Review of Economic Dynamics 49 (2023) 99–122

Contents lists available at ScienceDirect

Review of Economic Dynamics

journal homepage: www.elsevier.com/locate/red

Commitment versus flexibility and sticky prices: Evidence from life insurance ☆

Radoslaw Paluszynski [a,*], Pei Cheng Yu [b]

[a] University of Houston, United States of America
[b] University of New South Wales, Australia

ARTICLE INFO

Article history:
Received 16 March 2020
Received in revised form 13 July 2022
Available online 10 August 2022

JEL classification:
G22
L11
L14

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

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 dynamic model where policyholders are held-up by long-term insurance contracts, resulting in a time inconsistency problem for the insurer. The optimal contract balances commitment and flexibility and 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.

1. Introduction

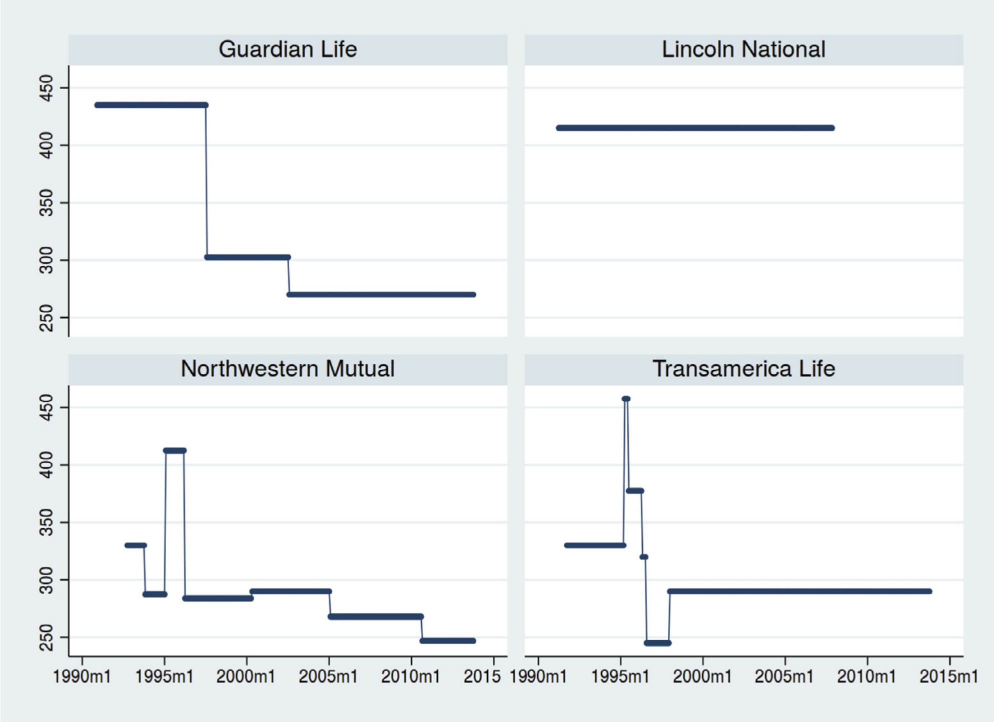
Traditional theories in finance assume that markets are efficient, so that prices of financial contracts respond to changes in the fundamentals. In contrast, this paper documents the high degree of price rigidity for a specific long-term financial contract, life insurance, the cost of which displays significant volatility. This finding has important implications for understanding the pricing of long-term financial services. We explain this unique pricing phenomenon using a quantitative model that features the commitment versus flexibility trade-off and generates price rigidity endogenously.

We show that life insurance premiums are characterized by long periods of rigidity with occasionally sizable adjustments (on average more than 10%). This is intriguing because the underlying marginal cost of life insurance is volatile over time,

☆ This paper was previously circulated under the title “Pay What Your Dad Paid: Commitment versus Flexibility in the Market for Life Insurance”. We are indebted to Motohiro Yogo for his encouragement and advice. We thank Martin Byford, Daniel Gottlieb, Igal Hendel, Kyle Herkenhoff, Anton Kolotilin, Edward Kung, Hongyi Li, Hodaka Morita, David Rahman, Breanne Richins, Jim Schmitz, Tzu-Hao Tsai, Tak Yuen Wong, and Kei-Mu Yi for many helpful comments, as well as Witold Paluszyński for his support with automating the collection of data. We also received helpful feedback from participants of the 2019 NBER Insurance workshop, 2018 EARIE annual conference, 2018 European Econometric Society summer meeting, 2018 Australasian Economic Theory workshop, 2016 Texas Theory Camp, 2014 China Econometric Society meeting, 2014 Asia Econometric Society meeting, 2014 North American Econometric Society summer meeting, 2014 Midwest Economic Theory meeting; as well as seminar participants at Academia Sinica, National Tsing Hua University, University of Technology Sydney, and Warsaw School of Economics. Pei Cheng Yu would like to thank the Institute of Economics at Academia Sinica for their hospitality during the pandemic.

* Corresponding author.
E-mail addresses: rpaluszynski@uh.edu (R. Paluszynski), pei-cheng.yu@unsw.edu.au (P.C. Yu).

https://doi.org/10.1016/j.red.2022.07.003
1094-2025/



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.

Fig. 1. ART premiums over time for selected companies (California Department of Insurance, 2004).

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). Fig. 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! More generally, our empirical findings indicate that life insurance companies tend to maintain stable profiles of premiums with respect to age. This means that over the life cycle, young policyholders will likely pay the same amount as what the older cohorts used to.

To explain the empirical findings, we construct an OLG model where the insurer faces a commitment versus flexibility trade-off, which stems from the consumer hold-up problem. Consumers live for three periods and buy one of two types of policies from monopolistically competitive insurers, renewable or non-renewable. They incur a transaction cost before purchasing, which represents the monetary expenses and opportunity cost of research and medical examination. They may also experience adverse health shocks in the second period, which could lead to significant premium hikes if they searched for a new policy. In the second period, non-renewable policyholders face all of these costs, while renewable policyholders are guaranteed coverage at no additional expense. Therefore, renewable policyholders are locked into a long-term relationship with the company, limiting their future options. This creates an incentive for the insurer to raise renewal prices, which also lowers the consumers' ex ante willingness to sign. In essence, the insurer is time-inconsistent and values commitment. The insurer also faces stochastic cost shocks in the second period, so it values flexibility. However, policyholders do not observe the shocks, so they are unsure if premium hikes are due to being held-up or due to changes in the cost.[1] Consequently, excessive flexibility in adjusting premiums could exacerbate the hold-up problem by discouraging consumers from purchasing.

To balance the need for commitment and the desire for flexibility, we show that the optimal premium as a function of cost follows a simple cutoff rule: Premiums are low and rigid for marginal costs below an endogenously determined

[1] The cost of providing life insurance is mainly determined by the mortality risk of its pool of policyholders and the interest rate. Typically, consumers do not observe the average mortality risk that the insurer bears. Section 2.4 shows how we use publicly available data to estimate the cost of providing life insurance.

threshold, while above this threshold premiums are high and initially rigid before full flexibility is possible. The reason is the renewal demand of locked-in policyholders is inelastic for low premiums. In the inelastic region, premiums can increase slightly without losing any consumers, rendering flexibility in adjusting premiums within this region non-credible. For marginal costs below the threshold, the unconstrained premiums are low and map to the inelastic region, so the insurer commits to a single low premium for all cost realizations sufficiently small to gain credibility. Premium adjustments have to be costly to the insurer for it to be credible. Therefore, increases from the low premium need to be significant to induce enough reduction in demand, which is optimal when marginal cost is large.

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. The main result is also consistent with the numerous premium drops observed in the data which may occur when the cost in the second period decreases significantly, as well as small premium changes which can be explained by the flexible part of the optimal schedule.

Having established the general properties of an optimal premium, 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 switches between the low and high regions, in line with what we observe on average in the data. 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.

In the final part of the paper we show that the life insurance premiums data supports the main predictions of our model. First, we show that as the level-term of a renewable policy increases, which weakens the hold-up problem due to a higher probability of policy termination before the renewal date, premiums are also more likely to be adjusted and exhibit smaller jumps. Second, 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. Third, we demonstrate that life insurance companies tend to respond to 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 pricing is based on a threshold rule. Fourth, we contrast life insurance premiums with prices of annuities, a related product whose buyers are not held-up by the insurer. We find that these prices adjust very frequently and by small margins, thus providing external validity to our theory. Finally, we test several alternative frictions that commonly lead to price stickiness, such as staggered contracts or menu costs, and show that their predictions are not consistent with the facts about life insurance premiums.

To summarize, our paper offers two main contributions. Empirically, we provide new evidence on the frequency and size of price changes in the life insurance market.[2] On the theoretical side, we explain this phenomenon with a model where the optimal incentive compatible contract necessarily features price rigidity and a discrete jump. 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 empirical finding provides support for a crucial assumption in the literature on life insurance contracts. In a seminal paper, Hendel and Lizzeri (2003) examine the front-loading of life insurance premiums, i.e., policyholders pay a surcharge when young to cover for expected future losses when they age. They analyze the cross-sectional data on premiums to show that when policyholders lack commitment and face health reclassification risk, the optimal insurance contracts exhibit front-loading. However, their analysis relies on the assumption that insurers keep their promises in that premiums for older cohorts are the future premiums. In essence, they 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. Several papers have since extended their framework.[3] Therefore, the findings in this paper allow us to empirically and theoretically validate the implicit assumption in Hendel and Lizzeri (2003).

Our model contributes to the literature on optimal delegation, which analyzes a principal-agent setting with no transfers and a biased agent who is better informed (Holmstrom, 1984; Melumad and Shibano, 1991; Alonso and Matouschek, 2008; Amador and Bagwell, 2013). In these models, the principal typically has full commitment and chooses a set of actions that the agent can take.[4] This is similar to our paper since the time-inconsistent insurer commits to a rule, i.e., a subset of renewal premiums that it can choose in the future.[5] In particular, our characterization of the optimal renewal premium function builds on the theoretical insights of Melumad and Shibano (1991) and Alonso and Matouschek (2008). Our paper

[2] The recent vast literature has focused on documenting the distribution of frequency and size of price changes in *consumer goods*, for example using the CPI or scanner data. On the other hand, very little such evidence is available for *financial services*, in particular in terms of the size of price changes.

[3] Daily et al. (2008) analyzed the effect of secondary markets on front-loading. Fang and Kung (2018) considered the consequences of introducing health-contingent cash surrender values, which work in a similar fashion to secondary markets. The front-loading of contracts motivated Fang and Kung (2021) to ask whether lapsation is driven by income, health or bequest shocks. Alternatively, Gottlieb and Smetters (2021) show how front-loaded contracts exploit policyholders who underestimate the probability of an adverse income shock. (Recent papers have focused on departures from the rational model, see Gottlieb (2018) for example.)

[4] The delegation framework has been applied to the analysis of savings mechanisms for present-biased agents (Amador et al., 2006), and to the optimal level of discretion for policymaking (Athey et al., 2005; Halac and Yared, 2014, 2018, 2022).

[5] Our model focuses on the disagreement between the insurer in the present and itself in the future. Compared to the present, the future insurer is biased because it wants to exploit the held-up policyholders, and it is also better informed than the present insurer because it knows the cost realization.

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.

also makes three novel contributions to the delegation literature. First, in our model, the time inconsistency of the insurer is endogenous. The insurer is able to decrease or even eliminate its intertemporal conflict, but we show quantitatively that it does not under empirically relevant parameters. This differs from the literature which analyzes an exogenously biased agent. Second, the optimal renewal premium will *always* feature a discontinuous jump if the insurer has discretion in adjusting the premiums in the future. This is in contrast to the previous literature which has found conditions for interval delegation to be optimal (Amador and Bagwell, 2013). Third, our paper provides empirical support for the trade-off between commitment and flexibility, which has not been quantified or tested in this literature.

This paper also contributes to the empirical literature on life insurance. Koijen and Yogo (2015) show that life insurers have recently been posting highly negative markups which can be explained by financial frictions around the 2008 crisis. Our paper provides an alternative theory for why many of these companies were reluctant to increase premiums in the presence of large marginal cost. Ge (2022) shows that insurers often adjust life insurance premiums in response to shocks to their divisions in other markets. Her story suggests that on their own, life insurance premiums may be even more rigid than the analysis in our paper indicates.

The remainder of the paper is structured as follows. Section 2 describes the construction of our dataset and summarizes the main findings about price dynamics 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 provides empirical support for the main predictions of the model. Section 6 concludes. The Supplementary Appendix contains the proofs of our theoretical results, 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 discuss our findings on premium rigidity and the magnitude of premium adjustments. We then show that marginal cost of life insurance is volatile over time, which presents a puzzle in light of the rigid premiums. We conclude by explaining how these findings motivate the construction of our model in Section 3.

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 one-year level-term insurance policy, commonly referred to as the Annual Renewable Term (ART), for the first 10 policy years. To help 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.

Since it is impossible for the future insurer to pay the present insurer, the only way for the present insurer to discipline the future insurer would be to restrict the renewal premiums that it can charge.

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. This in turn supports the assumption in Hendel and Lizzeri (2003) that consumers know the renewal contract upon signing.

We measure the insurance companies' adherence to these non-binding promises by collecting premiums data for a fixed-age customer, as described in the next section.[6] This approach is reasonable because we generally observe in the data that companies tend to adjust entire age schedules, rather than individual premiums separately. So while there is some measurement error involved, in Appendix A.2, we show that it is likely to be small. We thus assume that the pricing patterns for a fixed customer profile are a good approximation for how credible these non-binding projections are in the next section. For example, a 30-year-old customer signing an insurance contract as in Table 1 can expect to renew at $280 in 2005 (when he reaches the age of 31). Should the company deviate from this non-binding promise and charge him an amount greater than $280, we will observe a simultaneous change in the 30-year-old premium in our data since the entire age schedule of premiums shifts.

2.2. Data construction

We construct a sample of life insurance premiums from Compulife Software, a commercial quotation system used by insurance agents.[7] 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.[8] 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-. For the default consumer, we use a 30-year-old non-smoker male of the "regular" health category in California purchasing a policy with face value of $250,000.[9] The obtained sample consists of 55,829 observations on annual premiums for 578 different policies offered by 234 insurance companies.[10]

Naturally, over the course of 23 years, these insurers tend to disappear or merge, as well as discontinue their old products and launch new ones. We keep track of all such transformations whenever possible, merging the premium series of products with seemingly identical characteristics.[11] We also eliminate the seemingly duplicate products offered by the same company, always keeping the one with the lowest price. This is consistent with the assumption of rationality—consumers would only consider the policy that is offered at the lowest price for a fixed product characteristic. In the resulting sample, on average we observe each product for around 96 months (with a median of 84).

2.3. Historical premiums

Frequency of changes 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 2.59%, 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 could be because the insurance companies are not actively managing or promoting these products. 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 29 months.

Magnitude of changes Table 2 also shows that premium adjustments, whenever they occur, tend to be of large magnitude. We observe in total 1432 instances of premium adjustment, consisting of 580 hikes and 852 drops. The average size of these premium changes is close to 11%, although we also observe many small changes which yields a median change of around 8%. This pattern is broadly consistent with the shape of the distribution of price change sizes for CPI goods and services

[6] As explained in Section 2.2, we undertake this approach for practical reasons. Even though finding the full schedules of premium renewals with respect to age for each company in each month is theoretically possible, it would be prohibitively costly as we would not be able to automate the data collection.

[7] Since insurance agents need to provide up-to-date quotes for their consumers, Compulife reacts to changes in the life insurance market and updates the premiums accordingly, see Compulife Software, Inc. v. Newman (2017). Many of these changes are submitted by insurance companies themselves. Also, the quoted premiums are a credible source of actual pricing data. This is because it is illegal for agents to deviate from the quoted premiums in the 48 states with anti-rebating laws. Though California does not have an anti-rebating law, rebating is still prohibited due to its anti-trust statutes (Garsson, 2015).

[8] 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.

[9] The choice of this particular state is by Compulife's recommendation, due to a relatively large population and wide representation of insurance companies.

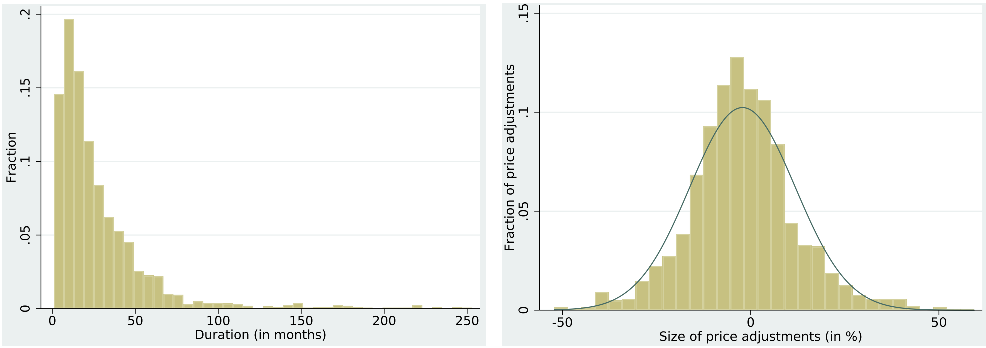
[10] Because of occasional incompleteness of Compulife data (especially in the 1990s), we impute the premiums whenever a discontinuity appears for up to at most 12 months. We also drop all the products that are observed for less than 12 continuous months. The imputed data represents roughly 1% of the final sample size and consists of 220 continuous intervals, of which about a half directly precede an observed adjustment. Our results on the overall rigidity would change very little even if we assumed that each of the remaining imputed intervals contained one premium adjustment.

[11] We do so to address a common concern that insurers may launch new products whenever they attempt to change premiums.

Table 2
Price rigidity in the sample.

Number of:			
observations	55,829	premium changes	1432
insurance products observed	578	premium hikes	580
products that change price	369	premium drops	852
Probability of price change (in %):		Whole sample	Conditional
Average		2.6	3.4
Median		1.7	3.3
Adjustment size (abs., in %):		Average	Median
All premium changes		10.7	7.9
Premium hikes only		10.6	7.4
Premium drops only		10.8	8.5

Note: The conditional sample is restricted to those products that change price at least once.



(a) Histogram of premium durations (b) Histogram of adjustment sizes

Fig. 2. Distribution of premium durations and adjustment sizes.

(Klenow and Kryvtsov, 2008), although life insurance premiums exhibit fewer small changes and fatter tails (with kurtosis of around 5.5).[12]

Illustration To visualize these findings, Figs. 2(a) and 2(b) plot 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 35% of premium spells last up to 12 months, while the majority last longer than a year, and some premiums stay constant for up to 20 years. The second chart presents the distribution of relative sizes of price adjustments, together with a fitted normal density plot. As is clear from the summary statistics in Table 2, premium drops occur more often and are of slightly larger magnitude. The size of adjustments reaches as much as 50% in both directions. The distribution also exhibits fatter tails and more concentration around zero than the normal one.[13]

2.4. Marginal cost estimation

In this section, we analyze the evolution of the marginal cost of life insurance. This is important, because premium rigidity may not appear puzzling unless we understand the dynamics of the underlying cost. Similar to Koijen and Yogo (2015), we approximate marginal cost by calculating the *actuarially fair value* of an insurance policy. A precise description of our method, applicable to renewable level-term insurance, is provided in Appendix A.3. Intuitively, actuarially fair value is a price that satisfies the insurance company's zero-profit condition and depends crucially on two factors: interest rates and mortality rates of the insured.

[12] 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 (Appendix A.4 provides a more comprehensive discussion of the effects of inflation). 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.

[13] While there are still many small price adjustments occurring in our data, the main takeaway is that the distribution of premium changes for life insurance has fatter tails than for common CPI goods and services as documented by, for example, Klenow and Kryvtsov (2008).

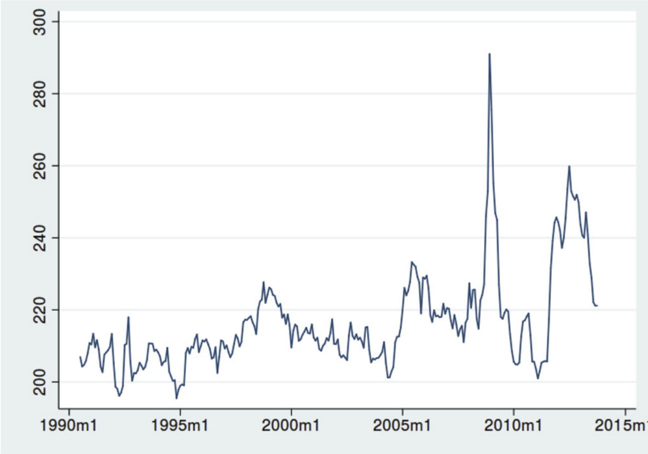


Fig. 3. Actuarially fair value for an Annual Renewable Term policy over time.

Fig. 3 plots the evolution of the actuarially fair value for an ART policy, from May 1990 until October 2013. 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. Notice the considerable fluctuations over time that result from high frequency movements in the interest rate and low frequency movements in mortality rates. A slight upward trend can be observed throughout the sample, which is a consequence of two opposing long-term empirical patterns—a decline in interest rates, and a decline in mortality of the insured. In particular, the actuarially fair value 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.

2.5. Using cross sectional data for life cycle estimations

Since this paper finds that renewable life insurance premiums exhibit extreme rigidity, it provides justification for using cross sectional data to infer premium changes over the life cycle. For example, Hendel and Lizzeri (2003) show that the optimal insurance contract is front loaded when policyholders lack commitment. However, to test their theory, they collect premiums for different ages from a fixed point in time—July 1997—to infer the change in premiums over the life cycle. In essence, they implicitly assume that, for a fixed profile of policyholders, the age schedule of premiums is held fixed or seldom adjust. Thus, our empirical finding that premiums for a fixed profile rarely change over time provides support for this methodology.

To better understand the difference, Fig. 4 illustrates the two dimensions of pricing being analyzed. Premiums are a function $P(t, a, x)$, where t is time, a is age, and x is the policyholder's initial profile—gender and health category—which is taken as given ($\bar{x}$). Fig. 4 shows how the premiums depend on t and a. The theoretical model of Hendel and Lizzeri (2003) focuses on the gray diagonal line, while their empirical exercise examines an age profile of premiums at a given point in time, $P(t_0, \cdot, \bar{x})$, which is the vertical area shaded in red. Hence, their model matches their empirical exercise only if policyholders can expect to pay the same premium in the future as the older cohorts—i.e., when the red-shaded age profile of premiums is mostly time-invariant. Our empirical analysis examines a panel of premiums for a fixed age, highlighted by the horizontal blue shading. Since our paper finds evidence of rigidity in premiums for a given age over time together with the fact that premium adjustments usually occur along the entire age schedule (see Appendix A.2), we provide empirical and theoretical support for the empirical approach in Hendel and Lizzeri (2003).

2.6. From data to model

We found that life insurance premiums tend to be rigid over time and exhibit infrequent large adjustments, while the marginal cost of issuing policies is volatile. To explain this, we develop a theory based on consumer hold-up and show that the optimal premium for renewing consumers is shaped by the trade-off between commitment and flexibility. Before formally presenting the theory, we now explain why the premiums shown in this section are the relevant premiums renewing consumers face.

Rigid premiums imply promise-keeping While our data tracks the premiums over time for a fixed-age consumer, its rigidity also implies that insurers tend to maintain their entire age profiles of premiums stable over time. This is because life insurance companies generally tend to adjust entire premium profiles, rather than individual premiums for different ages (in Appendix A.2 we discuss the potential scope for measurement error involved in this assumption). For example, referring back to Fig. 4, if premium $P(t, 31, x)$ is constant for all t, then a customer renewing at age 31 in 2011 pays the same amount as he anticipated at age 30 in 2010. In other words, the company keeps its non-binding promises towards the renewing

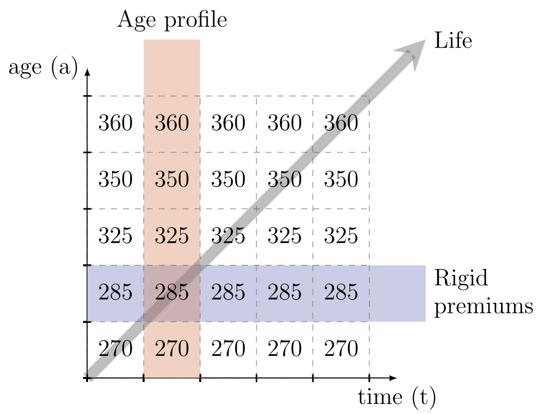


Fig. 4. Stylized illustration of premiums. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

customers. Our model, which follows a generation of aging policyholders who face renewal decisions in the future, will offer an explanation to why an insurance company has incentive to honor its non-binding commitment towards them.

Renewing vs. incoming consumers Although our data contains the initial premiums for all types of renewable policies, most renewing consumers pay the same premiums as the incoming consumers. Following the classification of Hendel and Lizzeri (2003), there are two types of renewable term policies. The first and most common type is aggregate level term policies, where premiums indeed vary in age but not across cohorts. For aggregate term policies, both renewing and incoming consumers of the same age pay the same price, so the premiums in our data are the renewal premiums. The other is Select and Ultimate (S&U) level term policies, where consumers of the same age may pay different premiums depending on when they had their last medical exam. For S&U, renewing consumers who do not get a new medical checkup pay a higher premium than newcomers of the same age. While we do not have an easy way to extract the S&U renewal premiums from our dataset, two remarks are in order. First, the supply of S&U policies is smaller, because insurers are less willing to issue it due to its high lapsation rate and severe adverse selection (Potasky et al., 1992). Second, anecdotal evidence suggests that insurance agents are reluctant to recommend S&U to consumers for fear that they may mistake S&U for aggregate (Steenwyk, 2007).

Effects of inflation 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.4, 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, which we document and verify quantitatively in Appendix A.4). 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*.

Regulation Like most of the insurance sector, life insurance companies are heavily regulated. A potential concern might be that the observations on price rigidity presented in this section are a result of the regulatory constraints. Hence, it should be emphasized that life insurance premiums are generally *not* subjected to regulatory approval of any sort, and the firms are allowed to set them freely.[14]

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 present the insurer's optimization problem. All proofs of our theoretical results are included in Appendix B.2.

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 at t as consumers of generation t. For each generation t, the young decide whether to purchase insurance at t and whether to renew, forgo coverage, or search for a new policy

[14] "State Insurance Regulation: History, Purpose and Structure", a brief by the National Association Of Insurance Commissioners.

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.

Mortality risk We assume that all young consumers are of the same health category, and face a population-average mortality risk $m_y \in (0,1)$. We denote the population-average mortality risk of the old as $m_o \in (0,1)$.

Private valuation We normalize the face value of all life insurance contracts to 1. Before purchasing life insurance, young consumers privately learn their reservation price for owning a policy when old, which is denoted as r_o. The reservation price is assumed to be the same for all consumers when young: $r_y = r$.[15] Private valuation r_o is drawn from a continuous and differentiable distribution $h(r_o)$ and c.d.f. $H(r_o)$ over support $[\underline{R}, \overline{R}]$. We assume the hazard rate is non-decreasing and $\overline{R}$ is sufficiently large so there is demand for insurance coverage when old even if the marginal cost of insurance is large. Only the distribution of r_o is common knowledge, so insurers are unable to write individual-specific contracts. Consumers have discount factor $\delta \in (0,1)$. We normalize the value of not owning life insurance to 0.

3.1.2. Life insurance company and contract

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 insurer faces a downward-sloping demand due to the imperfect substitutability of insurance policies.[16]

Marginal cost The insurer 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 $[\underline{c}, \overline{c}]$. The marginal cost for each date $c_{j,t}$, where $j \in \{y, o\}$ depends on the aggregate mortality rate and the interest rate: $c_{j,t} = \frac{m_{j,t}}{1+i_t}$, 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, movements in i_t do not affect $c_{y,t}$ much, so we assume $c_{y,t} = c_y$ for all t.[17]

A key assumption is that the cost is privately observed by the insurer. This is a natural assumption since the mortality rate of the insured pool is not observed by the consumer. Also, even if the insurer is well diversified so the mortality rate of the insured pool matches that of the population, the mapping from interest and mortality rates to the marginal cost of renewable life insurance contracts is complicated. This is evidenced by the complex formula for estimating the marginal cost of renewable policies detailed in Appendix A.3.

Life insurance contract As was mentioned, the face value is exogenously normalized to 1 for all life insurance contracts. The renewable life insurance contract consists of the premium for young consumers $P_{y,t}$ and the renewal premium $P_{o,t+1}(c_{o,t+1})$ as a *function of cost* $c_{o,t+1}$.[18] Since policyholders do not observe the cost of insuring old consumers, from their perspective, the contract consists of insurers choosing a premium from a *set of admissible renewal premiums*—the range of $P_{o,t+1}(c_{o,t+1})$. We express the set of admissible renewal premiums that generation t old consumers observe as $\{P_{o,t+1}(c_{o,t+1})\}$. Notice that by knowing $\{P_{o,t+1}(c_{o,t+1})\}$, consumers also know the renewal premium a profit maximizing insurer would choose for any c_o. We will show that $P_{o,t+1}$ is rigid and adjustments in $P_{o,t+1}$ are large, which matches the empirical evidence documented in the previous section.

3.1.3. Renewability, transaction cost and search frictions

Renewable vs. non-renewable There are two types of policies offered in the market: renewable and non-renewable. 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. For both types of policies, old consumers can lapse after one period (i.e. drop coverage altogether or switch to another insurer).[19]

Transaction cost 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

---

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

[16] 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 (see Appendix A.1). Section 3.1.3 introduces this assumption in more detail.

[17] The cost to insure the young is relatively stable in the data, for details see Figure 14 in Appendix A.3.

[18] This departs from the assumption in Hendel and Lizzeri (2003) and many papers that followed, where policyholders know the renewal premiums as a function of future health upon signing. Their paper is focused on a symmetric learning problem where the policyholder's health evolves over time. In contrast, we focus on a self-control problem where the insurer has private information on cost realizations in the future but is also biased towards exploiting the consumers.

[19] We do not consider the effects of the secondary market, because our quantitative exercise will focus on 30 and 40 year olds and the secondary market usually targets policyholders who have less than 15 years in life expectancy. See Daily et al. (2008); Gottlieb and Smetters (2021); Fang and Kung (2018); Fang et al. (2020) for in-depth analysis on long-term contract design with secondary markets.

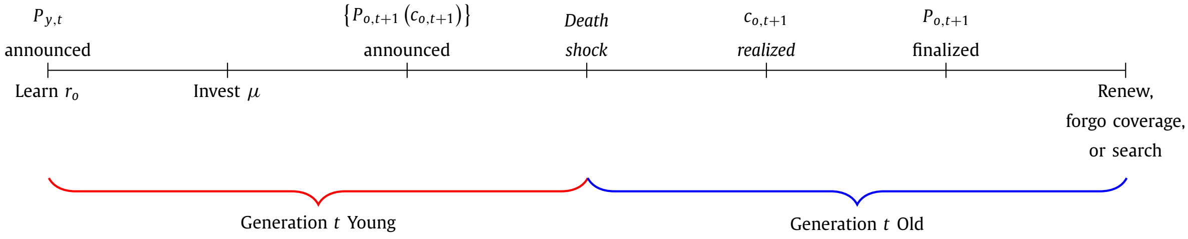


Fig. 5. Timing of events.

detailed questionnaires, as well as being exposed to the contestability period. If young consumers decide to purchase a non-renewable policy, then they must pay the transaction cost again to receive new coverage when old. This cost is avoided if consumers decide to extend their renewable policy.

Risk of searching In addition to paying μ again, old consumers do not know the non-renewable premium when they search for a new non-renewable policy. Non-renewable premiums are determined exogenously, and we assume they are equal to the marginal cost of providing the insurance, so insurers earn no profit from non-renewables. Specifically, 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 $(0, \infty)$. The shock ϵ encompasses two sources of risk associated with searching. First, a consumer faces the possibility of health deterioration which can result in much higher premium when purchasing non-renewable insurance.[20] The second source of risk comes from the search and information imperfections, resulting from ample dispersion of the premiums offered in the market, possibly coming from differences in mortality rates across varying insured pools. Hence, a consumer who decides to search may end up finding a worse alternative, even if the health status is unchanged. We will refer to ϵ as the health and search shock. Since c_y is constant, we assume that the non-renewable premium for the generation t young consumers is constant across time, $P^{NR}_{y,t} = P^{NR}_y$.

Most of the literature has focused on the optimal design of long-term contracts with risk-averse agents who face reclassification risk (Hendel and Lizzeri, 2003; Daily et al., 2008; Fang and Kung, 2018). Our paper departs from this literature by assuming that old consumers are risk neutral with respect to the risk of searching for a new policy ϵ. This is because, for a fixed transaction cost μ, the cost of switching insurers for consumers who are risk averse to ϵ is even higher than for their risk neutral counterparts. Since our theory hinges upon the fact that the consumers are held-up by the insurance company due to costly switching, assuming risk averse consumers would strengthen our theory.

3.1.4. Timing

At each t, the insurer announces $P_{y,t}$. Young consumers proceed to make their investment and purchasing decision. Prior to $t+1$, the insurer announces $\{P_{o,t+1}(c_{o,t+1})\}$. At $t+1$, $c_{o,t+1}$ is realized and the insurer sets $P_{o,t+1}$, which all surviving consumers observe. Then, existing policyholders decide whether to renew, forgo coverage, or search for a new offer.[21] Fig. 5 summarizes the timing for generation t.

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.

3.2. The insurer's problem

The environment is the same across generations, so the insurer can maximize total expected present-valued profit by optimizing the profit for each generation. Therefore, the optimal premium schedule is stationary. Subsequent analysis will thus focus on the profit maximization problem for a single generation. To simplify notation, we drop the date subscripts.

We consider a *sequentially optimal pricing rule*. The insurer chooses a premium function 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 insurer into consideration. The endogeneity of consumer behavior differs from the concept of sequential optimality in Halac and Yared (2014). To define the equilibrium, let $D_y(P_y, \{P_o(c_o)\})$ and $D_o(P_y, P_o(c_o))$ denote the demand for young and old consumers respectively. We will assume that all young consumers

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

[21] We do not model new consumers signing with the insurer when they are old. There are two reasons for this. First, if newcomers invest μ before the realization of c_o, then the hold-up problem persists, which is a likely concern since c_o is volatile as seen in Figure 14 in Appendix A.3. Second, our quantitative exercise will focus on 30-year olds renewing at 40. It is reasonable for the demand of renewing 40-year old policyholders to be larger than the demand of 40-year old new consumers. Therefore, the incentive to exploit held-up policyholders dominates the incentive to lower premiums and attract new consumers.

have a demand for life insurance coverage: $r \geq P_y^{NR} + \mu$.[22] In essence, if consumers do not purchase a renewable policy, they will purchase a non-renewable instead.

Definition 1. The sequentially optimal pricing rule is a contract $\left\{P_y, \{P_o(c_o)\}\right\}$ satisfying: (i) given P_y and the optimal response $\{P_o(c_o)\}$, consumers sign-up for renewable life insurance if and only if

$$B^{ren}(r_o) - B^{non}(r_o) \geq \frac{P_y - P_y^{NR}}{(1-m_y)\,\delta}, \tag{1}$$

where $B^{ren}(r_o)$ denotes the expected utility of an old renewable policyholder with private valuation r_o:

$$B^{ren}(r_o) = \int_{\underline{c}}^{\overline{c}} \max\left\{0, r_o - P_o(c_o), \int_{\epsilon} \max\left\{0, r_o - P_o(c_o), r_o - P_o^{NR}\right\} dZ(\epsilon) - \mu\right\} dG(c_o)$$

and $B^{non}(r_o)$ denotes the expected utility of and old non-renewable policyholder with private valuation r_o:

$$B^{non}(r_o) = \int_{\underline{c}}^{\overline{c}} \max\left\{0, \int_{\epsilon} \max\left\{0, r_o - P_o^{NR}\right\} dZ(\epsilon) - \mu\right\} dG(c_o);$$

(ii) taking P_y and the demand of renewing old consumers as given, $\{P_o(c_o)\}$ solves the static optimization problem:

$$\Pi_o(P_y) \equiv \max_{\{P_o(c_o)\}} \int_{\underline{c}}^{\overline{c}} (P_o(c_o) - c_o)\, D_o\left(P_y, P_o(c_o)\right) dG(c_o), \tag{2}$$

subject to the incentive compatibility constraints: for all $c_o, c_o' \in [\underline{c}, \overline{c}]$,

$$[P_o(c_o) - c_o]\, D_o\left(P_y, P_o(c_o)\right) \geq \left[P_o(c_o') - c_o\right] D_o\left(P_y, P_o(c_o')\right); \tag{3}$$

(iii) the premium P_y solves the optimization problem for a given c_y by taking the optimal response $\{P_o(c_o)\}$ as given:

$$\Pi = \max_{P_y} \left(P_y - c_y\right) D_y\left(P_y, \{P_o(c_o)\}\right) + \frac{1}{1+i}\Pi_o\left(P_y\right). \tag{4}$$

Part (i) of Definition 1 shows that consumers take prices as given and purchase renewable insurance if and only if the additional present value discounted expected benefit in owning a renewable when old is greater than premium difference of the two types of insurance when young.[23] By observing $B^{ren}(r_o)$, we can see that there are three benefits to purchasing a renewable policy. First, renewable policyholders do not incur a transaction cost μ if they renew. Second, renewable life insurance contracts are similar to options: Policyholders are not obligated to renew if premiums are high. They could instead forgo coverage or search for a new policy. Finally, policyholders can always renew if they are unable to find better deals after searching. From $B^{non}(r_o)$, non-renewable policyholders have the same options as renewable policyholders except for the option of renewing when they are old.

Importantly, part (i) of Definition 1 also helps us characterize the demand functions. First, notice that if $P_y \leq P_y^{NR}$, then (1) is automatically satisfied. In particular, the demand for renewables is independent of r_o, because all generation t young consumers would purchase renewables since $r \geq P_y^{NR} + \mu$. As a result, not all renewable policyholders expect to renew next period. Importantly, there is no hold-up problem if $P_y \leq P_y^{NR}$, because the renewal demand is elastic with respect to premium changes.

The more interesting case is when $P_y > P_y^{NR}$, because all renewable policyholders expect to renew. Intuitively, this implies that only consumers with large valuation for coverage when old—sufficiently high r_o—would purchase renewables and there is a lower bound in r_o for its pool of policyholders.[24] The insurer would then be tempted to increase P_o up to this lower bound, because the renewal demand is inelastic with respect to premium changes below it. The insurer encounters a hold-up problem when $P_y > P_y^{NR}$. This novel feature of our model is in contrast to the delegation literature, which has focused on exogenously determined time inconsistency that cannot be eliminated. Here, the insurer's level of

[22] This simplifies but does not change our analysis. As long as μ is large the hold-up problem persists.

[23] Since all consumers will either buy a renewable policy or a non-renewable policy, we can derive (1) from the following condition: $\left(r - P_y - \mu\right) + \left(1 - m_y\right)\delta B^{ren}(r_o) \geq \left(r - P_y^{NR} - \mu\right) + \left(1 - m_y\right)\delta B^{non}(r_o)$.

[24] Appendix B.1 provides a formal treatment of this intuitive result.

time inconsistency is determined by its own choice of premiums and can be eliminated when $P_y \le P_y^{NR}$. We will show numerically that despite this, the insurer sets $P_y > P_y^{NR}$ at the optimum.

Formally, let $\bar{r}_o\left(P_y, \{P_o(c_o)\}\right)$ be defined as the threshold valuation such that (1) holds with equality. Intuitively, this threshold is endogenously determined by the insurer's choice of premiums and is increasing with respect to both premiums. To streamline notation, we write $\bar{r}_o$ with the implicit understanding that it depends on P_y and $\{P_o(c_o)\}$. Since all consumers with $r_o \ge \bar{r}_o$ purchase the renewable insurance, the demand function of young consumers is

$$D_y\left(P_y, \{P_o(c_o)\}\right) = 1 - H(\bar{r}_o). \tag{5}$$

The demand for renewing is

$$D_o\left(P_y, P_o(c_o)\right) = \left(1 - m_y\right)\left[1 - H\left(\max\{\bar{r}_o, P_o(c_o)\}\right)\right]. \tag{6}$$

The demand is weakly decreasing in premiums. Most importantly, renewal demand becomes perfectly inelastic for any $P_o(c_o) \le \bar{r}_o$.

Next, we discuss part (ii) of Definition 1. From the analysis above, the insurer faces differing incentives before and after a young consumer purchases renewable insurance. By purchasing a renewable policy, policyholders reveal their valuation to be at least as large as $\bar{r}_o$. This discourages young consumers from buying renewable policies, since the insurer has an incentive to increase P_o up to $\bar{r}_o$. As a result, the insurer needs to commit to a low premium to attract young consumers. However, volatile cost shocks create an incentive to adjust premiums accordingly.[25] To resolve this tension, the insurer disciplines its pricing behavior by making sure that the renewal premiums are incentive compatible, i.e., the contract satisfies (3). In other words, the insurance company has a time-inconsistency problem which it manages by creating a rule for itself. The incentive constraints (3) restrict attention to renewal premium functions $P_o(c_o)$ that induce the insurer to report c_o truthfully. Effectively, the incentive constraints (3) reduce the set of admissible renewal premiums $\{P_o(c_o)\}$ that the insurer can choose from after realizing the cost. The set of admissible renewal premiums shares the same concept as *delegation sets* in Holmstrom (1984).[26] The set is announced prior to policyholders becoming old. For simplicity, we assume the insurer incurs a sufficiently large exogenous penalty for deviating from $\{P_o(c_o)\}$. In reality, the set of premiums is not legally binding, so in Appendix B.4 we explore a reputation mechanism to endogenize the penalty.

Furthermore, parts (i) and (ii) of Definition 1 also subtly require the insurer to take $\bar{r}_o$ as given and choose $\{P_o(c_o)\}$ such that (1) is satisfied for all $r_o \ge \bar{r}_o$. In essence, in addition to incentive compatibility, the insurer needs to deliver a minimal expected utility to the policyholders that corresponds to what the consumers expected when they were young.

The objective (2) 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 essence, the delegation literature has focused on the static optimization problem in (2). The novel feature of our model is that in (4), the insurer can completely eliminate its time inconsistency at a cost by setting $P_y \le P_y^{NR}$. To the best of our knowledge, we are the first to analyze such an empirically motivated two-stage delegation model: The insurer chooses the level of time inconsistency in the first stage before playing the standard delegation game in the second stage. However, we will show numerically that the trade-off in life insurance contracts is endogenously generated by the insurer optimally setting $P_y > P_y^{NR}$, so the time inconsistency problem is not eliminated at the optimum.

4. The optimal premium schedule

In this section, we characterize the set of incentive compatible renewal premiums 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.

We define the following cost regions: $\overline{\mathcal{C}}_o = \{c_o \mid P_o(c_o) \ge \bar{r}_o\}$ and $\underline{\mathcal{C}}_o = \{c_o \mid P_o(c_o) < \bar{r}_o\}$. The following lemma shows that the incentive compatible renewal premium follows a threshold rule, where it is rigid for cost shocks below the threshold.

Lemma 1. *An incentive compatible renewal premium 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 $\overline{\mathcal{C}}_o = \left[c^T, \bar{c}\right]$.*

[25] In a recent paper, L'Huillier (2020) also generates rigid prices in a model where firms are better informed of the state of economy, such as the inflation rate. He shows that prices are rigid when a sufficiently high proportion of consumers are uninformed of the state of the economy. In contrast to L'Huillier (2020), where price rigidity can be generated without long-term relationships, the rigidity in our paper relies on the hold-up problem that naturally stems from the multiple interactions between a policyholder and the insurer.

[26] Alonso and Matouschek (2008) showed that solving a direct mechanism design problem subject to (3) is equivalent to solving a delegation problem where insurers choose a delegation set to restrict the choice of renewal premiums.

Lemma 1 shows that the incentive compatible renewal premium is rigid when the marginal cost is small: $c_o < c^T$. The rigidity is caused by the inelastic renewal demand for prices below $\bar{r}_o$. The insurer knows the minimum valuation of the insured pool $\bar{r}_o$ 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 insurer would always announce the highest admissible premium below $\bar{r}_o$.

Denote $\bar{P}_o = P_o(c_o)$ for all $c_o \in \underline{\mathcal{C}}_o$. Lemma 1 allows us to rewrite some of the incentive constraints, and in particular, we have the following binding incentive constraint at c^T:

$$\left(\bar{P}_o - c^T\right)\left[1 - H(\bar{r}_o)\right] = \left(P_o\left(c^T\right) - c^T\right)\left[1 - H\left(P_o\left(c^T\right)\right)\right]. \tag{7}$$

If (7) is violated, then it is not incentive compatible for cost realizations within a neighborhood of c^T. Let $P_o^*(c_o)$ denote the optimal *frictionless premium* at c_o, which is the optimal premium when incentive constraints are not binding. The next lemma characterizes the incentive compatible premium for $\overline{\mathcal{C}}_o$.

Lemma 2. *An incentive compatible renewal premium 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 $\left(c_o', c_o''\right)$, then $P_o(c_o) = P_o^*(c_o)$ on $\left(c_o', c_o''\right)$.*
ii. *There is a discrete jump in premiums at c^T: $P_o\left(c^T\right) > \bar{r}_o > \bar{P}_o$. Furthermore, there exists $c^M > c^T$ such that $P_o(c_o)$ does not vary with $c_o \in \left[c^T, \min\left\{c^M, \bar{c}\right\}\right)$.*

Lemma 2 shows that the insurer charges the optimal frictionless price if it has full flexibility, but it is not incentive compatible for the insurer to have full flexibility for all costs in $\overline{\mathcal{C}}_o$. Crucially, Lemma 2 shows that the renewal premium function has a jump discontinuity at c^T, i.e., the set of admissible renewal premiums has a hole. In essence, the incentive compatible premium has to have a discrete jump if the insurer has flexibility to adjust premiums. What is interesting is the size of this jump. The jump at c^T is such that $P_o\left(c^T\right) > \bar{r}_o$. The reason is that for the premium increase at cost c^T to satisfy incentive compatibility (7), the premium hike has to induce a sufficient drop in demand: $1 - H\left(P_o\left(c^T\right)\right) < 1 - H(\bar{r}_o)$.

To see why the optimal renewal premium function will always contain a jump, consider a set of admissible renewal premiums without holes: $\{P_o(c_o)\} = [A, B]$. Amador and Bagwell (2013) provided general conditions for when *interval delegations* are optimal, but they fail here.[27] To see why, first notice that consumers with $r_o \le A$ would not purchase renewables, because all renewal premium realizations will be weakly greater than their private valuation. Next, notice that all consumers who purchase renewable insurance have valuation strictly greater than B, i.e., $\bar{r}_o > B$. This is because if $\bar{r}_o \in (A, B]$, then the insurer will always announce a premium of at least $\bar{r}_o$ since the demand function is inelastic below $\bar{r}_o$. As a result, by backward induction, none of the consumers with $r_o \in [A, B]$ would purchase renewable insurance. Therefore, $\bar{r}_o > B$ and the insurer would always charge a price of B. Intuitively, if the insurer chooses to implement $\{P_o(c_o)\} = [A, B]$, then the consumers believe that the insurer will exploit them because a renewal premium function that allows adjustments without jumps violates incentive compatibility.

Lemma 2 also states that incentive compatible premiums are rigid for $c_o \in \left[c^T, \min\left\{c^M, \bar{c}\right\}\right)$. This is because if the insurer chooses $P_o^*(c_o)$ for all $c_o \in \left[c^T, \bar{c}\right]$, then the insurer would deviate to the frictionless premium for cost realizations slightly below c^T. As a result, similar to the characterization of incentive compatible delegation rules with discontinuities in Melumad and Shibano (1991) and Alonso and Matouschek (2008), the insurer chooses a rigid premium within $\left[c^T, \min\left\{c^M, \bar{c}\right\}\right)$, which is the only other incentive compatible option. This implies the insurer is able to charge frictionless premiums only for sufficiently large marginal costs, because the corresponding frictionless premiums are large enough to cause a significant decrease in demand through lapsation to relax 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:*

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

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

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

[27] Halac and Yared (2020) is another paper that produces discontinuities even when conditions in Amador and Bagwell (2013) hold. Their jump is caused by a verification mechanism. Specifically, at the optimum, the principal chooses to verify the reported state when it is above a threshold. Hence, the principal's optimum is implemented for states above the threshold while the agent has some degree of flexibility for states below the threshold, which causes a discontinuity at the threshold. There is no costly state verification in our model.

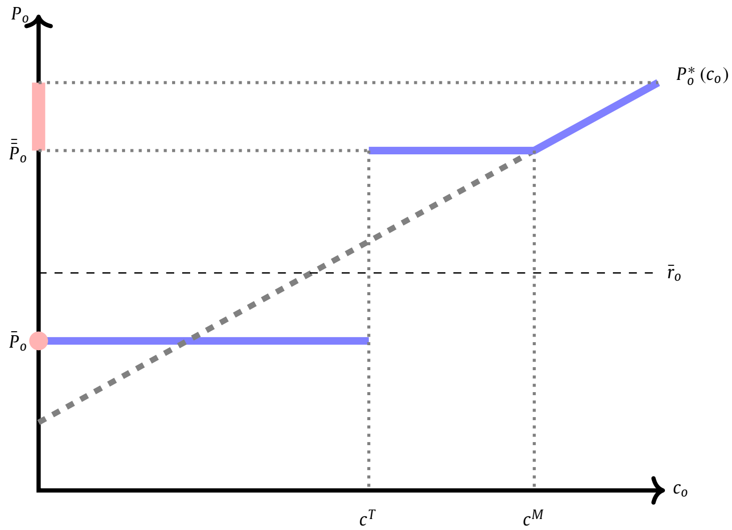


Fig. 6. Incentive compatible premium profile.

Theorem 1 shows that the incentive compatible set of admissible renewal premiums $\{P_o(c_o)\}$ is $\left\{\bar{P}_o, \left[\bar{\bar{P}}_o, P_o^*(\bar{c})\right]\right\}$, so the insurer does not announce premiums within the open interval $\left(\bar{P}_o, \bar{\bar{P}}_o\right)$. Fig. 6 illustrates an optimal renewal premium function (in blue) with the corresponding set of admissible renewal premiums on the vertical axis (in red). By Theorem 1, the insurer 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 (1) and (8).

Theorem 1 can explain the pricing phenomena documented in Section 2.[28] The low frequency of premium changes in Fig. 2(a) is explained by the intervals of cost shocks for which the optimal price is rigid. The discontinuous jump between the two rigid parts, $\bar{P}_o$ and $\bar{\bar{P}}_o$, accounts for the fact that life insurers often adjust premiums by large margins and the distribution of adjustment sizes in Fig. 2(b) exhibits fat tails. A jump can be both positive (a premium hike) or negative (a premium drop), which depends on how the future cost realization compares to what the current old generation is facing. The optimal renewal premium function also features a continuous part which allows for small changes, consistent with the high concentration of mass around zero on the histogram in Fig. 2(b). In the following Section 4.1, we take these predictions to the data by solving our model numerically and calibrating it to the US life insurance market.

4.1. Numerical analysis

4.1.1. Calibration

The model does not have an explicit closed-form solution. 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.

Marginal cost The marginal costs to insure consumers are computed directly from the data. Figure 14 in Appendix A.3 illustrates our measure of the cost shocks over time.[29] Cost of covering the young c_y corresponds to the average expected death benefit payout for a 30-year old male over the period of ten years.[30] 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 14, 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)

[28] We can enrich our baseline model with a private reservation price that is a function of the cost c_o that the insurer reports. Specifically, consumers privately learn their reservation price *function* $r_o(c_o)$, which is drawn from a known distribution $H(r_o)$. For this extension to work, we need to make two assumptions. First, the function r_o is increasing in c_o: Higher mortality rates or lower interest rates make owning life insurance more desirable. Second, an agent who values owning life insurance more does so for all costs: If $\hat{r}_o(c_o) > \tilde{r}_o(c_o)$ for some $c_o \in [\underline{c}, \bar{c}]$, then $\hat{r}_o(c_o') > \tilde{r}_o(c_o')$ for any $c_o' \in [\underline{c}, \bar{c}]$. Such an extension would not change the qualitative results presented in this section.

[29] In reference to the series presented in Fig. 3, here we consider a fixed 10-year insurance term only and do not convert the cost shocks to annual values.

[30] As discussed in Appendix A.3, we abstract here from the issue of voluntary lapsation and assume that the consumer will continue to pay the premium for the entire period.

Table 3
Parameter values in the model.

Symbol	Meaning	Value
μ	Transaction cost	970
c_y	Cost of 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		
$\underline{c}$	Lower bound	2700
$\bar{c}$	Upper bound	3500
Valuation distribution: Generalized Pareto		
γ	Scale parameter	700
κ	Shape parameter	0
θ	Threshold parameter	3580
Health and search shock distribution: lognormal		
$\bar{\epsilon}$	Mean	-0.0104
σ_ϵ	Standard deviation	0.14425

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 $[2700, 3500]$. Note that the realization of $c_{o,t+1}$ is independent from the value of $c_{o,t}$.

Private valuations 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 Appendix B.3). 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).

Health and search shock 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.[31] 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 ϵ 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.

Transaction cost 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 (given that our sample starts in the 1990s). An additional factor contributing to the dollar value of μ is getting exposed 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

[31] 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%.

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 (2018)	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.

estimates across other markets which also feature long-term contracts. Table 4 summarizes 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.

Other parameters 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 insurer 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 which encourages the insurer 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 premium for young consumers is 1052, which is greater than P_y^{NR}, so the insurer's time inconsistency problem is not eliminated. The lower cost threshold c^T is equal to 3026, just below the midpoint of the cost domain, and the upper cost threshold c^M is set to 3337. 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 3616 and 4037, 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 11.5%.

4.1.3. Comparative statics

Transaction cost We now analyze the mechanics of the model by performing several comparative statics exercises with respect to the key parameters. Fig. 7 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 insurer responds by increasing premiums (P_y, $\bar{P}_o$, and $\bar{\bar{P}}_o$) and restricting quantity, i.e. decreasing the pool of covered policyholders ($\bar{r}_o$ increases). As a result, the total profit of the insurer rises. Crucially though, an increase in the transaction cost also worsens the hold-up problem. To attract consumers *ex ante*, the insurer responds by increasing c^T (along with c^M, which rises even more) so that it is more committed to the rigid premiums $\bar{P}_o$ and $\bar{\bar{P}}_o$, and by raising the size of the jump between them. 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.

Health and search shock Fig. 8 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,

Table 5
Equilibrium of the model.

Symbol	Variable name	Value
Π	Total profit	312.73
Π_o	Profit from old	513.08
c^T	Lower cost threshold	3026.39
c^M	Upper cost threshold	3337.47
P_y	Price for young	1052.02
$\bar{P}_o$	Lower price for old	3615.87
$\bar{r}_o$	Threshold for renewables	3659.79
$\bar{\bar{P}}_o$	Upper price for old	4037.47
$\left(\bar{\bar{P}}_o/\bar{P}_o - 1\right) \times 100$	Jump between premiums (in %)	11.66
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 %)		11.43

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.

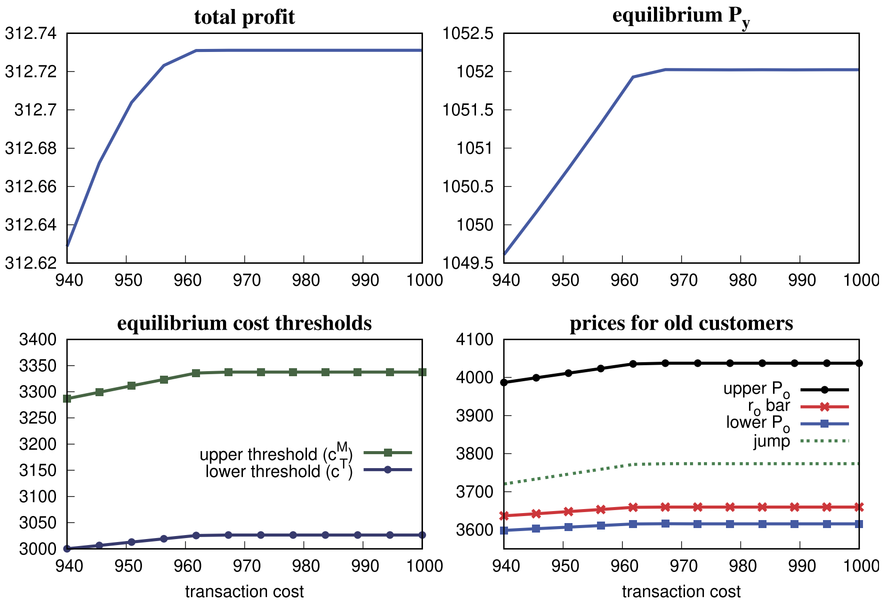


Fig. 7. Varying the transaction cost μ in the model.

the reduced competition from non-renewables leads the insurer 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 thresholds c^T and c^M go up, promising a wider interval of rigid-priced insurance, and the jump between $\bar{\bar{P}}_o$ and $\bar{P}_o$ widens.

Elasticity of demand Finally, in Fig. 9 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. As a result, the equilibrium cost thresholds fall, 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 ($c_o < c^T$) has increased with higher γ, so for (8) to hold, the profit for larger cost shocks ($c_o \geq c^T$) must also go up. Since $\bar{\bar{P}}_o$ is too high compared to the frictionless premium near c^T, the insurer lowers $\bar{\bar{P}}_o$ to increase profits for cost shocks near and above c^T.

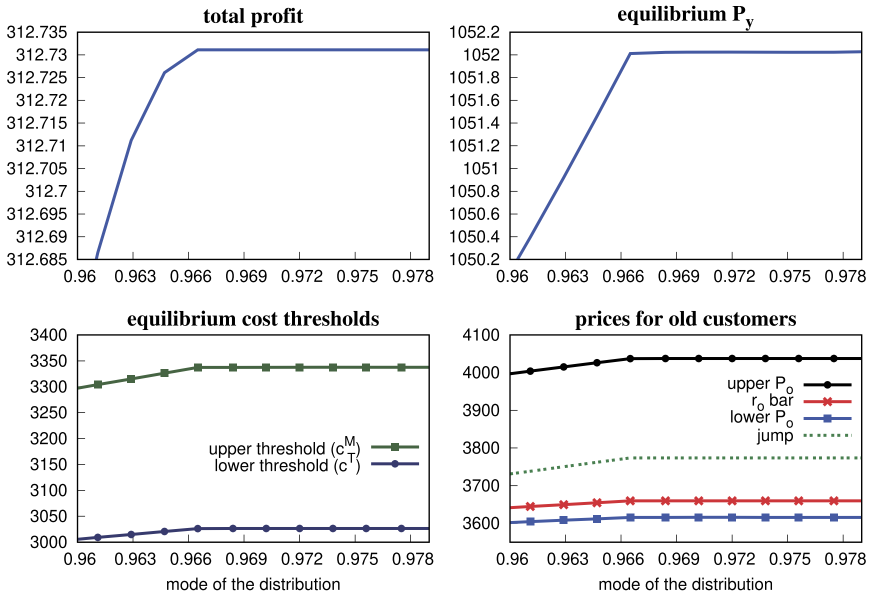


Fig. 8. Varying the mode of the health shock distribution in the model.

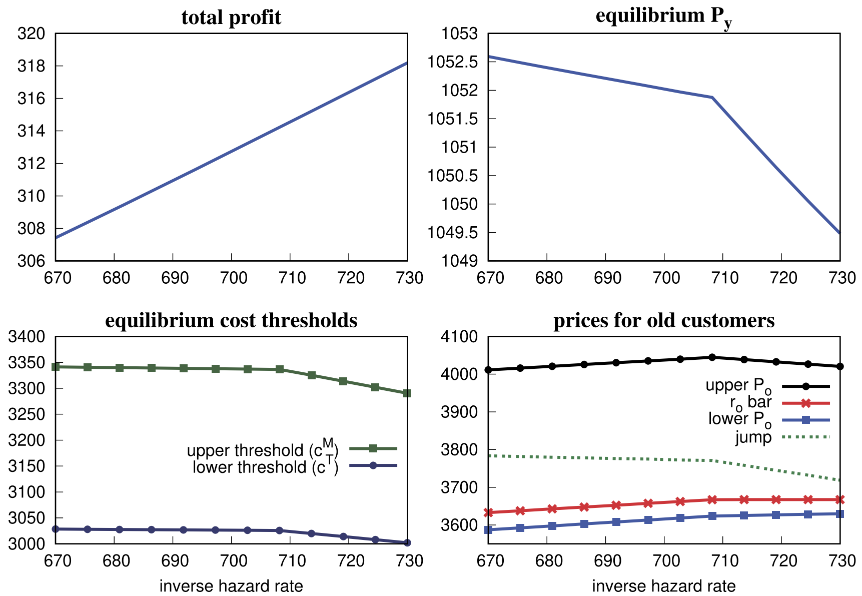


Fig. 9. Varying the inverse hazard rate γ in the model.

5. Empirical support

In this section, we show that the main predictions of our model are supported by broad trends in the life insurance premiums data. In particular, we look at how the average probability and size of premium adjustments vary across renewal terms and across time. Then, we show that the marginal cost is positively correlated with the hazard of premium adjustment, rather than size, indicating that life insurers tend to respond to cost shocks on the extensive margin. Finally, we contrast the price dynamics of life insurance with that of annuities, a related product but without the hold-up problem. We show that the latter change prices very frequently and by small margins. 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.

Table 6
Testing the difference in frequency and size of premium adjustments across terms.

Term length	N obs.	% adjusting	St. err.	Variance analysis	
1 year	13,499	1.08	0.01	MSB	1.70
5 years	6,440	2.10	0.03	MSW	0.001
10 years	20,763	3.01	0.03	F-stat	1795.85
20 years	14,549	3.61	0.03	p-value	0.00
Term length	N obs.	size (in %)	St. err.	Variance analysis	
1 year	146	15.01	0.83	MSB	1017.36
5 years	135	10.85	0.58	MSW	40.32
10 years	626	10.02	0.22	F-stat	25.23
20 years	525	10.38	0.26	p-value	0.00

5.1. Premium changes across renewal terms

First, 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 terms. 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. Fig. 7 shows that in the model this leads to a drop in both cost thresholds, more flexibility in adjusting premiums (c^M falls more than c^T), and a smaller jump between $\bar{P}_o$ and $\bar{\bar{P}}_o$.

Table 6 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 1.08% for ART policies up to 3.61% 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, term lengths above one year are 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 0.26).

5.2. Time trend in premium changes

In this section, 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 for two main 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.[32] 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 the cumulative 20-year death rate fell from 5.19% to 4.39% for a 40-year-old male policyholder. As our comparative statics exercises in Section 4.1.3 reveal, a reduction in the transaction cost, as well as a leftward shift in the distribution of health shocks in the model, both lead to more frequent (region of flexible pricing increases) and less sizable premium adjustments (the jump between the rigid premiums shrinks).

Table 7 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 2.45% to 2.78%, while the average size of the adjustment fell from 11.81% to 10.20%. Both differences are statistically significant at the 99% confidence level.

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. Fig. 10 plots the 13-month moving averages of the fraction of products that adjust their premiums and the average size of the adjustment, along with the marginal cost line from Fig. 3. Panel 10(a) shows that each of the episodes of high cost shocks (late 1990s, around 2005, late 2008, and 2012) was accompanied by an increase of at least one percentage point in the fraction of companies that adjusted their premiums, and the increase is statistically significant. On the other hand, panel 10(b) reveals

[32] 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.

Table 7
Testing the difference in frequency and size of premium adjustments over time.

Period	N obs.	% adjusting	St. err.	p-value
1990-1999	27,979	2.45	0.09	
2000-2009	21,976	2.78	0.11	
Difference		-0.33	0.14	0.01
Period	**N obs.**	**size (in %)**	**St. err.**	**p-value**
1990-1999	686	11.81	0.36	
2000-2009	612	10.20	0.37	
Difference		1.61	0.52	0.00

Note: One sided t-test for equality of means, with alternative hypotheses H_a: diff (% adjusting)<0 and H_a: diff(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.

(a) Fraction of products that adjust premium (b) Average size of premium adjustment

Note: 13-month moving averages applied. Gray areas depict the 95% confidence intervals.

Fig. 10. 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
1991-2000	0.65***	−0.07
2001-2010	0.19**	0.13
Full sample	0.28***	−0.24***

Note: correlations are between the marginal cost shock and 13-month moving averages of the frequency and size of premium changes. *** p<0.01, ** p<0.05, * p<0.1.

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 size of premium jumps discussed in the previous section.

Table 8 formalizes these findings by computing simple correlations between the cost shock and the line plotted on the two panels of Fig. 10 in different time periods. This relationship is generally positive for the fraction of insurers that adjust and was the strongest during the 1990s, when the consumer's hold-up problem was likely more severe (before the introduction of internet search which led to lower transaction costs (Brown and Goolsbee, 2002), and before a significant decline in the adversity of health shocks as measured by mortality rate). On the other hand, the average adjustment size turns out to be negatively related to the marginal cost shock (but not statistically significant). In light of our theory, this evidence suggests that life insurers respond to industry-wide shocks predominantly with an increased hazard of adjusting the premiums, rather than altering the magnitude of such an adjustment.

5.4. Comparison with annuities

So far we have demonstrated that renewable life insurance premiums are rigid, adjust by large margins and these properties tend to diminish as the hold-up problem gets weaker. It is then instructive to contrast these findings against the price dynamics of a related product that is free of the hold-up problem altogether. A good example of such a product is annuity, typically offered by the same life insurers, where consumers pay a single lump-sum amount up front in exchange for a

Table 9
Price changes in life annuities vs. life insurance: Jan 2007–Jul 2009.

	Life annuities	Life insurance
Number of observations	8140	4266
Number of insurers	19	76
Number of products	304	147
Probability of price change (in %):		
Average (weighted)	71.5	2.6
Average (unweighted) across insurers	68.7	2.6
Median (unweighted) across insurers	74.2	0
Distribution of change sizes (in %):		
Average	1.85	9.34
Standard deviation	1.56	9.18
Median	1.51	6.67

Note: the data is acquired from Koijen and Yogo (2015) and originally comes from the WebAnnuities Insurance Agency.

schedule of fixed benefits until death. Depending on policy, these benefits may carry a 10- or 20-year guarantee in case the insured dies earlier.

We use the annuity prices collected and made available by Koijen and Yogo (2015). This data was originally provided by the WebAnnuities Insurance Agency and comes at a monthly frequency from January 2007 to August 2009.[33] The quotes are available for males and females at all ages from 50 to 85 (in five year intervals). Table 9 presents the price rigidity in life annuities. Intriguingly, prices change in any given month with around 70% probability, and the magnitude of those adjustments is smaller than 2%. A natural concern then may be that this price dynamics is a result of the financial crisis in years 2007–2009. For this reason, in the right-hand side column, we also provide information about premium rigidity from our own sample, adapted to the time period of interest. The premiums during the financial crisis were as rigid as in the entire sample described in Table 2, while the adjustment sizes are slightly smaller.

We believe that the difference between the volatile pricing of annuities and the rigid pricing of life insurance supports our theory for price rigidity. It is possible that other theories of price rigidity can explain the sticky renewable prices. However, it would be difficult for other theories to explain the difference between the pricing of annuities and life insurance, which are often supplied by the same insurer. We discuss other prominent theories of price rigidity in the next section.

5.5. Alternative theories

We now investigate whether the rigidity of life insurance premiums could potentially be explained by existing models of price stickiness. We first consider the two most popular theories: Calvo-type staggered contracts and menu costs. Then, we investigate if a model of competition can generate realistic premium rigidity.

5.5.1. Staggered contracts à la Calvo

In 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 such a pricing setup, we would expect to observe variation on the *intensive margin*, and much less so on the *extensive margin* which is determined exogenously. Fig. 10 shows that the opposite appears to be the case in the life insurance market, where the firms tend to respond to cost shocks predominantly on the extensive margin.

In addition, Fig. 11 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 10 confirms this in a regression analysis. The relationship between premium duration and size of the adjustment is rather weak (albeit positive), with a slope of 0.07% and R-squared of less than 0.02.[34] We conclude that models based on Calvo-type frictions are not a promising alternative to explain premium rigidities observed in the life insurance market.

5.5.2. Menu costs

In models such as 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. The basic prediction of traditional menu cost models is that the hazard of price change is an increasing function of duration.[35] This occurs because incoming shocks move the optimal price

[33] The full dataset is semi-annual and covers the years from 1989 until 2011. However, the panel is highly unbalanced, making it difficult to make a compelling case on price rigidity.

[34] The regression results here, as well as in Section 5.5.2, are robust to controlling for cost shocks and firm fixed effects.

[35] It should be noted that several papers have recently proposed frameworks that generate different predictions. For example, Alvarez et al. (2011) feature a non-monotonic hazard function, and it is decreasing in Ilut et al. (2020).

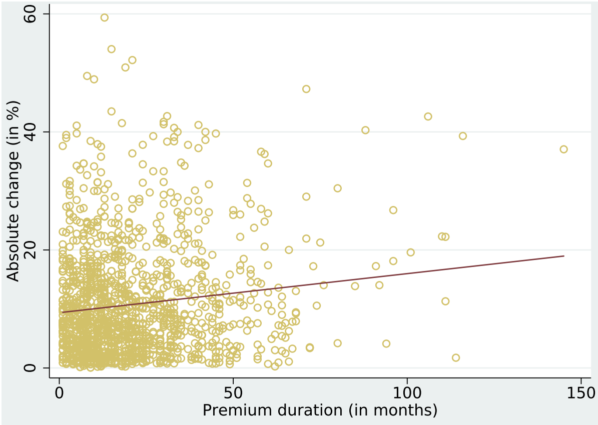


Fig. 11. Premium duration and adjustment size in the life insurance market.

Table 10
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

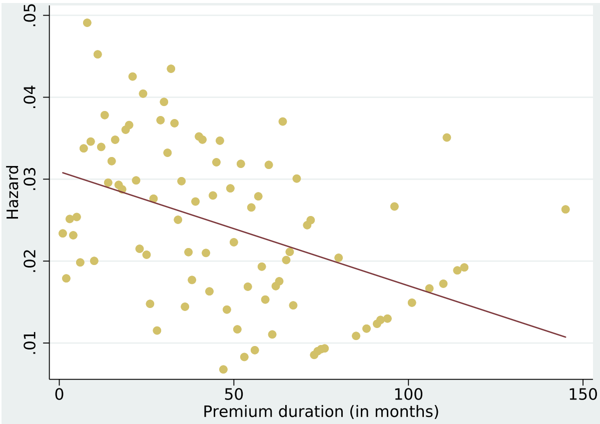


Fig. 12. Premium duration and adjustment hazard in the life insurance market.

Table 11
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

away from the one currently posted. Fig. 12 plots the hazard of premium adjustment in our dataset at all durations for which we observe multiple changes. Notice that the points are dispersed and the relationship is generally *negative*, with a correlation of −0.44. Table 11 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.

As a second step of this analysis, we approximate the size of potential menu cost that would be needed to explain the frequency and size of price changes consistent with life insurance data. Appendix B.5 describes in detail two, starkly different, quantitative exercises we conduct. In the first exercise, we use the simplest model of i.i.d. marginal cost shocks and *physical costs* to adjusting prices and apply it to our calibration from Section 4.1.1. We find that the physical adjustment costs needed to achieve the frequency and size of price changes consistent with our data correspond to 1.6-2.2% of firm

Table 12
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.

revenue. This interval, while not implausible, is considerably higher than the range of estimates found by the empirical literature (0.3-1.3%). In the second exercise, by contrast, we adapt a discrete-time version of the Alvarez et al. (2011) model based on a random walk marginal cost series and *observation costs*. We find that the observation cost needed to achieve the required frequency of price change corresponds to around 25% of firm's revenue, which is consistent with empirical estimates that show how various managerial costs related to optimizing prices tend to dwarf the physical costs of updating them. These exercises suggest that menu costs would not be a straightforward explanation for the behavior of life insurance premiums.

Furthermore, while menu costs can potentially play a role in the pricing behavior of life insurers, our paper provides a detailed theory of a type of pricing friction in the insurance industry. For financial regulators, a model of pricing that fits specifically to life insurance companies may be more useful than a model that can explain the price dynamics of any general product.

5.5.3. Competition

Our model accounts for market competition in a reduced form, allowing consumers to search for an outside option in the non-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 12 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 6, 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).

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 6 shows, even the 20-year level-term premiums are quite rigid which leaves room for complementary theories.

Economists and policymakers who study the prices of life insurance products may be tempted to conclude that this market exhibits low competition and results in suboptimal provision of risk-sharing in the economy. Such a conjecture would then naturally warrant calls for government intervention. Our work cautions against such immediate conclusions. In particular, we show that the price rigidity arises endogenously as a solution to a time inconsistency problem that could otherwise deter consumers from entering a long-term contract. In other words, in the absence of such a pricing pattern, the provision of risk sharing might be inhibited even further. 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.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.red.2022.07.003.

References

Alonso, Ricardo, Matouschek, Niko, 2008. Optimal delegation. The Review of Economic Studies 75, 259–293.
Alvarez, Fernando, Lippi, Francesco, Paciello, Luigi, 2011. Optimal price setting with observation and menu costs. The Quarterly Journal of Economics 126 (4), 1909–1960.
Amador, Manuel, Bagwell, Kyle, 2013. The theory of optimal delegation with an application to tariff caps. Econometrica 81 (4), 1541–1599.
Amador, Manuel, Bagwell, Kyle, Werning, Ivan, Angeletos, George-Marios, 2006. Commitment vs. flexibility. Econometrica 74 (2), 365–396.
Athey, Susan, Atkeson, Andrew, Kehoe, Patrick, 2005. The optimal degree of monetary policy discretion. Econometrica 73, 1431–1476.
Bils, Mark, Klenow, Peter J., 2004. Some evidence on the importance of sticky prices. Journal of Political Economy 112 (5), 947–985.
Brown, Jeffrey R., Goolsbee, Austan, 2002. Does the Internet make markets more competitive? Evidence from the life insurance industry. Journal of Political Economy 110 (3), 481–507.
California Department of Insurance, 2004. California Life & Annuity Insurance Industry: 2003 Market Share Report. Rate Specialist Bureau, Department of Insurance.
Calvo, Guillermo A., 1983. Staggered prices in a utility-maximizing framework. Journal of Monetary Economics 12 (3), 383–398.
Compulife Software, Inc. v. Newman, 9:16-cv-81942, 2017.
Cullen, Joseph, Shcherbakov, Oleksandr, 2010. Measuring consumer switching costs in the wireless industry. Unpublished manuscript.
Daily, Glenn, Hendel, Igal, Lizzeri, Alessandro, 2008. Does the secondary life insurance market threaten dynamic insurance? The American Economic Review: Papers and Proceedings 98 (2), 151–156.
Dotsey, Michael, King, Robert, Wolman, Alexander, 1999. State-dependent pricing and the general equilibrium dynamics of money and output. The Quarterly Journal of Economics 114 (2), 655–690.
Fang, Hanming, Kung, Edward, 2018. Life insurance and life settlements: the case for health-contingent cash surrender values. The Journal of Risk and Insurance 87 (1), 7–39.
Fang, Hanming, Kung, Edward, 2021. Why do life insurance policyholders lapse? The roles of income, health and bequest motive shocks. The Journal of Risk and Insurance 88 (4), 937–970.
Fang, Hanming, Kung, Edward, Wu, Zenan, 2020. Life insurance and life settlement markets with overconfident policyholders. Journal of Economic Theory, 189.
Garsson, Frederic, 2015. Five things insurers need to know before transacting an insurance business in California. Federation of Regulatory Counsel Journal 26 (4).
Ge, Shan, 2022. How do financial constraints affect product pricing? Evidence from weather and life insurance premiums. The Journal of Finance 77 (1), 449–503.
Golosov, Mikhail, Lucas, Robert E., 2007. Menu costs and Phillips curves. Journal of Political Economy 115 (2), 171–199.
Gottlieb, Daniel, 2018. Prospect Theory, Life Insurance, and Annuities. Working Paper.
Gottlieb, Daniel, Smetters, Kent, 2021. Lapse-based insurance. The American Economic Review 111 (8), 2377–2416.
Halac, Marina, Yared, Pierre, 2014. Fiscal rules and discretion under persistent shocks. Econometrica 82 (5), 1557–1614.
Halac, Marina, Yared, Pierre, 2018. Fiscal rules and discretion in a world economy. The American Economic Review 108 (8), 2305–2334.
Halac, Marina, Yared, Pierre, 2020. Commitment vs. flexibility with costly verification. Journal of Political Economy 128 (12), 4523–4572.
Halac, Marina, Yared, Pierre, 2022. Fiscal rules and discretion under limited enforcement. Econometrica. Forthcoming.
Handel, Benjamin R., 2013. Adverse selection and inertia in health insurance markets: when nudging hurts. The American Economic Review 103 (7), 2643–2682.
Hendel, Igal, Lizzeri, Alessandro, 2003. The role of commitment in dynamic contracts: evidence from life insurance. The Quarterly Journal of Economics 118 (1), 299–328.
Holmstrom, Bengt, 1984. On the theory of delegation. In: Boyer, Marcel, Kihlstrom, Richard (Eds.), Bayesian Models in Economic Theory. North Holland, New York.
Honka, Elisabeth, 2014. Quantifying search and switching costs in the US auto insurance industry. The Rand Journal of Economics 45 (4), 847–884.
Hortacsu, Ali, Syverson, Chad, 2004. Product differentiation, search costs, and competition in the mutual fund industry: a case study of S&P 500 index funds. The Quarterly Journal of Economics 119, 403–456.
Illanes, Gastón, 2016. Switching Costs in Pension Plan Choice. Unpublished manuscript.
Ilut, Cosmin, Valchev, Rosen, Vincent, Nicolas, 2020. Paralyzed by fear: rigid and discrete pricing under demand uncertainty. Econometrica 88 (5), 1899–1938.
Klemperer, Paul, 1987. Markets with consumer switching costs. The Quarterly Journal of Economics 102 (2), 375–394.
Klenow, Peter J., Kryvtsov, Oleksiy, 2008. State-dependent or time-dependent pricing: does it matter for recent U.S. inflation. The Quarterly Journal of Economics 123 (3), 863–904.
Koijen, Ralph S.J., Yogo, Motohiro, 2015. The cost of financial frictions for life insurers. The American Economic Review 105 (1), 445–475.
L'Huillier, Jean-Paul, 2020. Consumer imperfect information and endogenous price rigidity. American Economic Journal: Macroeconomics 12 (2), 94–123.
LIMRA, 1994. What's happening in the term marketplace. LIMRA's MarketFacts, p. 13.
Melumad, Nehum, Shibano, Toshiyuki, 1991. Communication in settings with no transfers. The Rand Journal of Economics 22 (2), 173–198.
Miller, Daniel P., Yeo, Jungwon, 2018. Estimating switching costs with market share data: an application to Medicare Part D. International Journal of Industrial Organization 61, 459–501.
Nishimura, Kiyohiko, 1986. Rational expectations and price rigidity in a monopolistically competitive market. The Review of Economic Studies 53, 283–292.
Nosal, Kathleen, 2012. Estimating Switching Costs for Medicare Advantage Plans. Unpublished manuscript.
Pauly, Mark, Withers, Kate, Subramanian-Viswanathan, Krupa, Lemaire, Jean, Hershey, John, Armstrong, Katrina, Asch, David, 2003. Price Elasticity of Demand for Term Life Insurance and Adverse Selection. NBER Working Paper. p. 9925.
Potasky, Sandra, Feinberg, Melvin, Katcher, Mitchel, Querfeld, Erica, 1992. Individual life product update. Record of Society of Actuaries 18 (4B), 2039–2061.
Shcherbakov, Oleksandr, 2016. Measuring consumer switching costs in the television industry. The Rand Journal of Economics 47 (2), 366–393.
Steenwyk, Jason Van, January 2007. The Decline and Fall of the Term Life Insurance Premium.
Weiergräber, Stefan, 2014. Network Effects and Switching Costs in the US Wireless Industry. Disentangling Sources of Consumer Inertia. SFB/TR 15 Discussion Paper. p. 512.