). The scale parameter γ is selected to match the existing evidence on elasticity of demand for term life insurance. Specifically, Pauly et al. (2003) use the Compulife data from January 1997 and find the price elasticity of demand to be 0.475 for a median company. Given our distributional assumption, and the median yearly premium for an ART in January 1997 of \$335, we set the value of the scale parameter γ to be 700. The threshold parameter θ is then calibrated to match the fraction of non-renewable policies among all term policies underwritten, equal to 11% as reported by LIMRA (1994).

The distribution of the health and search shock ϵ is assumed to be lognormal, which conveniently allows us to calculate the integrals inside of $B^{ren}(\cdot)$ and $B^{non}(\cdot)$ of (1) analytically, rather than numerically. The right skewness of the distribution captures the idea that a consumer who decides to search for a new policy when old may find a better deal in the market, but is also at a risk of prohibitive premium increases should his health have deteriorated or lifestyle habits changed.[26] We do not have compelling evidence on the probabilities of getting a table rating. For this reason, we set the parameters of the lognormal distribution to match two moments: i.) $E(\epsilon) = 1$, and ii.) $Var(\epsilon) = 0.145^2$ where 0.145 is the average coefficient of variation across time for all the prices of 10-year renewable term policies in our sample. In other words, we calibrate the variance of search shocks to match the observed dispersion of premiums in the data, and allow the right-skewness of the lognormal distribution to determine the likelihood of adverse health shocks. Notice that by calibrating the distribution of ε to the empirical price dispersion, we introduce to the model a reduced-form effect of market competition (Section 5.4.3 provides more evidence on this).

The transaction (or switching) cost μ is the key parameter in our model, and at the same time probably the most controversial one. It comprises the opportunity cost of researching the products on the market (and not working or using leisure), the opportunity and monetary cost of attending the medical examination, the opportunity cost of meeting with an insurance agent and filling out the paperwork (in particular given that our sample starts in 1990s). Additional factors contributing to the dollar value of μ consist of the cognitive cost of shopping for an insurance policy and undertaking a long-term commitment (for example, overcoming cognitive dissonance), as well as getting exposure to the contestability period, i.e. a possibility that the insurance company may reject a benefit claim if death occurs in a short period after signing the contract. Direct estimation of switching costs in the life insurance market is beyond the scope of this paper. Instead, we survey the literature for similar recent estimates across other markets which also feature long-term contracts. Table 4 summarizes

[26] In practice, below the regular health category life insurance companies use the so-called table ratings to determine a premium hike. The consumer's health and lifestyle is evaluated with respect to several categories and each one may raise the standard rate by 25%.

our investigation. The switching cost estimates vary significantly for different studies and markets, ranging between \$40 and \$700 for markets such as auto insurance, wireless or cable TV, as well as between \$1200 and \$5000 for health and retirement plans. In order not to rely on possibly irrelevant outliers, we adopt a median value between these two groups of \$970, and we analyze the importance of this parameter by performing comparative statics exercises in the following section.

Table 4: Switching cost estimates in the related literature

Reference	**Market**	**Dollar value**
Honka (2014)	Auto insurance	45-190
Cullen and Shcherbakov (2010)	Wireless	255
Shcherbakov (2016)	Television and satellite	227-395
Weiergräber (2014)	Wireless	337-672
Illanes (2016)	Pension plans	1285
Miller and Yeo (2012)	Medicare	1700-1930
Handel (2013)	Health insurance	2250
Nosal (2012)	Medicare	4990

Note: Relative to the amounts quoted in original papers, we convert them to 2012 US dollars.

The remaining parameters of the model are calculated directly from the data. The mortality rate m_y is the cumulative ten-year probability of dying for the insured 30-year old male; we find it to be 0.7% using the 2001 Select and Ultimate mortality tables. The annual interest rate i is assumed to be 4%, which yields the ten-year discount factor δ of 0.68. Finally, we assume that the non-renewable insurance premium P_y^{NR} is \$1040, which implies that such policies are priced competitively and sold at a slight discount relative to the cost of renewables c_y. This assumption is useful in the model due to the fact that the firm may choose $P_y = P_y^{NR}$ to eliminate its time inconsistency problem. However, with $c_y > P_y^{NR}$, this will cause a loss in covering the young encouraging the firm to seek an interior solution instead. Empirically, the assumption that non-renewable insurance is sold at a discount relative to the ten-year marginal cost is plausible for at least two reasons. First, consumers who purchase such policies are likely to actually need it for a shorter time, resulting in higher lapsation rates. Second, the average health status of renewable policyholders at any given time tends to be worse than non-renewable policyholders. This is because all non-renewable policyholders recently had health exams, while some renewable policyholders have renewed without undergoing a health exam and have likely deteriorated in health. For example, in the data the cost to insure a pool of 30-year-olds who have held their policies for 10 years increases the marginal cost by 55%. While in reality such vintage policyholders are likely a

minority in the pool of 30-year-old customers, their existence naturally elevates the average cost c_y relative to P_y^{NR}, a price only available to the newcomers.

4.1.2 The Equilibrium

Table 5 presents the equilibrium of our model under the discussed calibration. The upper cost threshold c^M is optimally set equal to the upper bound of the cost distribution, $\overline{c}$. This means that the premiums can only be one of two values: $\bar{P}_o$ and $\bar{\bar{P}}_o$. The lower cost threshold c^T is equal to 3054.47, roughly in the middle of the cost domain. This frequency of price adjustments is reasonable given that we calibrate the model to 10-year renewable level-term insurance. The predicted rigid premiums, $\bar{P}_o$ and $\bar{\bar{P}}_o$, are equal to 3626.22 and 4074.47, respectively. As the lower panel of Table 5 shows, this is well in the ballpark of what the renewing 40-year-olds can expect to pay in the data (expressed in cumulative ten-year terms). Also, the size of the jump predicted by our model, $\left(\bar{\bar{P}}_o/\bar{P}_o - 1\right) \times 100$, matches closely the average size of the premium adjustment observed in the data, around 12%.

Table 5: Equilibrium of the model

Symbol	**Variable name**	**Value**
Π	Total profit	310.93
Π_o	Profit from old	511.21
c^T	Lower cost threshold	3054.47
c^M	Upper cost threshold	3500.00
P_y	Price for young	1050.89
$\bar{P}_o$	Lower price for old	3626.22
$\bar{r}_o$	Threshold for renewables	3669.27
$\bar{\bar{P}}_o$	Upper price for old	4074.47
$\left(\bar{\bar{P}}_o/\bar{P}_o - 1\right) \times 100$	Jump between premiums (in %)	12.36
Premiums in the data:		
40-year-old average		3549.14
40-year-old median		3489.35
40-year-old standard deviation		817.93
Average change (in %)		11.47

Note: the data section summarizes the premiums for 40-year-old males in regular health category, for 10-year renewable insurance. Annual premiums are expressed here as present expected value of the entire ten-year period, until renewal. Similarly as in section 2.4, we ignore the issue of lapsation.

4.1.3 Comparative Statics

We now analyze the mechanics of the model by performing several comparative statics exercises with respect to the key parameters. Figure 9 illustrates how the optimal pricing rule changes with transaction cost μ. Renewable life insurance contracts become more attractive compared to non-renewables as μ increases. The firm responds by increasing premiums (P_y, $\bar{P}_o$ and $\bar{\bar{P}}_o$) and restricting quantity, i.e. the pool of covered policyholders ($\bar{r}_o$ up). As a result, the total profit of the firm rises. Crucially though, an increase in the transaction cost also worsens the hold-up problem. In order to attract consumers *ex ante*, the firm responds by increasing c^T so that it is more committed to the low premium $\bar{P}_o$, and by raising the size of the jump up to the high premium $\bar{\bar{P}}_o$. Notice that beyond a certain level of μ, the terms of the optimal contract become invariant. This is due to the fact that search becomes too expensive for a vast majority of consumers and only the ones with high enough demand for coverage when old decide to buy.

Figure 10 shows a similar exercise when we vary the mode of the lognormal health and search shock distribution, while holding the mean equal to 1. Higher mode means that consumers face more adverse health shocks and become more locked-in to the contract, but it also makes renewables more desirable than non-renewables. Once again, the decrease in competition from non-renewables leads the firm to raise equilibrium premiums and restrict the quantity supplied. By the same logic as with the transaction cost, to attract more consumers *ex ante* the equilibrium cost threshold c^T goes up, promising a wider interval of rigid low-price insurance, and the jump between $\bar{\bar{P}}_o$ and $\bar{P}_o$ widens.

Finally, in Figure 11 we vary the inverse hazard rate of the distribution of consumers' valuations. As γ goes up, the price elasticity of demand decreases and the company enjoys more monopoly power resulting in a higher $\bar{P}_o$. Importantly though, lower demand elasticity does not alter the strength of the consumers' hold-up problem and as a result the equilibrium cost threshold falls, imposing high premiums over a wider range of cost shocks, while the jump between $\bar{\bar{P}}_o$ and $\bar{P}_o$ becomes smaller. The increase in profit from higher renewal premiums would be bigger with a larger pool of policyholders, which is why P_y falls as γ rises to attract more customers in the spirit of the switching cost literature (Klemperer, 1987). However, it is important to note that $\bar{\bar{P}}_o$ can be inversely related to γ. This is because the profit for smaller cost shocks ($c_o \in \underline{\mathcal{C}}_o$) has increased with higher γ, so for (9) to hold, the profit for larger cost shocks ($c_o \in \overline{\mathcal{C}}_o$) must also go up. Since $\bar{\bar{P}}_o$ is too high compared to the frictionless premium near c^T, the firm lowers $\bar{\bar{P}}_o$ to increase profits for cost shocks near and above c^T.

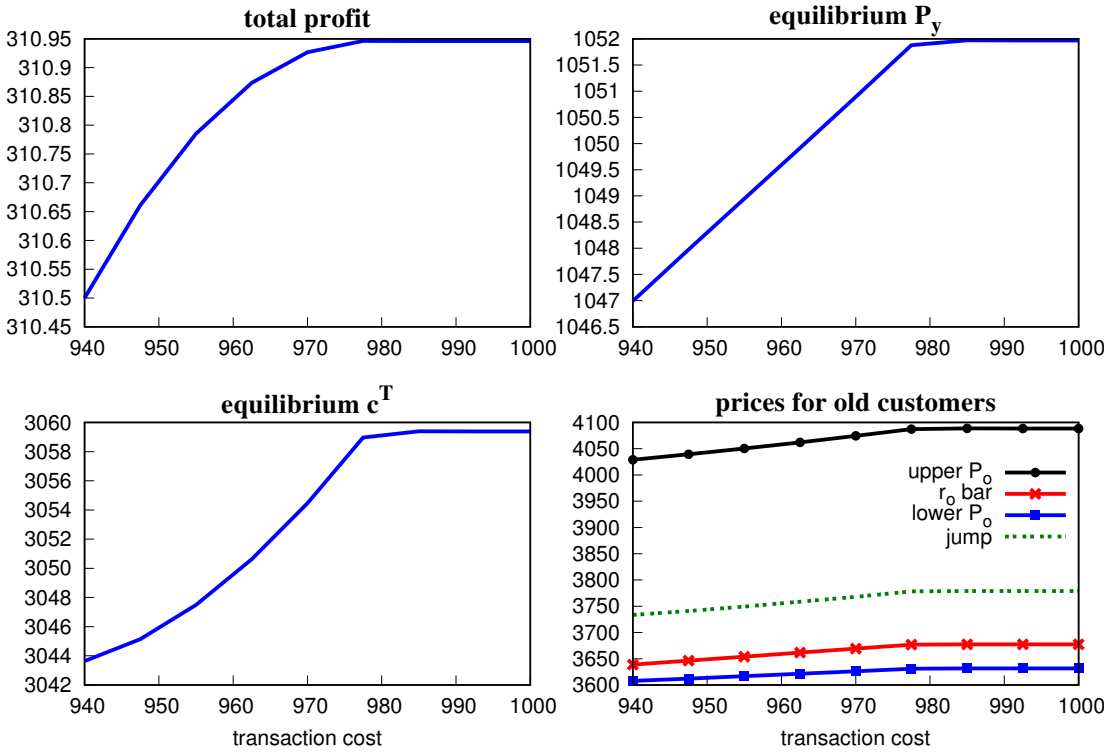


Figure 9: Varying the transaction cost μ in the model

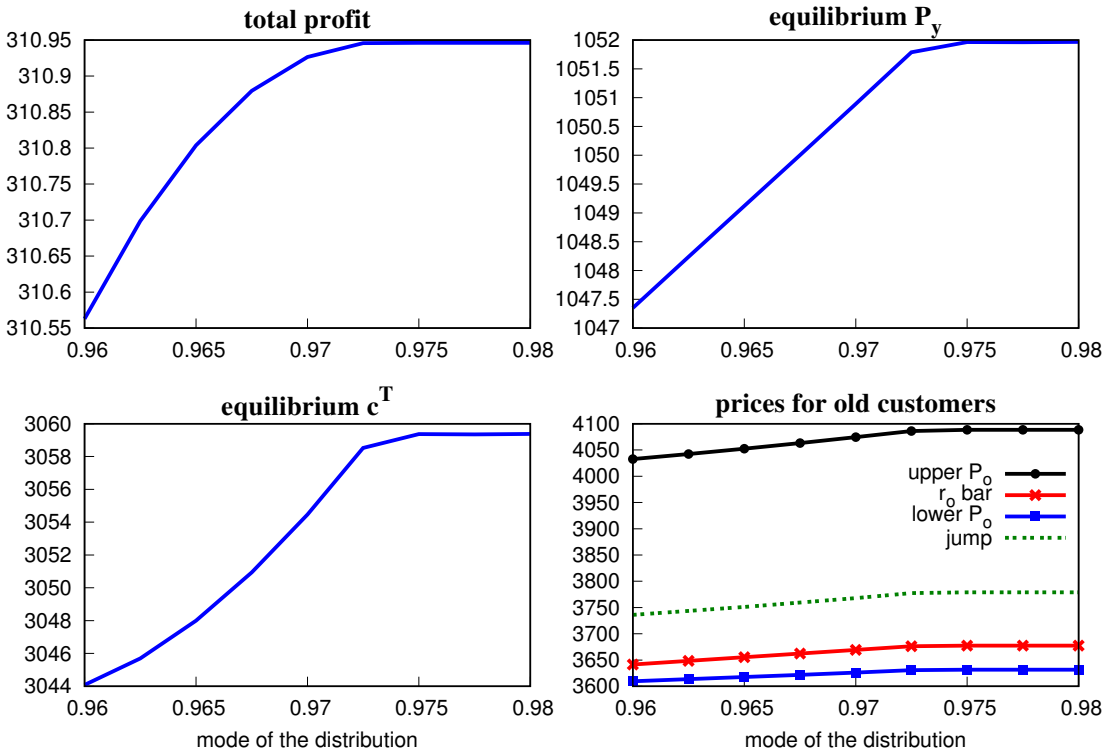


Figure 10: Varying the mode of the health shock distribution in the model

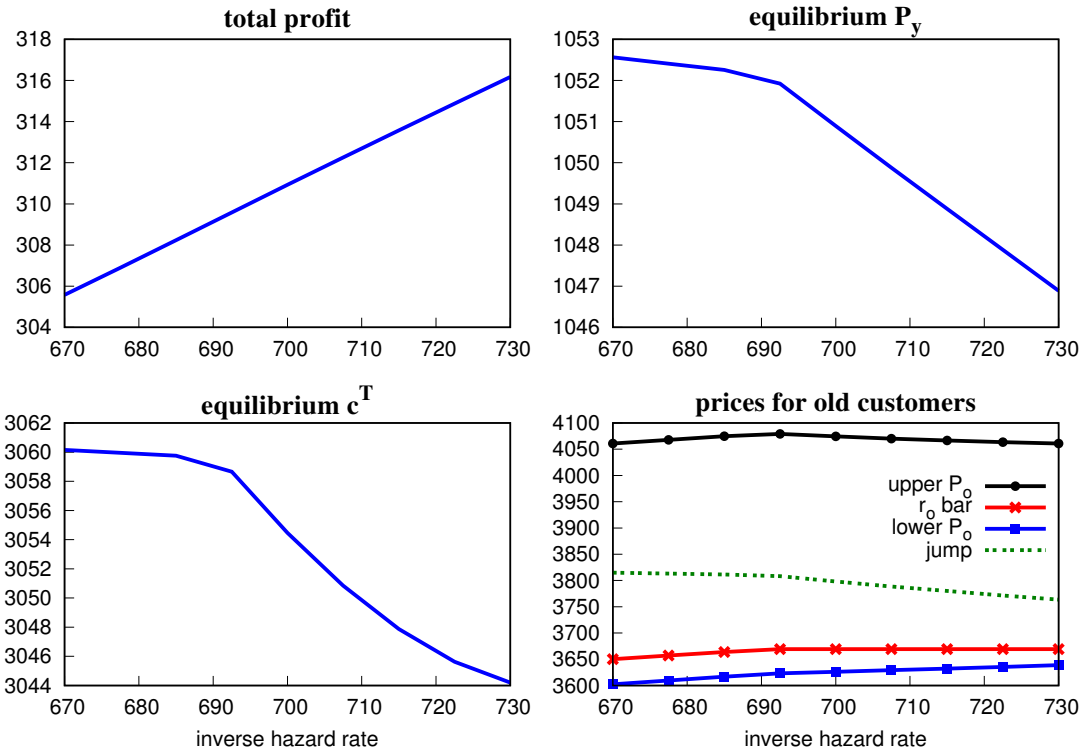


Figure 11: Varying the inverse hazard rate γ in the model

5 Testing the Model's Predictions

Now we turn directly to our dataset of insurance premiums to test the main predictions of the model. In particular, as the consumer's hold-up problem becomes weaker (through a lower transaction cost μ or less adverse distribution of health shocks), the equilibrium cost thresholds c^T and c^M in the model fall, leading to more frequent price adjustments. At the same time, a weaker hold-up problem leads to a reduced magnitude of premium changes due to both a smaller jump between the rigid intervals of the premium schedule, and a higher likelihood of adjustments within the frictionless pricing region. We test these predictions by looking at average probability and size of premium adjustments across time and across renewal terms. Finally, we show that the marginal cost presented in Figure 5 is positively correlated with the hazard of premium adjustment, rather than size, indicating that life insurance firms tend to respond to cost shocks on the extensive margin. In the concluding part of this section, we use the data to address common alternative theories of price rigidity in the context of life insurance.

5.1 Premium Changes Across Time

In the first test, we divide our sample into two sub-periods: 1990-1999 and 2000-2009. It can be argued that between these two time intervals, the consumer hold-up problem became weaker mainly for two reasons. First, the emergence of on-line pricing tools led to a reduced transaction cost needed to search for a life insurance policy and compare premiums across different companies and products. The internet also enabled customers to purchase policies directly from the insurance firms, avoiding the need to meet an agent physically.[27] Second, mortality rates among the insured dropped significantly in the 2000s as documented by the two vintages of Select and Ultimate mortality tables issued in 2001 and 2008. For example, a cumulative 20-year probability of death for a 30-year-old male policyholder decreased from 2.36% to less than 1.96%, while for a 40-year-old male policyholder the cumulative 20-year death rate fell from 5.19% to 4.39%. Analogous to a reduction in the transaction cost, a leftward shift in the distribution of health shocks in the model leads to more frequent and less sizable premium adjustments.

Table 6 presents the average fraction of premium changes and the average size of adjustments for the two sub-periods. Between the 1990s and the 2000s, the average fraction of life insurance policies adjusting premiums in any month increased from 2.31% to 2.76%, while the average size of the adjustment fell from 12.02% to 10.22%. Both differences are statistically significant at the 5% confidence level.

Table 6: Testing the difference in frequency and size of premium adjustments over time

Period	**Months**	**% adjusting**	**St. err.**	**p-value**
1990-1999	116	2.31	0.19	
2000-2009	120	2.76	0.20	
Difference		−0.45	0.27	0.05
Period	**Months**	**size (in %)**	**St. err.**	**p-value**
1990-1999	106	12.02	0.60	
2000-2009	111	10.22	0.59	
Difference		1.80	0.84	0.02

Note: One sided t-test for equality of means, with alternative hypotheses $H_a : diff(\%\text{ adjusting}) < 0$ and $H_a : diff(\text{size}) > 0$, respectively. Average premium changes are computed only for the months where at least one is observed, hence the difference in the number of observations.

[27] A similar assertion is made by Brown and Goolsbee (2002) to argue that the popularization of internet pricing tools over that period of time led to an overall decrease in the *level* of life insurance premiums.

5.2 Premium Changes Across Renewal Terms

In the second test, we investigate whether the frequency and size of price adjustments vary with the length of renewability term. As the term extends, the level of marginal cost, premiums, and consumers' valuations increase, while the one-time transaction cost remains unchanged (it takes the same amount nominally to invest in purchasing ART or 10-year level-term). In other words, the transaction cost falls relative to the size of the consumer's surplus as we move from one-year term to 10- and 20-year term. Suppose we calibrate our model to three different term lengths, and normalize the lower rigid price $\bar{P}_o$ to be equal to 100 across all calibrations. The normalized value of the transaction cost parameter μ would then decrease as the term extends. Figure 9 shows that in the model this leads to a drop in both cost thresholds, more flexibility in adjusting premiums, and a smaller jump between the two rigid parts of the premium schedule.

Table 7 presents the frequency of premium changes, along with the average magnitude of adjustments, for the four standard lengths of level-term renewable insurance. Two stark observations arise from this test. First, as the term extends, we indeed observe a larger frequency of price changes, climbing monotonically from 1.07% for ART policies up to 3.57% for 20-year level-term. The analysis of variance between and within the groups confirms that these differences are statistically significant. Second, ART policies exhibit a significantly higher average size of the premium adjustment at 15%, compared to roughly 10% for all the remaining term lengths. Interestingly, the term length above one year is not informative about the expected size of a premium change and the mean of squares within these three groups exceeds the means of squares between them (with p-value of the F test equal to 0.56).

Table 7: Testing the difference in frequency and size of premium adjustments across terms

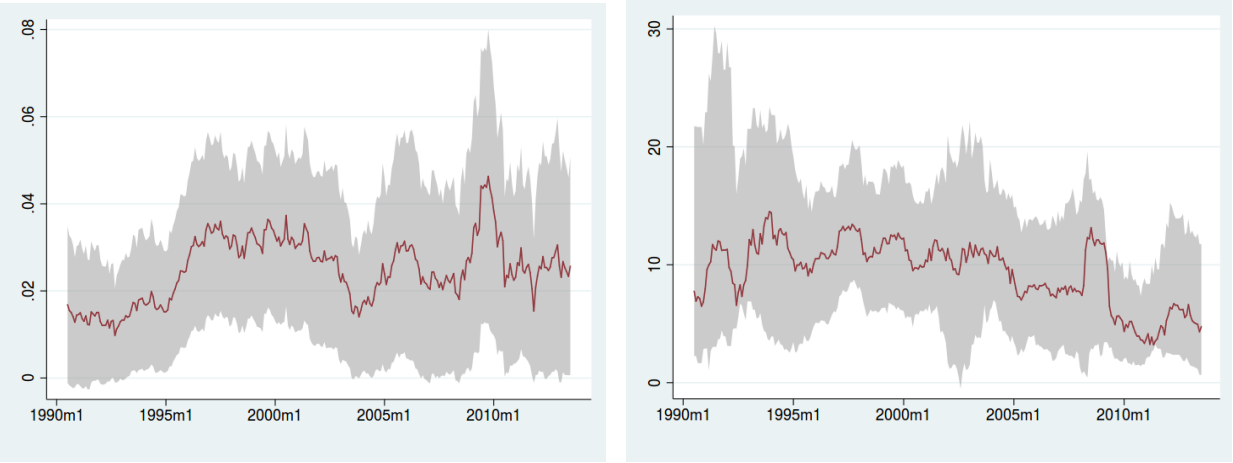
Term length	**N obs.**	**% adjusting**	**St. err.**	**Variance analysis**	
1 year	13,644	1.07	0.08	MSB	1.69
5 years	6,537	2.07	0.18	MSW	0.02
10 years	20,959	2.99	0.12	F-stat	68.01
20 years	14,689	3.57	0.15	p-value	0.00

Term length	**N obs.**	**size (in %)**	**St. err.**	**Variance analysis**	
1 year	146	15.01	1.00	MSB	1017.36
5 years	135	10.85	0.75	MSW	84.04
10 years	626	10.02	0.34	F-stat	21.11
20 years	525	10.38	0.40	p-value	00.00

5.3 Relationship Between Premium Changes and Cost Shocks

We now focus on the dynamics of life insurance premiums over time using the entire available sample. Figure 12 plots the 12-month moving averages of the fraction of products that adjust their premium and the average size of the adjustment. Recall from Figure 5 that the largest shocks to marginal cost in our sample roughly occur in the late 1990s, around 2005, late 2008, and 2012. Figure 12(a) shows that each of those episodes was accompanied by an increase of at least one percentage point in the fraction of companies that adjusted their premium schedules, and the increase is statistically significant. On the other hand, Figure 12(b) reveals no such apparent correlation between the cost shock and the average magnitude of changes (except for late 2008, in the presence of a record-high cost). The most apparent observation from that time series is probably the gradual decline in the average size of premium jumps discussed in the previous section.

Table 8 formalizes these findings by computing correlations between the cost shock and the frequency and average size of premium changes. This relationship is generally positive for the fraction of firms that adjust and was the strongest during the 1990s, when the consumer's hold-up problem was likely to be more severe, as discussed in Section 5.1. On the other hand, the average adjustment size turns out to be negatively related to the marginal cost shock. In light of our theory, this evidence suggests therefore that life insurers respond to industry-wide shocks predominantly with an increased hazard of adjusting the premium schedule, rather than altering the magnitude of such an adjustment.



(a) Fraction of products that adjust premium

(b) Average size of premium adjustment

Note: 6-month moving averages applied. Gray areas depict the 95% confidence intervals.

Figure 12: Fraction and average size of premium changes

Table 8: Correlation between the cost shock and frequency/size of premium changes

Correlation with cost:	**% adjusting**	**adjustment size**
1990-1999	0.67	0.24
2000-2009	0.20	−0.03
Full sample	0.28	−0.18

5.4 Alternative theories

We now investigate whether the rigidity of life insurance premiums could potentially be explained by existing models of price stickiness. Following Klenow and Kryvtsov (2008), we consider two broad classes of models: time-dependent and state-dependent pricing models. Then, we investigate if a model of competition can generate realistic premium rigidity.

5.4.1 Time-Dependent Models of Sticky Prices

In time-dependent pricing frameworks, for example Taylor (1980) or Calvo (1983), firms adjust prices with exogenous frequency. The spell duration is subject to a random shock and every period a fixed number of firms reoptimize their price. The longer an individual price remains staggered, the more shocks accumulate in the meantime, resulting in larger average size of the adjustment. In a time-dependent pricing setup, we would expect to observe variation on the *intensive margin*, and much less so on the *extensive margin* which is determined exogenously. Figure 12 shows that the opposite appears to be the case in the life insurance market, where the firms tend to respond to cost shock predominantly on the extensive margin. In addition, Figure 13 shows a plot of all premium changes in our data (expressed in absolute value) as function of premium duration. As can be noticed, the points are scattered with no clear pattern and the correlation of the two variables is about 0.13. Table 9 confirms this in a regression analysis. The relationship between premium duration and size of the adjustment is rather weak (albeit positive), with the slope of 0.07% and R-squared of less than 0.02. We conclude that the standard time-dependent models are not a promising alternative to explain premium rigidities observed in the life insurance market.

Table 9: Regression results (dependent variable: abs_change)

	coefficient	s.e.	$P > \|z\|$		
constant	9.372	0.372	0.000	No. of obs.	1432
duration	0.066	0.014	0.000	Adj. R-sq.	0.016

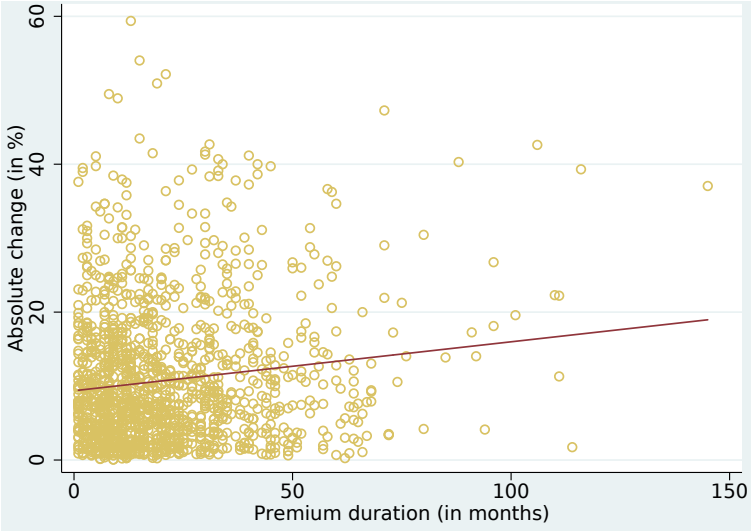


Figure 13: Premium duration and adjustment size in the life insurance market

5.4.2 State-Dependent Models of Sticky Prices

By contrast, in state-dependent pricing models, for example Dotsey et al. (1999) or Golosov and Lucas (2007), individual firms are subject to heterogeneous "menu costs" and can choose when to adjust their prices. Changes then occur predominantly on the extensive margin, and the hazard of price change is an increasing function of duration because incoming shocks move the optimal price away from the one currently posted. Figure 14 plots the hazard of premium adjustment in our dataset at all durations for which we observe multiple changes. Notice that the relationship is generally *negative*, with the correlation of -0.44. Table 10 summarizes the regression results of change hazard on duration. An additional month of premium duration tends to reduce the hazard by 0.01%, but only 18% of the variation in hazard can be explained through this channel. Interestingly, it appears that the increasing hazard may be a local feature of the premiums that have remained staggered for more than 50 months. We do not have enough observations to make this assertion robust though. We conclude that state-dependent pricing theories are not readily capable of explaining premium rigidity in the life insurance market.

Table 10: Regression results (dependent variable: hazard)

	coefficient	s.e.	$P > \|z\|$		
constant	0.0309	0.00176	0.000	No. of obs.	88
duration	−0.0001	0.00003	0.000	Adj. R-sq.	0.184

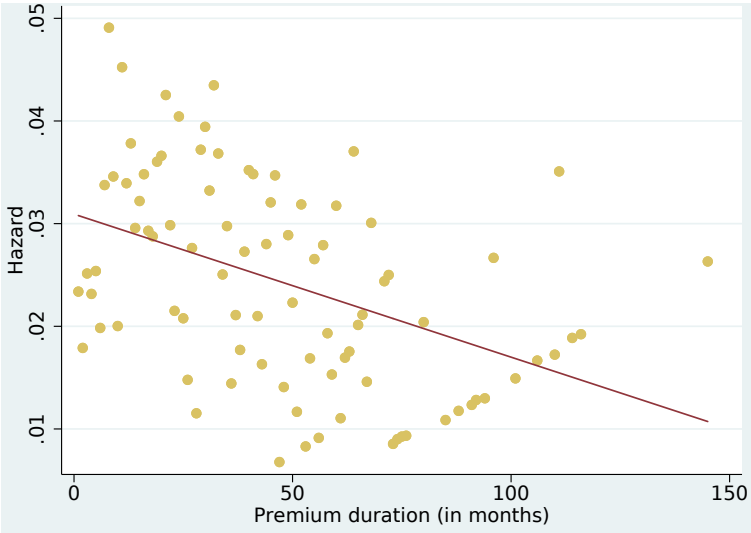


Figure 14: Premium duration and adjustment hazard in the life insurance market

5.4.3 Competition

Our model accounts for market competition in a reduced form, allowing consumers to search for an outside option in the renewable insurance market. It has been shown in the literature though that explicit modeling of competition can also generate price rigidity. In a monopolistically competitive market, Nishimura (1986) shows that prices become rigid as the price elasticity of demand approaches infinity if a firm cannot infer whether a transient cost shock is market-wide or firm-specific. Firms set prices based on the expectation of other firms' prices. If a firm responds to the cost shock by increasing its price then, with elastic demand, it will attract few consumers when the shock is firm-specific. On the other hand, if a firm responds to the cost shock by lowering its price, then it would attract many consumers when the shock is firm-specific, but the lower price is not profitable. As a result, equilibrium prices become less sensitive to transient shocks as markets become more competitive.

Even though competition can generate price rigidity in an incomplete information environment, it also entails price concentration. Indeed, if prices were dispersed, then firms with high prices would not be competitive. However, the dispersion in life insurance premiums (measured by the coefficient of variation) is generally high suggesting that life insurance markets are not competitive. Table 11 summarizes the premium dispersion for our four standard term lengths. The dispersion decreases with term length, with an exception for the 20-year level-term which also has a much higher standard deviation so its premiums may not be statistically more dispersed than shorter-term contracts. From Table 7, premiums adjust more frequently as the term length increases. This suggests that the relationship between price dispersion and rigidity runs counter to the prediction of Nishimura (1986).

Table 11: Coefficient of variation of premiums across terms

Term length	Mean	St. dev.	Min.	Max.
1 year	0.18	0.02	0.15	0.23
5 years	0.17	0.03	0.11	0.28
10 years	0.15	0.03	0.11	0.21
20 years	0.17	0.05	0.06	0.28

Note: For each term length, this table shows the distribution of cross-sectional coefficients of variation over time. The total number of observations is 282 months.

6 Conclusion

We show that the market for life insurance has exhibited a remarkable degree of price rigidity since 1990. Firms that changed premiums in the analyzed sample did so on average every 39 months, preferring one-time jumps of large magnitude to more frequent and gradual price adjustments. We build a theoretical model to explain this phenomenon, based on the assumption that consumers are locked-in due to a relationship-specific investment. In line with what we find in the data, the model predicts that premiums remain constant for a wide range of cost shock realizations, while potential changes take the form of discrete jumps.

Our hypothesis is obviously not the only explanation for the observed rigidity of life insurance premiums. As Table 7 shows, even the 20-year level-term premiums are quite rigid which leaves room for complementary theories. At the same time however, we provide guidelines for future research by showing that life insurance data does not support the standard price stickiness models.

The broader implications of our work include providing microfoundations for price stickiness observed in the markets characterized by long-term commitment. In particular, we show that the price rigidity arises endogenously as a solution to a time inconsistency problem that could otherwise deter consumers from buying the good. Future research should investigate the pricing behavior of other financial or contractual services, and find out if similar products also exhibit pricing anomalies such as the ones found in life insurance contracts.

There are also several other potential applications of the model developed in this paper for future research. One of them is to apply our theory to other environments plagued by the hold-up problem. As we have noted, the theoretical model is quite general and can be adapted to other settings to explain why adjustments may be drastic rather than gradual. For example, if it is difficult for employees to switch jobs, our theory may potentially explain why employers prefer layoffs over wage cuts during economic downturns, in order to attract quality employees *ex ante*.

References

Alonso, Ricardo and Niko Matouschek, "Optimal Delegation," *Review of Economic Studies*, 2008, *75*, 259–293.

Amador, Manuel and Kyle Bagwell, "The Theory of Optimal Delegation with an Application to Tariff Caps," *Econometrica*, 2013, *81* (4), 1541–1599.

_ , Ivan Werning, and George-Marios Angeletos, "Commitment vs. Flexibility," *Econometrica*, 2006, *74* (2), 365–396.

Athey, Susan, Andrew Atkeson, and Patrick Kehoe, "The Optimal Degree of Monetary Policy Discrestion," *Econometrica*, 2005, *73*.

Bils, Mark and Peter J. Klenow, "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 2004, *112* (5).

Brown, Jeffrey R. and Austan Goolsbee, "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy*, 2002, *110* (3), 481–507.

Bureau of Consumer Protection and Bureau of Economics, *Life Insurance Cost Disclosure* 1979.

California Department of Insurance, "California Life & Annuity Insurance Industry: 2003 Market Share Report," *Rate Specialist Bureau, Department of Insurance*, 2004.

Calvo, Guillermo A., "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 1983, *12* (3).

Cawley, John and Tomas Philipson, "An Empirical Examination of Information Barriers to Trade in Insurance," *American Economic Review*, 1999, *89* (4).

Cullen, Joseph and Oleksandr Shcherbakov, "Measuring consumer switching costs in the wireless industry," *Unpublished manuscript*, 2010.

Dotsey, Michael, Robert King, and Alexander Wolman, "State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output," *Quarterly Journal of Economics*, 1999, *114* (2).

Golosov, Mikhail and Robert E. Lucas, "Menu Costs and Phillips Curves," *Journal of Political Economy*, 2007, *115* (2).

Gottlieb, Daniel and Kent Smetters, "Lapse-Based Insurance," *Working Paper*, 2016.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, 2007, *54* (8).

Halac, Marina and Pierre Yared, "Fiscal Rules and Discretion Under Persistent Shocks," *Econometrica*, 2014, *82* (5).

_ **and** _ , "Fiscal Rules and Discretion under Self-Enforcement," *Working Paper*, 2017.

_ **and** _ , "Fiscal Rules and Discretion in a World Economy," *American Economic Review*, 2018, *forthcoming.*

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

Hendel, Igal and Alessandro Lizzeri, "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *Quarterly Journal of Economics*, 2003, *118* (1).

Holmstrom, Bengt, "On the Theory of Delegation," in Marcel Boyer and Richard Kihlstrom, eds., *Bayesian Models in Economic Theory*, New York: North Holland, 1984.

Honka, Elisabeth, "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 2014, *45* (4).

Hortacsu, Ali and Chad Syverson, "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *Quarterly Journal of Economics*, 2004, *119.*

Huntington, Henry S., "Derivation of Premium Rates for Renewable Term Insurance," *Transactions of Society of Actuaries*, 1958, *10* (28).

Illanes, Gastón, "Switching Costs in Pension Plan Choice," *Unpublished manuscript*, 2016.

Klemperer, Paul, "Markets with Consumer Switching Costs," *Quarterly Journal of Economics*, 1987, *102* (2).

Klenow, Peter J. and Oleksiy Kryvtsov, "State-Dependent or Time-Dependent Pricing: Does it Matter for Recent U.S. Inflation," *Quarterly Journal of Economics*, 2008, *123* (3).

Koijen, Ralph S. J. and Motohiro Yogo, "The Cost of Financial Frictions for Life Insurers," *American Economic Review*, 2015, *105* (1).

_ **and** _ , "Shadow Insurance," *Econometrica*, 2016, *84* (3).

LIMRA, "What's happening in the term marketplace," *LIMRA's MarketFacts*, 1994, *13* (2).

Melumad, Nehum and Toshiyuki Shibano, "Communication in Settings with No Transfers," *RAND Journal of Economics*, 1991, *22* (2).

Miller, Daniel P. and Jungwon Yeo, "Estimating Dynamic Discrete Choice Models of Product Differentiation: An Application to Medicare Part D with Switching Costs," *Unpublished manuscript*, 2012.

Nishimura, Kiyohiko, "Rational Expectations and Price Rigidity in a Monopolistically Competitive Market," *Review of Economic Studies*, 1986, *53.*

Nosal, Kathleen, "Estimating Switching Costs for Medicare Advantage Plans," *Unpublished manuscript*, 2012.

Pauly, Mark, Kate Withers, Krupa Subramanian-Viswanathan, Jean Lemaire, John Hershey, Katrina Armstrong, and David Asch, "Price Elasticity of Demand for Term Life Insurance and Adverse Selection," *NBER Working Paper*, 2003, *9925.*

Shcherbakov, Oleksandr, "Measuring consumer switching costs in the television industry," *The RAND Journal of Economics*, 2016, *47* (2).

Taylor, John B., "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, 1980, *88* (1).

Weiergräber, Stefan, "Network Effects and Switching Costs in the US Wireless Industry. Disentangling sources of consumer inertia," *SFB/TR 15 Discussion Paper*, 2014, *512.*

Appendices (for online publication)

A Nominal and Real Rigidity

This section analyzes the effect of inflation on life insurance premiums. We first show that nominal rigidity implies real rigidity *per dollar of real face value*, and vice versa. Next, we discuss how inflation affects the pricing behavior of life insurers and how it relates to the results of our paper.

Consider a simple framework where perfectly competitive life insurance companies face a constant mortality rate m and a constant interest rate r, The (nominal) face value of a generic policy is F_t, and its real counterpart is F_t^R, where $t = 0$ is the base year. Let P_t^N denote the nominal premium for a one-period (non-renewable) insurance and P_t^R be the real premium. Let π be a (constant) inflation rate. Then, Lemma 4 shows that nominal premiums are rigid if and only if real premiums per dollar of real face value are rigid.

Lemma 4 *For any t, $P_t^N = P^N$ if and only if for all t, $\frac{P_t^R}{F_t^R} = \bar{P}^R$.*

Proof To show sufficiency, suppose that in a competitive market for non-renewable insurance, nominal premiums are rigid (i.e. the mortality rate, interest rate and nominal face value are fixed): for all t,

$$P_t^N = \frac{m}{1+r} F_t = \frac{m}{1+r} F.$$

The real premium is given by

$$\begin{aligned} P_t^R &= \frac{P^N}{(1+\pi)^t} \\ &= \frac{m}{1+r} \frac{F}{(1+\pi)^t} \\ &= \frac{m}{1+r} F_t^R. \end{aligned}$$

The real premium *per dollar of real face value* is

$$\frac{P_t^R}{F_t^R} F_0 = \frac{m}{1+r} F_0 \equiv \bar{P}^R,$$

where F_0 is the face value in the base year. This shows that nominal rigidity of premiums implies that premiums are also rigid for a policy with fixed real face value.

The proof for necessity follows exactly the steps shown above in the reverse order. ■

From the proof of Lemma 4, notice that as the nominal premium is rigid by assumption, the real premium decreases over time at the constant rate of inflation π. However, this is not the insurance product that we consider in our model, because its real face value also decreases over time. Our model assumes a constant real face value, so the focus of Lemma 4 is on the real premium *per dollar of real face value*.

Now, consider a two-period renewable insurance policy. Let $P^N_{t,a}$ denote the nominal premium in period t dollars for age a individual, and let $P^R_{t,a}$ be the real premium. The nominal payment over time is $\left(P^N_{t,a}, P^N_{t+1,a+1}\right)$. The next lemma uses Lemma 4 to establish the equivalence relation between nominal and real rigidity in a competitive market for renewable insurance. The result can be generalized to any n-period renewable insurance.

Lemma 5 *For any t and age a, $P^N_{t,a} = P^N_a$ if and only if for any t and any age a, $\frac{P^R_{t,a}}{F^R_t} = \bar{P}^R_a$.*

Proof To prove sufficiency, we define the actuarially fair premiums for renewable insurance according to the backward induction approach of Huntington (1958) (for details, see Appendix D). In $t+1$, a policy bought in t at age a, becomes non-renewable and the actuarially fair premium satisfies $P^N_{t+1,a+1} = \frac{m}{1+r}F$. In t, nominal premiums satisfy a zero-profit condition

$$P^N_{t,a} + \frac{1-m}{1+r}P^N_{t+1,a+1} = F\left[\frac{m}{1+r} + \frac{m\,(1-m)}{(1+r)^2}\right]. \tag{10}$$

Converting nominal premiums to real ones yields

$$\begin{aligned} P^R_{t,a}(1+\pi)^t + \frac{1-m}{1+r}P^R_{t+1,a+1}(1+\pi)^{t+1} &= F\left[\frac{m}{1+r} + \frac{m\,(1-m)}{(1+r)^2}\right] \\ \iff \frac{P^R_{t,a}}{F^R_t}F + \frac{1-m}{1+r}\frac{P^R_{t+1,a+1}}{F^R_{t+1}}F &= F\left[\frac{m}{1+r} + \frac{m\,(1-m)}{(1+r)^2}\right]. \end{aligned}$$

The second lines follows from the fact that $F^R_t = \frac{F_0}{(1+\pi)^t}$ and $F_0 = F$, since $t = 0$ is the base year. By Lemma 4, $P^N_{t+1,a+1} = P^N_{a+1}$ if and only if $\frac{P^R_{t+1,a+1}}{F^R_t} = \bar{P}^R_{a+1}$ for any t and a. As a result, if for all t and a, $P^N_{t,a} = P^N_a$ and $P^N_{t+1,a+1} = P^N_a$, it immediately follows that $\frac{P^R_{t,a}}{F^R_t} = \bar{P}^R_a$.

The proof for necessity follows exactly the steps above in the reverse order. ■

Lemmas 4 and 5 show that the nominal rigidity observed in the data is equivalent to real rigidity *per dollar of real face value* in the simplest case of actuarially fair premiums. As inflation erodes the value of premiums paid by consumers over time, it does so with the value of death benefits as well. From this point of view, a theory of real price rigidity (such as the one presented in this paper) is informative for explaining the nominal rigidity observed in the data.

Naturally, the presence of inflation may affect the pricing strategy of insurers by inducing them to change the age profile of premiums. As pointed out by Hendel and Lizzeri (2003), life insurance premiums are typically front-loaded which implies that the majority of the profit comes from initial premiums. For example, an increase in expected inflation may lead to more front-loading and higher profit because the nominal value of death benefits will be eroded more than the value of premiums. This suggests that a model in real terms may not capture the full dynamics of front-loaded nominal premiums. In other words, in a world without inflation we would expect premiums to be even more rigid than what we already observe in the data and document in Section 2. Therefore, a theory of real premium rigidity is needed to explain these facts.

B Proofs

Proof of Lemma 1: First note that $B^{ren}(r_o;t) \geq B^{non}(r_o;t)$ and $B^{ren}(r_o;t), B^{non}(r_o;t) \geq 0$ for all r_o and for all t. This is because there are more options available for renewable policyholders and consumers can always choose to forgo coverage.

Next, we show that $B^{non}(r_o;t)$ and $B^{ren}(r_o;t)$ are weakly increasing in r_o for any t. Note that there exists $\tilde{r}_o$ such that for all $r_o < \tilde{r}_o$, $B^{non}(r_o;t) = 0$ and for $r_o \geq \tilde{r}_o$, we have

$$\begin{aligned} B^{non}(r_o;t) &= \int_{\underline{c}}^{\overline{c}} \left[\int_{\epsilon} \max\{0, r_o - \epsilon c_{o,t+1}\} dZ(\epsilon) - \mu \right] dG(c_{o,t+1}) \\ &= \int_{\underline{c}}^{\overline{c}} \int_{0}^{\frac{r_o}{c_{o,t+1}}} (r_o - \epsilon c_{o,t+1}) g(c_{o,t+1}) z(\epsilon) d\epsilon dc_{o,t+1} - \mu. \end{aligned}$$

Differentiating $B^{non}(r_o;t)$ with respect to r_o yields $\int_{\underline{c}}^{\overline{c}} Z\left(\frac{r_o}{c_{o,t+1}}\right) dG(c_{o,t+1})$, so $B^{non}(r_o;t)$ is strictly increasing for $r_o \geq \tilde{r}_o$. Since $B^{ren}(r_o;t) \geq B^{non}(r_o;t)$, $B^{ren}(r_o;t)$ is also strictly increasing for $r_o \geq \tilde{r}_o$.

Suppose there exists valuation $\hat{r}_o$ such that $B^{ren}(\hat{r}_o;t) > B^{non}(\hat{r}_o;t)$. We will show that $B^{ren}(r_o;t) - B^{non}(r_o;t)$ is increasing in $r_o \geq \hat{r}_o$. If $B^{ren}(\hat{r}_o;t) > B^{non}(\hat{r}_o;t)$, then there exists a set $\hat{\mathcal{C}}$ with strictly positive measure defined as $\hat{\mathcal{C}} = \{c_{o,t+1} | \hat{r}_o > P_{o,t+1}(c_{o,t+1})\}$. Since if $\hat{\mathcal{C}}$ is empty or measure zero, then it cannot be the case that $B^{ren}(\hat{r}_o;t) > B^{non}(\hat{r}_o;t)$. For

$r_o \geq \hat{r}_o$, we have

$$B^{ren}(r_o;t) - B^{non}(r_o;t) \geq$$
$$\int_{\hat{\mathcal{C}}} \max\left\{ r_o - P_{o,t+1}(c_{o,t+1}), \int_{\epsilon} \max\left\{ r_o - P_{o,t+1}(c_{o,t+1}), r_o - \epsilon c_{o,t+1} \right\} dZ(\epsilon) - \mu \right\} dG(c_{o,t+1})$$
$$- \int_{\hat{\mathcal{C}}} \left[\int_{\epsilon} \max\left\{ 0, r_o - \epsilon c_{o,t+1} \right\} dZ(\epsilon) - \mu \right] dG(c_{o,t+1}). \quad (11)$$

The inequality comes from the fact that for higher valuations the set of costs such that $r_o > P_{o,t+1}(c_{o,t+1})$ should be weakly larger. Let $\hat{C}^r$ denote the set of cost realizations where the policyholder would renew immediately and $\hat{C}^s$ denote the set where the policyholders search. Define the two sets such that they are mutually exclusive (if policyholders are indifferent, they renew), then the right-hand side of (11) can be rewritten as

$$\underbrace{\int_{\hat{\mathcal{C}}^r} \left[r_o - P_{o,t+1}(c_{o,t+1}) \right] dG(c_{o,t+1})}_{\text{renew immediately}}$$
$$+ \underbrace{\int_{\hat{C}^s} \left[r_o Z\left(\frac{P_{o,t+1}(c_{o,t+1})}{c_{o,t+1}} \right) - c_{o,t+1} \int_0^{\frac{P_{o,t+1}\left(c_{o,t+1}\right)}{c_{o,t+1}}} \epsilon dZ(\epsilon) \right] dG(c_{o,t+1})}_{\text{search and sign non-renewable}}$$
$$+ \underbrace{\int_{\hat{C}^s} \left[r_o - P_{o,t+1}(c_{o,t+1}) \right] \left[1 - Z\left(\frac{P_{o,t+1}(c_{o,t+1})}{c_{o,t+1}} \right) \right] dG(c_{o,t+1})}_{\text{search and then renew}}$$
$$- \underbrace{\int_{\hat{\mathcal{C}}} \int_0^{\frac{r_o}{c_{o,t+1}}} (r_o - \epsilon c_{o,t+1}) \, dZ(\epsilon) \, dG(c_{o,t+1})}_{\text{sign non-renewable}}.$$

Differentiating the above expression with respect to r_o yields $\int_{\hat{\mathcal{C}}} \left[1 - Z\left(\frac{r_o}{c_{o,t+1}} \right) \right] dG(c_{o,t+1})$, which is strictly positive so $B^{ren}(r_o;t) - B^{non}(r_o;t)$ is strictly increasing in $r_o \geq \hat{r}_o$.

Finally, since $B^{ren}(r_o;t) - B^{non}(r_o;t)$, $B^{ren}(r_o;t)$ and $B^{non}(r_o;t)$ are increasing for sufficiently large r_o, consumers of any generation would purchase renewables if their valuation is sufficiently large. ■

Proof of Lemma 2: For part (i.), by (4) we have the following demand for $c_o \in \underline{\mathcal{C}}_o$,

$$D_o(P_y, P_o(c_o)) = (1 - m_y)\left[1 - H(\bar{r}_o) \right].$$

The insurance company takes $\bar{r}_o$ as given, so for costs in $\underline{\mathcal{C}}_o$, the demand is independent of the variations in $P_o(c_o)$.

Suppose $c_o \in \underline{\mathcal{C}}_o$ is the actual cost and the insurance company reports $c_o' \in \underline{\mathcal{C}}_o$, then the incentive compatible premium schedule requires $P_o(c_o) \geq P_o(c_o')$. Now suppose c_o' is the actual cost and the insurance company reports c_o, then the incentive compatibility requires $P_o(c_o) \leq P_o(c_o')$. Therefore, we have $P_o(c_o) = P_o(c_o')$ for any $c_o, c_o' \in \underline{\mathcal{C}}_o$.

Finally, to show that $\underline{\mathcal{C}}_o = \{c_o \mid P_o(c_o) < \bar{r}_o\}$ has strictly positive measure, first assume that $\underline{\mathcal{C}}_o$ is measure zero. This implies that for almost all $c_o \in [\underline{c}, \overline{c}]$, $P_o(c_o) \geq \bar{r}_o$. By (2), $0 = \frac{P_y - P_y^{NR}}{1 - m_y}$, which is a contradiction when $P_y > P_y^{NR}$.

For part (ii.), if the premium schedule is incentive compatible, then for any $c_o \in [\underline{c}, \overline{c}]$ and $\epsilon > 0$ such that $c_o + \epsilon \in [\underline{c}, \overline{c}]$ we have

$$[P_o(c_o) - c_o] D_o(P_y, P_o(c_o)) \geq [P_o(c_o + \epsilon) - c_o] D_o(P_y, P_o(c_o + \epsilon)),$$

$$[P_o(c_o + \epsilon) - (c_o + \epsilon)] D_o(P_y, P_o(c_o + \epsilon)) \geq [P_o(c_o) - (c_o + \epsilon)] D_o(P_y, P_o(c_o)).$$

Summing the incentive constraints yields $D_o(P_y, P_o(c_o)) \geq D_o(P_y, P_o(c_o + \epsilon))$. Since D_o is weakly decreasing in P_o, incentive compatible premiums have to be weakly increasing in cost.

Finally, the existence of a cutoff follows immediately from the fact that P_o is weakly increasing in cost and $P_o(c_o) \geq P_o(c_o')$ for any $c_o \in \overline{\mathcal{C}}_o$ and $c_o' \in \underline{\mathcal{C}}_o$. ■

Proof of Lemma 3: For part (i.), suppose $P_o(c_o) < P_o^*(c_o)$ for some $c_o \in (c_o', c_o'')$. Since $P_o(c_o)$ is strictly increasing and continuous, then there exists $\epsilon > 0$ such that $P_o(c_o) < P_o(c_o + \epsilon) < P_o^*(c_o)$. The hazard rate is non-decreasing, so $(P_o - c_o)(1 - H(P_o))$ is single peaked. This implies

$$[P_o(c_o + \epsilon) - c_o][1 - H(P_o(c_o + \epsilon))] > [P_o(c_o) - c_o][1 - H(P_o(c_o))],$$

which violates incentive compatibility at c_o. The argument also applies for $P_o(c_o) > P_o^*(c_o)$.

For part (ii.), it is trivial to show that $\bar{r}_o > \bar{P}_o$ from (2) when $P_y > P_y^{NR}$. To show that $P_o(c^T) > \bar{r}_o$, we only need to rule out $\bar{r}_o = P_o(c^T)$. If $\bar{r}_o = P_o(c^T)$, then by (8), we have $\bar{P}_o = \bar{r}_o$. By (2), it implies that $0 = \frac{P_y - P_y^{NR}}{1 - m_y}$, which is a contradiction when $P_y > P_y^{NR}$.

Next, we will establish the fact that the frictionless premium is not incentive compatible for $\overline{\mathcal{C}}_o$. By (2), $\bar{r}_o > \bar{P}_o$. Suppose $P_o(c_o) = P_o^*(c_o)$ for all $c_o \in \overline{\mathcal{C}}_o$, then (8) implies

$$LHS \equiv (\bar{P}_o - c^T)[1 - H(\bar{r}_o)] = RHS \equiv (P_o^*(c^T) - c^T)[1 - H(P_o^*(c^T))].$$

Notice the following: $LHS < (\bar{P}_o - c^T)[1 - H(\bar{P}_o)]$. Since RHS is the optimal frictionless

profit, it follows that $RHS > LHS$, and is only equal when $P_y \leq P_y^{NR}$ and with strict inequality when $P_y > P_y^{NR}$. Hence, it cannot be the case that $P_o(c_o) = P_o^*(c_o)$ for all $c_o \in \overline{\mathcal{C}}_o$.

Finally, we will show there is rigidity in $\overline{\mathcal{C}}_o$. Define $P_o^+\left(c^T\right) \equiv \lim_{c_o \to c^T} P_o(c_o)$ and $c^M = P_o^{*-1}\left(P_o^+\left(c^T\right)\right)$. Consider $c_o \in \left(c^T, \min\left\{c^M, \overline{c}\right\}\right)$, then $P_o^+\left(c^T\right) > P_o^*(c_o)$. By Lemma 2, P_o is weakly increasing, so $P_o(c_o) \geq P_o^+\left(c^T\right)$ for any $c_o \in \left(c^T, \min\left\{c^M, \overline{c}\right\}\right)$. Suppose there exists $\hat{c}_o \in \left(c^T, \min\left\{c^M, \overline{c}\right\}\right)$ such that $P_o(\hat{c}_o) > P_o^+\left(c^T\right)$. This implies the following ordering: $P_o(\hat{c}_o) > P_o^+\left(c^T\right) > P_o^*(\hat{c}_o)$. However, $(P - \hat{c}_o)(1 - H(P))$ is single peaked around $P_o^*(\hat{c}_o)$, so

$$\left(P_o^+\left(c^T\right) - \hat{c}_o\right)\left(1 - H\left(P_o^+\left(c^T\right)\right)\right) > \left(P_o(\hat{c}_o) - \hat{c}_o\right)\left(1 - H\left(P_o(\hat{c}_o)\right)\right).$$

This violates incentive compatibility, so $P_o(c_o)$ is rigid for $c_o \in \left(c^T, \min\left\{c^M, \overline{c}\right\}\right)$. ■

Proof of Theorem 1: From Proposition 1 of Melumad and Shibano (1991), the premium schedule is globally incentive compatible. From Lemma 1 and Lemma 3, the optimal incentive compatible premium has to take this form. ■

C Solving the Model

To solve the model, we use the theory presented in section 4 and formulate the profit maximization problem for insuring old consumers as follows:

$$\begin{aligned}\Pi_o^* = \max_{\bar{P}_o, \bar{\bar{P}}_o, c^T, c^M} & \int_{\underline{c}}^{c^T} \left(\bar{P}_o - c_o\right)\left[1 - H\left(\bar{r}_o\right)\right] g(c_o)\, dc_o \\ & + \int_{c^T}^{c^M} \left(\bar{\bar{P}}_o - c_o\right)\left[1 - H\left(\bar{\bar{P}}_o\right)\right] g(c_o)\, dc_o \\ & + \int_{c^M}^{\overline{c}} \left(P_o^*(c_o) - c_o\right)\left[1 - H\left(P_o^*(c_o)\right)\right] g(c_o)\, dc_o\end{aligned}$$

subject to (9) and $\bar{\bar{P}}_o = P_o^*(c^M)$. Let λ denote the Lagrange multiplier on (9). Given the distributional assumptions: $h(r_o) = \frac{1}{\gamma}\exp^{-\frac{(r_o - \theta)}{\gamma}}$ and $g(c_o) = \frac{1}{\overline{c} - \underline{c}}$, from the first-order conditions, we can derive the following premiums:

$$\bar{P}_o = \frac{\gamma}{\frac{\partial \bar{r}_o}{\partial \bar{P}_o}} + c^T - \frac{0.5(c^T - \underline{c})^2}{\lambda(\overline{c} - \underline{c}) + c^T - \underline{c}}, \tag{12}$$

$$\bar{\bar{P}}_o = \gamma + c^T - \frac{0.5(c^M - c^T)^2}{\lambda(\bar{c} - \underline{c}) - c^M + c^T}. \tag{13}$$

The monopoly premium is $P_o^*(c_o) = \gamma + c_o$. Since $\bar{\bar{P}}_o = P_o^*(c^M)$, we have

$$c^M = c^T + 2\lambda(\bar{c} - \underline{c}). \tag{14}$$

To solve for λ and c^T, we need (9) and the first-order condition on c^T :

$$\frac{c^T - \underline{c}}{\bar{c} - \underline{c}} \left[\bar{P}_o - 0.5\left(c^T + \underline{c}\right)\right] \frac{\partial \bar{r}_o}{\partial c^T} = \lambda \left[\gamma \left(e^{-\frac{\bar{\bar{P}}_o - \bar{r}_o}{\gamma}} - 1\right) - \frac{\partial \bar{r}_o}{\partial c^T}\left(\bar{P}_o - c^T\right)\right]. \tag{15}$$

To solve for the model equilibrium, we apply the following numerical algorithm:

1. Choose P_y to maximize the firm's total profit (7).
2. Given the choice of P_y, select c^T such that the FOC (15) holds under two cases:
   (a) the upper cost threshold hits a corner, i.e. $c^M = \bar{c}$.
   (b) the upper cost threshold is interior and calculated according to (14).

   Compare the resulting profit from the old (6) for each case and select the lower cost threshold c^T for which it is maximized.
3. Given $\{P_y, c^T, c^M\}$, find the value of λ for which the IC constraint (9) holds.
4. Given $\{P_y, c^T, c^M, \lambda\}$, compute the upper rigid price $\bar{\bar{P}}_o$ using (13), and select the value of lower rigid price $\bar{P}_o$ for which the FOC (12) holds.
5. Given $\{P_y, c^T, c^M, \lambda, \bar{P}_o, \bar{\bar{P}}_o\}$, find the value of $\bar{r}_o$ that makes condition (2) hold with equality. Approximate partial derivatives $\{\frac{\partial \bar{r}_o}{\partial \bar{P}_o}, \frac{\partial \bar{r}_o}{\partial c^T}\}$ necessary to compute the FOCs.

The algorithm is executed backwards, effectively nesting a sequence of five optimization or root-finding problems.

D Estimating the Marginal Cost of Life Insurance

In what follows, let $m_{t,n,\bar{n}}$ denote the period t mortality rate of age n individuals who bought life insurance at age $\bar{n}$,[28] and let N be the maximum attainable age according to the corresponding mortality tables. Let $R_t(i)$ be the (annualized) interest rate on zero-coupon risk-free securities with maturity i at time t. The schedule of net premiums for an ART policy acquired at age n for ages up to N per dollar of death benefit is defined as $\{P_t(i)\}_{i=n,n+1,...,N}$ and obtained by solving the following equation

$$\sum_{i=1}^{N-s} \frac{\prod_{j=0}^{i-2}(1-m_{t,s+j,n})m_{t,s+i-1,n}}{R_t^i(i)} = P_t(s) + \sum_{i=1}^{N-s-1} \frac{\prod_{j=0}^{i-1}(1-m_{t,s+j,n})P_t(s+i)}{R_t^i(i)} \tag{16}$$

recursively for every age $s = N-1, N-2, ..., n$. Following the method presented by Huntington (1958), we calculate the full schedule of net premiums backwards, starting from the highest admissible age. Formula (16) can further be augmented to account for two additional features of renewable term policies. First, at certain age $N_c < N$ the consumer may choose to convert to a universal life insurance and pay a fixed premium for all the remaining periods up to N. Second, the premium may be renewed at frequencies lower than one year, in particular in 5-, 10- or 20-year intervals.

Notice that formula (16) does not take into account potential lapsation of policies, that is the possibility that a consumer may choose not to renew it. This is because there is currently no industry-wide standard for insurance pricing with lapsation, and data lapsation is scarce and varies widely across different policies and time. Similarly as in Koijen and Yogo (2015, 2016), for simplicity we ignore lapsation in our analysis.

In our calculation of the net premium we use the mortality tables issued by the American Society of Actuaries. We apply the 1980 Commissioners Standard Ordinary (CSO) table for all years prior to January 2001, the 2001 Valuation Basic Table (VBT) prior to January 2008 and the 2008 VBT for the time period following January 2008. We use geometric averaging on the monthly basis to smooth the transition between any two vintages of the mortality tables. It is important to emphasize that these tables are created based on the actual mortality rates among the insured rather than the general population. For this reason, they account for a potential adverse selection in the market for life insurance.[29] For the risk-free interest rate we use the U.S. Treasury zero-coupon yield curve.[30]

[28] It is important to keep track of different cohorts of the insured due to adverse selection, i.e. individuals who have already held a policy tend to have significantly higher mortality rates than the same-age newcomers.

[29] Cawley and Philipson (1999) found no strong evidence of adverse selection in the term life insurance.

[30] Taken from Gurkaynak et al. (2007) and averaged for each month.