Lapse-Based Insurance

Daniel Gottlieb and Kent Smetters*

February 25, 2021

Abstract

Most individual life insurance policies lapse, with lapsers cross-subsidizing non-lapsers. We show that policies and lapse patterns predicted by standard rational expectations models are the opposite of those observed empirically. We propose two behavioral models consistent with the evidence: (i) consumers who forget to pay premiums and (ii) consumers who understate future liquidity needs. We conduct two surveys with a large insurer. New buyers believe that their own lapse probabilities are small compared to the insurer's actual experience. For recent lapsers, forgetfulness accounts for 37.8 percent of lapses while unexpected liquidity accounts for 15.4 percent.

JEL No. D03, G22, G02

*London School of Economics and Wharton School, University of Pennsylvania. Daniel Gottlieb: D.Gottlieb@lse.ac.uk. Kent Smetters: smetters@wharton.upenn.edu. This paper was previously titled "Narrow Framing and Life Insurance." We thank four anonymous referees, Nicholas Barberis, Daniel Bauer, Roland Bénabou, Pedro Bordalo, Sylvain Chassang, Keith Crocker, Kfir Eliaz, Erik Eyster, Hanming Fang, Xavier Gabaix, Nicola Gennaioli, Quentin Graham, Michael Grubb, Paul Heidhues, Botond Kőszegi, Lee Lockwood, Ted O'Donoghue, Pietro Ortoleva, Matthew Rabin, David Richardson, Andrei Shleifer, Paul Siegert, Justin Sydnor, Jeremy Tobacman, Jean Tirole, Daniel Sacks, Georg Weizsäcker, Richard Zechauser, and seminar participants at the Central European University, Cornell, EGRIE, ESMT, the European Behavioral Economics Meeting, Federal Reserve Board/George Washington University, Harvard University, the London School of Economics, NBER Insurance, NBER Household Finance, Penn State, Princeton, Rice, the Risk Theory Society, UC Santa Barbara, University of Southern California, University of Utah, University of Pennsylvania, University of Pittsburgh, Washington University in St. Louis, and the University of Wisconsin-Madison for comments. James Finucane provided outstanding research assistance.

1 Introduction

Life insurance is the most valuable method for individuals to financially protect their loved ones upon death. It is also an enormous industry. Over 70 percent of U.S. families own life insurance (LIMRA 2016), and annual premiums exceed $110 billion (IIC 2018). Between 1990 and 2017, $42.3 trillion in coverage was issued in the individual life insurance market.[1] The average face value of policy sold in 2017 was $165,460 (ACLI 2018, Table 7.2), double the median household net worth including home equity and retirement accounts (Eggleston and Munk, 2018).

However, most individual policies are terminated by the policyholder—known as "lapsing"—before the policies expire or pay a death benefit. Specifically, most *term* policies, which offer coverage for a fixed number of years, lapse before death or the end of the term. Term policies generally lapse after the policyholder stops making scheduled premium payments. At that point, the policyholder pays no future premiums and no death benefit or related value is paid back to the policyholder. Similarly, most *permanent* policies are terminated before death or their expiration at age 100 or older. Permanent policies generally lapse after the policyholder stops making scheduled premium payments or by explicitly electing to lapse. At that point, the policyholder pays no future premiums and no death benefit is paid back to the policyholder but a "cash value" might be paid.[2] For permanent policies with cash value, lapsing is called "surrendering." For both types of policies, because initial premiums exceed actuarially-fair prices early into the policy term and are surpassed by actuarially-fair prices later on, lapsing is costly to policyholders.[3] About $24 trillion of in-force coverage was dropped in the U.S. between 1990 and 2010, equal to almost 78 percent of all coverage issued during that period.[4]

Starting as far back as Linton (1932), a vast empirical literature has documented the relationship between policy terminations and other variables. But three puzzles remain (Section 2). First, the conventional view is that insurers should use front loads to reduce lapses (Hendel and Lizzeri, 2003). Without income shocks, the optimal load will be large enough to prevent any lapses, thereby enforcing continued

[1] Life insurance is also provided as an employer-based voluntary group benefit. Group policies are generally not portable across employers and, therefore, are priced differently. This paper focuses on individual (non-group) policies. 44 percent of American households have individual life policies and 49 percent have group policies. Individual policies are over twice in size than group policies, which have an average face value of $74,935 for policies sold in 2017.

[2] With many permanent policies, premiums are often collected only for part of a person's life. As a result, for the same death benefit, permanent policies are more expensive than term policies. This premium difference adds savings to a policyholder's "cash value," after front loads are deducted. The cash value typically increases for a while and eventually declines as the payment of the death benefit approaches. Upon lapsing or "surrendering," the cash value is returned, but because of the front load, the cash value is smaller than the premiums paid in excess of mortality risk. If the permanent policy is not surrendered, the death benefit is paid upon death or when the policyholder reaches age 100, 105, 110, 120, or 121.

[3] Front loads take on many forms, including level premiums, single premiums, limited-pay whole life, and decreasing term insurance policies. We could not find any insurer that offers back-loaded policies, and no major trade organization tracks sales information regarding back-loaded policies.

[4] Drops include coverage issued before 1990. In some cases, policies were dropped based on factors other than failure to pay (lapses)—for example, if the insurer believes that the policy terms were not satisfied.

participation in the insurance pool as policyholders learn more about their mortality over time ("risk reclassification").[5] With income shocks, lapses may occur in equilibrium after large shocks. Quantitatively, Hambel et al. (2017) simulate life insurance demand in a calibrated rational-expectations lifecycle model with idiosyncratic income shocks, health shocks, liquidity constraints, reclassification risk, and industry-average markups. They find that little to no lapses across a wide wealth and income distribution. Lapsing forfeits the front load, so many rational households would buy little insurance, or lapses require large swings in income.[6] This finding is also consistent with the results in Krebs, Kuhn, and Wright (2015), who model endogenously binding borrowing constraints in the context of life insurance purchases and macroeconomic shocks. Second, a related puzzle is the identity of the lapsers themselves: they are not more likely to lapse after a positive health shock, as predicted by reclassification risk. Third, it is common for policyholders to take loans against permanent policies, which, despite being fully collateralized by the life insurance policy itself, are typically more expensive than equivalent forms of secured credit. Outstanding loans totaled $135 billion in 2016.

As we show, life insurance companies make positive earnings on clients who lapse and negative earnings on those who keep their policies. However, insurers do not seem to make extraordinary profits. Rather, policyholders who lapse cross-subsidize those who do not. Making a profit from policies that lapse is a taboo topic in the life insurance industry. As one of their main trade groups recently put it, "[t]he life insurance business vigorously seeks to minimize the lapsing of policies" (ACLI 2017: 64). Still, as we show herein, insurers seem to compete on this margin. This result is the opposite of the conventional ("rational") view that insurers use front loads to reduce lapses.

We propose two behavioral mechanisms that can account for both the structure of life insurance policies and the pattern of lapses. First, many policyholders forget to pay their premiums.[7] Presumably, as a way to mitigate the effect of forgetfulness, most U.S. states impose a grace period of 30 or 31 days during which a life insurance policy does not lapse even though the payment is past due. However, despite mandating a grace period, most states do not require companies to notify customers who miss their payments. We show that, when faced with consumers who may forget to pay their premiums, equilibrium policies *endogenously* lapse after a policyholder misses a payment, and lapsers cross-subsidize non-lapsers.

In the second mechanism, individuals do not fully account for future liquidity needs when purchas-

[5]More precisely, with health shocks only, front loads will be large enough to prevent any lapses if the initial wealth is large enough. If the initial wealth is not large enough, individuals will only purchase annual policies at actuarially-fair prices in each period.

[6]In particular, younger households will often not purchase life insurance unless they have enough income. Middle-aged households will never lapse unless their income swings to very low or high amounts. Since the authors focus on individual insurance demand, they do not calculate a measure of households at different income and wealth combinations to aggregate their results. But their income and wealth bands are wide, and they conclude: "Our results show that in a neoclassical model, households have a low surrender rate since they rationally anticipate their future liquidity demands." (P. 1173)

[7]Previous work has documented the prevalence of forgetfulness in other settings, including consumption and saving decisions (Ericson, 2011; Sussman and Alter, 2012; Karlan, McConnell, Mullainathan, and Zinman, 2016). More generally, there is an emerging literature that studies individuals with imperfect memory (c.f., Bénabou and Tirole, 2016 and references therein).

ing life insurance. This mechanism can be a manifestation of either narrow framing or biased beliefs (such as overconfidence or optimism). Narrow framing states that when an individual evaluates a risky prospect, "she does not fully merge it with her preexisting risk but, rather, thinks about it in isolation, to some extent; in other words, she frames the gamble narrowly" (Barberis, Huang, and Thaler, 2006).[8] Alternatively, individuals may underweight their future liquidity needs because of overconfidence or optimism, a behavior that is prevalent in other markets.[9] In our model, consumers face two sources of risk: mortality risk that motivates the purchase of life insurance, and a possible "background" shock that produces a subsequent demand for liquidity. Examples of background shocks include unemployment, medical expenses, stock market fluctuations, real estate prices, new consumption opportunities, and the needs of dependents. Consumers correctly account for mortality risk when buying life insurance but fail to sufficiently account for background risks. Indeed, a large empirical literature reviewed later documents the strong effect that income and unemployment shocks have on life insurance lapses. Consistent with actual life insurance policies, consumers in our models are allowed to drop their policy and replace it with a new one. This lack of commitment introduces a pure state constraint that makes the optimal control problem non-standard. Our solution can be useful to other researchers interested in solving mechanism design problems with both incentive constraints and renegotiation proofness. We show that, when individuals underweight future liquidity needs when purchasing life insurance, policies offer expensive loans, endogenously lapse after large income shocks, and provide cross-subsidies from lapsers to non-lapsers.

To test these two mechanisms and other potential explanations directly, we implemented two surveys with customers of a large U.S. life insurer (abbreviated as "LLI" herein). LLI regularly receives an A.M. Best Company rating of A++ and its life insurance is widely marketed to the general U.S. population. The first survey (LLI, 2018) considers all people who purchased life insurance between October 2013 and November 2017 and were still active premium payers. This survey focuses on beliefs about lapsing. Over 94% of respondents indicated that they either did not anticipate stopping their policy before its expiration (80.4%) or had not thought about it before (12.0%). Only 2.4% of all respondents reported a 50% or greater chance of stopping their policy. In sharp contrast, based on this insurer's historical experience with these same policies, approximately 60% will likely lapse, similar to the industry average. The survey also elicits policyholders' beliefs about specific reasons for lapsing and their beliefs that others will lapse.

In the second survey (LLI, 2018), we probe the lapsing motivation of LLI policyholders who lapsed their policies between January 2012 and November 2017. Forgetfulness accounts for 37.8% of lapses

[8] See Read, Loewenstein, and Rabin (1999) for a survey on narrow framing, and Rabin and Weizsäcker (2009) for theoretical and empirical results on how narrow framing causes violations of stochastic dominance. Relatedly, in the context of health insurance, there is evidence that people weigh different contract features unevenly (Abaluck and Gruber, 2011; Ericson and Starc, 2012; Handel and Kosltad, 2015; and Bhargava, Loewenstein, and Sydnor, 2015). See also Baicker, Mullainathan, and Schwartzstein (2015) for an insurance model where buyers make behavioral mistakes.

[9] In the context of unemployment insurance, Spinnewijn (2015) finds that the unemployed vastly overestimate how quickly they will find work. Grubb (2009) shows that overconfidence accounts for the prevalence of three-part tariffs in cellular phone plans, Malmendier and Tate (2005) show that managerial overconfidence can account for investment distortions, and Ortoleva and Snowberg (2015) find that overconfidence can explain ideology and voter turnout. Bénabou and Tirole (2002) study endogenously optimistic beliefs. And, more recently, Conlon, Pilossoph, Wiswall, and Zafar (2018) and Mueller, Spinnewijn, and Topa (2019) find evidence of miscalibrated beliefs in the labor market.

while liquidity shocks account for 15.4%.

In addition to the literature noted above, our paper is related to an emerging literature that studies how firms respond to consumer biases.[10] See Ellison (2005), Kőszegi (2014), and Grubb (2015) for surveys of the behavioral industrial organization and behavioral contract theory literatures.

The rest of the paper is organized as follows. Section 2 describes key features of the life insurance industry, with a focus on those stylized facts that challenge rational models. Section 3 presents evidence on the mechanisms for lapsing based on the surveys we conducted with LLI. Section 4 presents models of a competitive life insurance market in which consumers either forget to make payments or underweight income shocks. Section 5 discusses other models, including present bias, and explains why they are unable to explain the structure of life insurance policies. Section 6 concludes. The Appendix contains an outline to Theorem 1 that requires solving a non-standard optimal control problem with incentive constraints and renegotiation proofness, as mentioned above. All proofs are contained in Online Appendix A. Online Appendix B contains more detailed information about the LLI surveys, including each question, responses tallied by question, as well as a comparison of responders and non-responders attributes using matched administrative data. Online Appendix C contains derivations corresponding to competing models referenced in the text. Online Appendix D collects data analysis related to the stylized facts discussed in Section 2, including premium information from two national insurers, which can be used to indirectly distinguish between the comparative static predictions across different models.

2 Key Stylized Facts

This section describes some important features of the life insurance industry. Rational expectations models of reclassification risk face several challenges for being the primary explanation for life insurance contracts and the patterns of lapsing observed in practice. This contrasts with health insurance markets, where there is evidence for the importance of reclassification risk (Handel, Hendel, and Whinston 2015; Diamond et al. 2018).

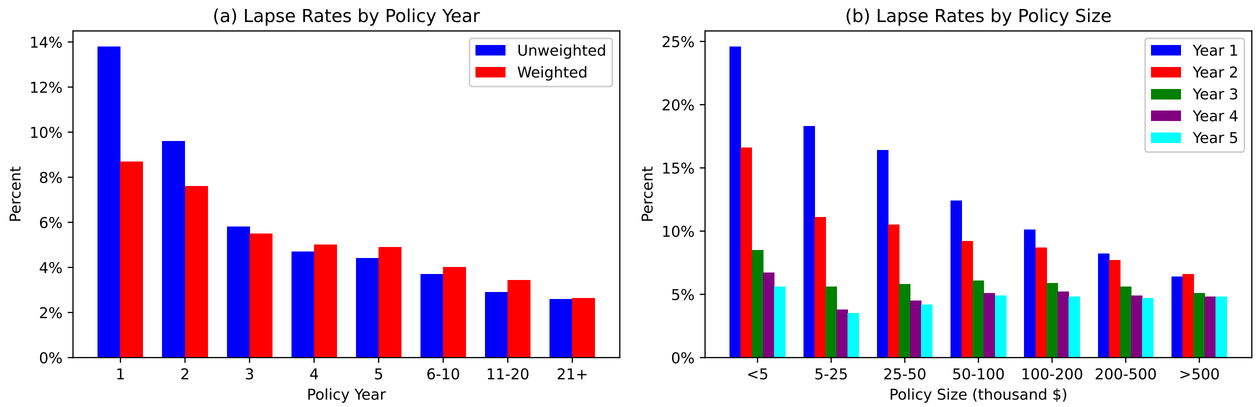
2.1 Substantial Lapsing

LIMRA, a large life insurer trade association, and the Society of Actuaries define an insurance policy lapse as "termination for nonpayment of premium, insufficient cash value or full surrender of a policy, transfer to reduced paid-up or extended term status, and in most cases, terminations for unknown reason" (LIMRA 2011A, P. 7).

As Figure 1(a) shows, 29% of *permanent* insurance policyholders lapse within just three years of first purchasing the policies; within 10 years, 57% have lapsed. In particular, nearly 88% of universal

[10] For example, Squintani and Sandroni (2007), Eliaz and Spiegler (2008), and Grubb (2009) study firms that face overconfident consumers. DellaVigna and Malmendier (2004), Eliaz and Spiegler (2006), and Heidhues and Kőszegi (2010) consider consumers who underestimate their time inconsistency. Eliaz and Spiegler (2011) and Bordalo, Gennaioli, and Shleifer (2016) study competition in markets where consumer attention is endogenously determined.

Figure 1: Annual Policy Lapse Rates by Different Characteristics



Notes: Figure (a) shows annual lapses of permanent insurance policies by policy year (the number of years since policy was purchased). Lapse rates are shown as a simple (unweighted) rate as well as weighted by face value. Figure (b) shows annual lapse rates of permanent policies by size and policy year. Source: LIMRA (2012).

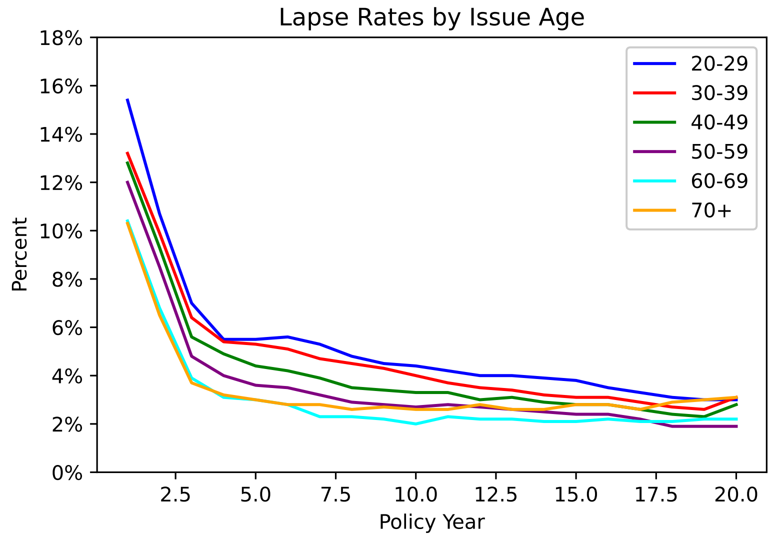
life policies, a popular type of permanent insurance, do not terminate with a death benefit claim. While the majority of policies issued are permanent, the majority of face value now takes the term form (ACLI 2018, Table 7.2), which contractually expire after a fixed number of years if death does not occur. *Term* policies lapse at an even higher annual rate, at about 6.4% lapse each year. In fact, for policies sold to seniors at age 65, 74% of term and 76% of permanent (universal life policies) never pay a claim (Millian USA 2004). As noted in Section 1, these lapse rates are substantially larger than predicted by realistic calibrations of rational models.

2.2 Who Lapses

Using the comparatively old population in the Health and Retirement Study (HRS), where health shocks are likely to be more prevalent, Fang and Kung (2012, P.11) find that lapses are uncorrelated with health shocks. In fact, while not statistically significant at conventional levels, people who lapse after a health shock tend to be *less healthy* than those who keep their policies.[11] This result is more consistent with the need for liquidity to cover medical expenses than with reclassification risk, according to which healthier people would drop their coverage to buy a cheaper policy. Online Appendix D updates their analysis for more recent waves in the HRS and, quite consistently, shows that lapsing tends to *increase* (not decrease) after an increase in the number of health conditions, although this relationship is again not statistically significant. Similarly, also using data from the HRS, three recent papers find that lapses of *long-term care* policies are not driven by reclassification risk, but are correlated with other factors such as the need

[11] See, in particular, their Table 6 (p. 11), which shows the determinants of lapses in a multinomial logit regression. As they argue, "individuals who have experienced an increase in the number of health conditions are somewhat more likely to lapse all coverage, though the effect is not statistically significant." In their structural model, which assumes that individuals choose coverage rationally, they find that younger individuals (among the relatively old population in the HRS) mostly lapse due to i.i.d. shocks. As individuals age, however, the importance of health shocks grows.

Figure 2: Annual Lapse Rates by Age of Buyer



Note: Annual lapse rate per policy year for whole life policies by the number of years since policy was purchased. Each line represents an average of lapse rates within the shown age range. Source: LIMRA (2012).

for resources or cognitive ability (Konetzka and Luo 2011, Basu 2016, and Friedberg et al. 2017).

Additional evidence points to the role of liquidity needs. Figure 1(b) shows that lapses are more prevalent for *smaller* policies, which are typically purchased by lower-income households who are more exposed to liquidity shocks. Similarly, Figure 2 shows that *younger* policyholders lapse almost three times more often than older policyholders, presumably reflecting less precautionary savings when young. Consistent with liquidity needs, lapse rates also vary with the business cycle, increasing during recessions.[12] While this evidence is indirect, Section 3 directly tests the mechanisms behind lapses using LLI survey data.

2.3 Lapsed-Supported Pricing

There is substantial evidence that insurers take profits from lapses into account when setting their premiums. Dominique LeBel, actuary at Towers Perrin Tillinghast, defines a lapse-supported product as a

[12] See Outreville (1990) and Kuo, Tsai, and Chen (2003) for studies using aggregate data from the U.S. and Dar and Dodds (1989) for British data. Lapse rates spike during times of recessions, high unemployment, and increased poverty. For example, while $600B of coverage was dropped in 1993, almost $1 trillion was dropped in 1994 (a year with record poverty) before returning to around $600B per year through the remainder of the decade. After the 2000 stock market bubble burst, over $1.5 trillion in coverage was forfeited, more than double the previous year (ACLI 2011). Similarly, Hoyt (1994) and Kim (2005) document the importance of unemployment for surrendering decisions using firm-level data. Jiang (2010) finds that both lapsing and policy loans are more likely after policyholders become unemployed. Liebenberg, Carson, and Dumm (2012) find that households are twice more likely to surrender their policy after a spouse becomes unemployed. Fier and Liebenberg (2013) find that the probability of voluntarily lapsing a policy increases after large negative income shocks, especially for those with higher debt. Using detailed socio-demographic data from Germany, Inderst and Sirak (2014) find that income and unemployment shocks are leading causes of lapses. They also find that the correlation between age and lapses disappears once one controls for income shocks and wealth.

"product where there would be a material decrease in profitability if, in the pricing calculation, the ultimate lapse rates were set to zero (assuming all other pricing parameters remain the same)." (LeBel, 2006) Precisely measuring the extent to which life insurance policies are lapse-supported is challenging since insurers do not report the underlying numbers. One reason is regulatory: for determining the insurer's reserve requirements, the historic NAIC "Model Regulation XXX" discouraged reliance on significant income from lapses for those policies surviving a certain threshold of time.[13] A second motivation is competitive: insurers are naturally tight-lipped about their pricing strategies.

Nonetheless, numerous pieces of evidence confirm the widespread use of lapse-supported pricing. We start with some anecdotal evidence. Like economists, actuaries employed by major insurers write papers and give seminars to their peers. The Society of Actuaries 2006 Annual Meetings held a session on lapse-supported pricing that included presentations from actuaries employed by several leading insurance companies and consultants. Kevin Howard, Vice President of Protective Life Insurance Company, for example, demonstrated the impact of lapses on profit margins for a representative male client who bought a level-premium secondary guarantee universal life policy, with the premium set equal to the average amount paid by such males in August 2006 in the company's sample. Assuming a zero lapse rate, the insurer projected a substantial negative profit margin, equal to -12.8%. However, at a typical four percent lapse rate, the insurer's projected profit margin was +13.6%, or a 26.4 percentage point increase relative to no lapsing.[14] Similarly, at the 1998 Society of Actuaries meeting, Mark Mahony, marketing actuary at Transamerica Reinsurance, presented calculations for a large 30-year term insurance policy often sold by the company. The insurer stood to gain $103,000 in present value using historical standard lapse rate patterns over time. But, if there were no lapses, the insurer was projected to *lose* $942,000 in present value. He noted: "I would highly recommend that in pricing this type of product, you do a lot of sensitivity testing." (Society of Actuaries 1998, p. 11)[15]

In Canada, life insurance policies are also supported by lapsing.[16] As A. David Pelletier, Executive Vice President of RGA Life Reinsurance Company, argues:

> What companies were doing to get a competitive advantage was taking into account these higher projected future lapses to essentially discount the premiums to arrive at a much more competitive premium initially because of all the profits that would occur later when people lapsed. (Society of Actuaries 1998, P. 12)

Other anecdotal evidence comes from bankruptcy proceedings, which often force a public disclosure

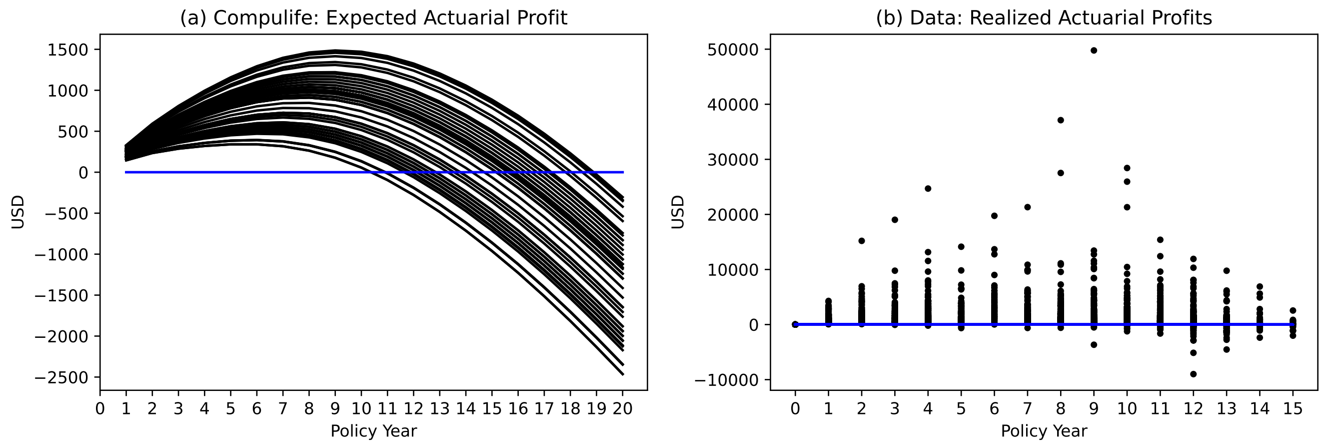
[13] Most recently, principles-based regulations (PBR) have emerged, which are widely regarded to allow for more consideration of policy lapses for purposes of reserve calculations. European insurers are generally allowed to consider lapse income under Solvency II requirements (Society of Actuaries, 2015, p. 31).

[14] For less popular single-premium policies, the swing was lower, from -6.5% to +8.7%.

[15] In explaining the rise in secondary life insurance markets, where primary policies can be resold other firms, the National Underwriter Company writes: "Policy lapse arbitrage results because of assumptions made by life insurance companies. Policies were priced lower by insurance companies on the assumption that a given number of policies would lapse." (NUC 2008, P.88)

[16] See, for example, Canadian Institute of Actuaries (2007).

Figure 3: Actuarial Profits by Policy Year



Notes: Figure (a): Each black line corresponds to a 20-year term policy with $500k coverage across 56 insurers operating in the State of California. The expected actuarial profit in each period corresponds to the difference between the present value of premiums and the expected death benefit up to and including that period. The calculations above assume a nominal interest rate of 6.5%. Results for different interest rates are qualitatively similar and are presented in the Online Appendix D. Source: Compulife (2013) quotation system, retrieved July, 2013. Figure (b): Each point represents the realized actuarial profit, calculated under the same assumptions, for 2,854 term policies of varying policy terms (10-year, 15-year, 20-year and 30-year) that lapsed between January 2012 and November 2017. Policy year represents the number of years before the policy lapsed for the shown point. Data provider is discussed in Section 3. Online Appendix D further decomposes the results based on the length of the policy terms.

of pricing strategies in order to determine the fair distribution of remaining assets between permanent life policyholders with cash values and other claimants. For example, the insurer Conseco relied extensively on lapse-based income for their pricing; they also bet that interest rates earned by their reserves would persist throughout their projected period. Prior to filing for bankruptcy, they tried to increase required premiums—in fact, tripling the amounts on many existing customers—in an attempt to effectively reduce the cash values for their universal life policies (and, hence, reduce their liabilities). In bankruptcy court, they rationalized their price spikes based on two large blocks of policies that experienced lower-than-expected lapse rates (InvestmentNews 2011).[17] Bankruptcy proceedings have also revealed substantial lapse-based pricing in the long-term care insurance market (Wall Street Journal 2000); most recently, several large U.S. long-term care insurers dropped their coverage without declaring bankruptcy, citing lower-than-expected lapse rates, which they originally estimated from the life insurance market (InvestmentNews 2012).

To more formally evaluate the importance of lapse-supported pricing, we gathered data from Compulife (2013) in 2013, a quotation system for American life insurance companies. In calculating in-

[17] Premiums for universal life permanent policies can be adjusted under conditions outlined in the insurance contract, usually pertaining to changes in mortality projections. However, in this case, the bankruptcy court ruled that the Conseco contract did not include provisions for adjusting prices based on lower interest rates or lapse rates. Conseco, therefore, was forced into bankruptcy.

surance profits, we used the most recent Society of Actuaries (2008) mortality table that existed before 2013. These tables are based on the actual mortality experience of insured pools to correct for selection; they are also used by insurers for regulatory reporting purposes. Our calculations are discussed in more detail in Online Appendix D. The results confirm an enormous reliance on lapse income. Consider, for example, a standard 20-year term policy with $500,000 in coverage for a 35-year old male in good health ("preferred plus" category). Figure 3(a) shows the projected actuarial profits for 56 such policies available in February 2013 in the state of California.[18] All 56 insurers are projected to earn between $177 and $1,486 in present value if the consumer lapses between the fifth and the tenth years of purchasing insurance while losing between $304 and $2,464 if the consumer never lapses. Figure 3(b) reports similar calculations with the universe of term policies from LLI that lapsed between January 2012 and November 2017 (see Section 3). Out of the 2,854 policies that lapsed in this period, only 121 (about 4.2%) lapsed with negative actuarial profits.

2.4 Expensive Borrowing

Instead of lapsing, permanent policies often allow policyholders to borrow against a cash value. Policy loan balances totaled $135 billion in 2016 (ACLI, 2017, Table 2.1 and P. 10). Despite being fully collateralized by the policy, these loans are usually more expensive than other forms of secured credit. For example, during the first quarter of 2018, many cash-value plans offered fixed annual borrowing rates around 8% (Gigante 2018). In contrast, a 30-year fixed-rate U.S. mortgage averaged around 4.25% at the time, while a 10-year Treasury constant maturity rate was around 2.75%. As we show in Section 4.2.5, a model with rational expectations and liquidity shocks predicts that consumers with *small* enough shocks lapse, while those with large shocks borrow from their policies. Moreover, policies offer loans at *subsidized* rates to promote consumption smoothing. In contrast, empirically, loans are offered at above-market rates up to a cap.

3 Survey Evidence

To elicit policyholder ex-ante beliefs about the probability of lapses and the ex-post reasons for lapsing, we developed and implemented two surveys with customers from a large life insurer that wished to remain anonymous ("LLI"). LLI is a large U.S. life insurer that regularly receives an A.M. Best Company rating of A++. LLI life insurance is widely marketed to the general U.S. population. Over the years, LLI has developed considerable expertise in surveying its clients, including how to word questions to avoid confusion. The first survey (LLI, 2018) was delivered to all customers from LLI who purchased term life

[18] We chose California because it is the state with the largest number of available policies. The coverage level was set to the Compulife software's default level ($500,000). The extent of lapse-based pricing, however, is extremely robust to different terms, ages, coverage levels, and states.

insurance between October 2013 and November 2017.[19] The second survey (LLI, 2018) was delivered to all customers from LLI who lapsed on their life insurance policies between January 2012 and November 2017. Online Appendix B presents each survey in detail, including comparing key characteristics of responders and non-responders using LLI matched administrative data.

3.1 New Buyers Survey

Our first survey examines the attitudes of recent buyers of term life insurance. The survey asked customers up to eight questions regarding their beliefs about their chances of lapsing, potential reasons for lapsing, beliefs about future income shocks, and their beliefs about *other* people lapsing. We also have detailed matched administrative data about customers and their policies, including gender, age, risk class, marital status, education, employment sector, job tenure, and policy type and size. The survey was accessible through an e-mail sent to all customers who purchased life insurance between October 2013 and November 2017 and were still active premium payers. Those who did not respond were sent two reminders, the first one a week after the original email was sent and the second one a day before the survey deadline.[20] The response rate was around 13% (1,689 responders), in line with LLI's typical rate.

For our purposes, survey Question 1 is the most important, which asks:

> *1. Your term life insurance policy has about N years left on it. What is the chance that you might stop your policy (sometimes called lapsing) before then?*

The value N was set equal to the actual value for that customer. As options, participants could choose (with allocated percent of answers indicated along with number of responders):

> *1.1. I have not given it much thought (12.0%, 202 responders)*
> *1.2. I do not currently anticipate stopping my policy (80.4%, 1357)*
> *1.3. I currently anticipate stopping my policy with a 10 percent or lower chance (1.4%, 24)*
> *1.4. I currently anticipate stopping my policy with a chance of 10 – 25 percent (1.7%, 28)*
> *1.5. I currently anticipate stopping my policy with a chance of 25 – 50 percent (2.1%, 35)*
> *1.6. I anticipate stopping my policy with a chance greater than 50 percent (2.4%, 41)*

Notice that 94.2% of respondents[21] indicated that they did not anticipate stopping their policy (80.4%) or had not thought about it before (12.0%). Only 6.2% of responders believed that they had a 10% or greater chance of lapsing, and only 2.4% indicated having a 50% or greater chance of stopping their

[19] We only survey those who purchased term policies because LLI permanent policies include a unique type of universal life policy that is being used by LLI's wealth management group as a tax-efficient investment vehicle for people who have maxed out their tax preferred (e.g., 401(k), 403(b), or IRA) contributions. Officials at LLI pointed out that their unique investment-focused permanent policies are substantially different from their other permanent policies. In contrast, all LLI's term policies provide a pure form of life insurance, allowing us to test our models more directly.

[20] Ideally, one would want to elicit beliefs at the moment of buying insurance. However, concerns about how asking these questions might affect their purchasing decision prevented us from being able to implement this approach. We, therefore, focused on customers who bought insurance recently.

[21] The 95% confidence interval is (93.1%, 95.3%).

policy.[22] In contrast, LLI's historical experience with these policies suggests that lapses are similar to the industry as a whole, averaging 5.2% per year during the past 15 years on these types of policies, with an average face value of $516,585 (median $500,000). As a back-of-the-envelope calculation, suppose policyholders face a constant lapse rate of 5.2% per year. Since policies in our sample have, on average, 16.2 years left, approximately 60% of these policies will lapse.[23]

Question 3, shown in more detail in Online Appendix B, asks a similar question but about the respondent's beliefs about the chance of *other* people people lapsing. Interestingly, while only 6.2% of responders believed that they had a 10% or greater chance of *themselves* lapsing (Question 1), this value increases to 18.6% for their beliefs of *other* insured people lapsing (Question 3). Moreover, the portion indicating that they have not "given it much thought" increases from 12.0% for their own policies (Question 1) to 67.1% for policies of *other* people (Question 3). Therefore, while most respondents appear to be aware of the possibility of lapses, most do not believe that it is very relevant for their own situation.

We also asked the respondent's beliefs that they would lapse broken down by three particular reasons: (i) "because [they] will need the money, maybe due to lower income or increased expenses"; (ii) "due to divorce or death of a spouse"; and (iii) "because [they] feel healthier than expected and would prefer to purchase a different policy." 81.9% of respondents indicated a 5% chance or less of stopping their policy due to a need for money (63.6%) or had not thought about it before (18.3%). 89.7% of respondents reported a 5% chance or less of stopping their policy due to divorce of death of a spouse (55.0%) or had not thought about it before (34.7%). And 94.3% reported a 5% chance or less of stopping their policy because they felt healthier than expected and would prefer to purchase a different policy (70.2%) or had not thought about it before (24.1%).

We also asked customers about their income fluctuations. Out of the 1,689 survey respondents, 32.5% reported an income loss in the last 5 years, whereas 42.8% estimated a 5% or greater chance "that at some point in the next 5 years, your total household income would decrease substantially" and 27.4% estimated a 10% or greater chance. The prevalence of income losses and their expectation, however, did not translate into significant beliefs about lapsing. For those individuals indicating a 10% or greater chance of a substantial income loss during the next 5 years, 90.4% indicated that they did not plan on stopping their policy (80.3%) or had not thought about it before (10.1%). In fact, beliefs about future income losses are not only uncorrelated with one's beliefs about lapsing in general but also with beliefs about lapsing due to income losses. These results suggest that individuals were not overconfident or optimistic about future income shocks. Instead, they appeared not to take their beliefs about future income shocks into account when evaluating their chances of future lapses, a finding that is consistent with a narrow framing mechanism.

[22]These beliefs were consistent across types of policies, including 30-year terms, where a 94.7% of respondents indicated that they did not anticipate stopping their policy (78.9%) or had not thought about it before (15.8%). Only 1.2% of those holding a 30-year term policy indicated having a 50 percent or greater chance of stopping their policy.

[23]Taking into account the entire distribution of years left (rather than using the average) while keeping the assumption of a constant annual 5.2% chance of lapsing, we find that approximately 57% of these policies would lapse.

3.2 Recent Lapsers Survey

Our second survey was sent to 3,229 former LLI policyholders who lapsed their policies between January 2012 and November 2017. Maybe not surprisingly, the response rate was at lower than with active customers (4.9 percent, or 157 responders), given the separation in relationship. For our purposes, survey Question 1 is the most important, which asks:

> *1. You have recently canceled (or let "lapse") your life insurance policy. Many people cancel / lapse their policies for one or more of the reasons listed below. Which choice best reflects your reason?*

As options, participants could choose (with allocated percent of answers indicated):

1.1. My income decreased (7.6%, 12 responders)
1.2. I needed the money (1.3%, 2)
1.3. My family situation changed due to divorce (4.5%, 7)
1.4. My family situation changed due to death of spouse (0.6%, 1)
1.5. I recently retired (12.7%, 20)
1.6. I was healthier than expected and bought another policy (1.3%, 2)
1.7. I forgot to make my insurance premium payments (12.1%, 19)
1.8. I believe that I didn't cancel my policy (12.7%, 20)
1.9. Other (please explain) (47.1%, 74)

Income or cost shocks (1.1 and 1.2) and forgetfulness (1.7 and 1.8) directly accounts for about 33.7% of answers. Divorce or death of a spouse (1.3 and 1.4) accounts for 5.1% of the answers, while health risk reclassification (1.6) accounts for 1.3%. Notice that the open-ended answer "Other" accounts for almost half of answers at 47.1%. While somewhat subjective, many of the explanations for 1.9 can be bucketed into one of the 1.1 - 1.8 answers including, in some cases, policyholders who accidentally lapsed and wanted a chance to explain the details. Based on the answers, we also added a new answer: "1.10. Family situation changed for reasons other than divorce or death of spouse." So, we did our best to recode the written explanations given for 1.9 to answers 1.1 - 1.9 and 1.10. The online data store presents the original data file we were given by LLI, which includes the raw answers reported above, the written explanations provided under "1.9 Other" and our subjective recoding, added as a separate column. Our new mapping is as follows (with allocated percent of answers indicated):

1.1. My income decreased (8.3%, 13 responders)
1.2. I needed the money (7.1%, 11)
1.3. My family situation changed due to divorce (4.5%, 7)
1.4. My family situation changed due to death of spouse (0.6%, 1)
1.5. I recently retired (13.5%, 21)
1.6. I was healthier than expected and bought another policy (6.4%, 10)

1.7. I forgot to make my insurance premium payments (23.1%, 36)

1.8. I believe that I didn't cancel my policy (14.7%, 23)

1.9. Other (please explain) (12.8%, 20)

1.10 Family situation changed for reasons other than divorce or death of spouse (9.0%, 14)

Under the recoding map, income or cost shocks (1.1 and 1.2) account for 15.4 percent of lapses and forgetfulness (1.7 and 1.8) accounts for 37.8 percent of lapses, thereby suggesting that most lapses are not consistent with the rational model which predicts very little lapsing even in the presence of realistically calibrated shocks (Section 1). While the sample size is small, the margin of error is still just 7.8%, suggesting a large role for shocks and forgetfulness. Moreover, 82.5% (130 out of 157) responders indicated that they did not purchase a new policy after their previous policy lapsed. Of those 17.5% (27 responders) that did, 9.1% (14) purchased a smaller policy while 8.4% (13) purchased a larger one.

4 Theory

The results from the surveys described in Section 3 indicate that two behavioral mechanisms—forgetting to pay and unanticipated income shocks—play important roles in explaining empirical lapses. In this section, we present competitive life insurance models in which policyholders display each of these two behavioral mechanisms. Both models predict endogenous lapsing, with lapsers cross-subsidizing non-lapsers. The models differ in terms of the possibility of borrowing from a policy. With forgetfulness, equilibrium policies always lapse if the benefit is not paid. With income shocks, policies also allow for loans. As noted in Section 2.4, policy loans total $135 billion in 2016.

4.1 Forgetting to Make a Payment

We first consider a market in which consumers may forget to make future payments. Our focus is on consumers who do not anticipate that they may forget to pay the premium (i.e., they are "naive"). We discuss the case of consumers who are aware that they may forget to make a payment (they are "sophisticated") in Subsection 4.1.4.

There are $N \geq 2$ insurance firms indexed by $j = 1, ..., N$ and a continuum of households.[24] Each household consists of one head (the "consumer") who makes all the decisions and at least one heir. We do not impose exogenous restrictions on the space of contracts. Since we are interested in explaining the pattern of life insurance contracts observed in practice, it is important that policy lapses and lapse-based pricing emerge *endogenously* in equilibrium, rather than through exogenous restrictions on the contract space. The only constraints that firms in our model face when designing their contracts are due to legal

[24]We assume that firms maximize profit. In practice, some life insurance firms are "mutuals" that, in theory, operate in the best interests of their customers. In a previous version of this paper (Gottlieb and Smetters, 2012), we show that, with perfect competition, the equilibrium of our model remains unchanged even if some firms are fully paternalistic. With market power, paternalistic firms may be able to sell different contracts than the ones offered by for-profits.

or economic realism: consumers are allowed drop their policies ("one-sided commitment")[25] and can forget, maybe strategically, to make a payment (asymmetric information).[26]

4.1.1 Timing

There are three periods. In period 1, consumers purchase an insurance policy. In period 2, each consumer forgets to make the required payment with probability $l \in (0,1)$. Then, each consumer chooses whether to replace the old insurance policy by a new one. In period 3, each consumer dies with probability $\alpha \in (0,1)$. To simplify notation, we assume that there is no discounting and normalize the net interest rate to zero.

An insurance contract is a vector of (possibly negative) state-contingent payments

$$\mathbf{T}_j \equiv \left(t_{1,j}, t_{2,j}^{R}, t_{3,j}^{R,A}, t_{3,j}^{R,D}, t_{3,j}^{F,A}, t_{3,j}^{F,D}\right) \in \mathbb{R}^6,$$

where $t_{1,j}$ is the first-period payment; $t_{2,j}^{R}$ is the second-period payment for consumers who remember (R) to pay (those who forget pay zero); and $t_{3,j}^{R,A}$, $t_{3,j}^{R,D}$, $t_{3,j}^{F,A}$, and $t_{3,j}^{F,D}$ are the last-period payments for consumers who are alive (A) or dead (D) conditional on whether they remembered (R) or forgot (F) to make a payment in period 2. These state-contingent payments can be interpreted as follows. The insurance contract specifies an immediate premium of $t_{1,j}$ and a future premium of $t_{2,j}^{R}$. Consumers who remember to pay the premium in period 2 receive $-t_{3,j}^{R,A}$ if they survive and $-t_{3,j}^{R,D}$ if they die in period 3. Those who forget to pay in period 2 receive $-t_{3,j}^{F,A}$ and $-t_{3,j}^{F,D}$ in period 3 instead. The difference between payments in period 3 corresponds to a penalty for forgetting to pay the premium in period 2. A firm's expected profits from an insurance policy is the expected state-contingent payments it gets from the consumer.

The timing of the game is as follows:

t=1: Each firm $j = 1,...,N$ offers a policy T_j. Consumers start with an initial wealth $I_1 > 0$ and decide which policy (if any) to accept. After they consume, firms offer new policies and consumers decide whether to switch to a new policy.

t=2: Consumers earn an income $I_2 > 0$ and, if they purchased a policy in period 1, forget to make the required second payment with probability $l \in (0,1)$. Forgetting to make an insurance payment means that consumers pay zero in that case ($t_{2,j}^{F} = 0$). Since we are assuming for simplicity that other assets are

[25] Whereas creditors owed repayment typically have legal authority to force two-sided commitment subject to bankruptcy law, life insurers are generally unable to do so under standard legal doctrines pertaining to individual consumers (Keeton, 1970). Lack of commitment on the policyholder side is precisely why lapses can occur, both in the models considered here and in the existing literature on reclassification risk.

[26] One could expand the model, allowing firms send messages reminding consumers to pay. This possibility would not affect the equilibrium of our model, as firms would never choose to use such reminders. Since naive consumers do not think that they will forget to pay, they would not accept worse contract terms in exchange for being reminded to make a payment. If instead consumers were sophisticated, they would prefer contracts with reminders, which insure against the risk of forgetting to pay.

illiquid, this means that they consume their entire period-2 income $c^F_{2,j} = I_2$.[27] Firms observe whether the consumer made the payment but cannot tell whether a failure to pay was strategic or because the consumer forgot.[28] After consumption happens, firms offer new policies and consumers decide whether to switch to a new policy.

t=3: Each consumer dies with probability $\alpha \in (0,1)$. Consumers who survive have income $I^A_3 > 0$, and those who die have income $I^D_3 \geq 0$, so that their expected income equals $I_3 \equiv \alpha I^D_3 + (1-\alpha) I^A_3$.

4.1.2 Consumer Utility

Let $u_A(c)$ denote the utility of household consumption when the consumer is alive, and let $u_D(c)$ denote the utility of bequests (the "joy of giving" resources to survivors). As usual, $u_A(c)$ and $u_D(c)$ are strictly increasing, strictly concave, twice differentiable, and satisfy the Inada conditions: $\lim_{c \searrow 0} u'_A(c) = +\infty$ and $\lim_{c \searrow 0} u'_D(c) = +\infty$.

Since other assets are illiquid, there is a one-to-one mapping between state-contingent payments and state-contingent consumption $\mathbf{C}_j \equiv \left(c_{1,j}, c^R_{2,j}, c^{R,A}_{3,j}, c^{R,D}_{3,j}, c^{F,A}_{3,j}, c^{F,D}_{3,j} \right)$, and there is no loss of generality in assuming that a contract specifies a vector of state-contingent *consumption* rather than state-contingent *payments*.[29] Note that a contract could specify that consumers pay a fee and remain with their old policy after missing a payment. However, as we will see, the equilibrium policy will endogenously lapse after the consumer forgets to pay. Therefore, the model *endogenously* explains why policies do not allow consumers to keep their policies after missing a payment. Because no payments are made after a consumer lapses, the first-period premium, $t_{1,j}$, corresponds to the lapse fee or, analogously, the cross-subsidy from lapsers to non-lapsers.

4.1.3 Equilibrium

An *equilibrium* of the game is a vector of policies offered by each firm in periods 1 and 2, a consumer acceptance decision in periods 1 and 2, and a consumer payment decision conditional on the policy and on remembering to pay in period 2 with the following properties:

1. Each firm's policy maximizes the firm's expected profits in each period;

[27] The assumption of full illiquidity greatly simplifies the exposition and is consistent with previous work. Daily, Lizzeri and Hendel (2008) and Fang and Kung (2010) also assume that no credit markets exist in order to generate lapses. Our results remain valid if some (but not all) assets could be reallocated.

[28] The assumption that individuals are not subject to mortality risk in periods 1 and 2 is only made to simplify notation. Our results would remain unchanged if individuals could die in all periods. Similarly, the assumption that individuals can strategically forget to pay is made for realism only. Since the equilibrium policies with naive consumers will punish those who forget to pay as harshly as possible, they cannot benefit from strategically missing a payment (this incentive constraint does not bind).

[29] Specifically, $t_{1,j} = I_1 - c_{1,j}$ is the premium paid in period 1, $t^R_{2,j} = I_2 - c^R_{2,j}$ is the premium in period 2 if the consumer remembers to make a payment. Similarly, $-t^s_{D,J} = c^{s,D}_{3,j} - I^D_3$ is the death payment and $-t^s_{A,J} = c^{s,A}_{3,j} - I^A_3$ is the payment in case the consumer outlives the policy, where $c^{s,D}_{3,j}$ corresponds to the consumer's bequest. Note that since the period-2 payment of someone who forgets is zero, $t^R_{2,j} = I_2$.

2. Each consumer chooses a policy that maximizes his/her perceived utility, randomizing with strictly positive probabilities when multiple policies give the same perceived utility;

3. If a consumer remembers to pay, he/she chooses whether or not to pay to maximize his/her period-2 continuation utility.

Competition between firms forces at least two of them to offer policies that maximize the consumer's perceived utility subject to three types of constraints.

First, because firms do not know which consumers remembered to make the payment in period 2, policies have to induce consumers to pay if they remember:

$$u_A(c_{2,j}^R) + \alpha u_D(c_{3,j}^{R,D}) + (1-\alpha)\,u_A(c_{3,j}^{R,A}) \geq u_A(I_2) + \alpha u_D(c_{3,j}^{F,D}) + (1-\alpha)\,u_A(c_{3,j}^{F,A}). \qquad \text{(IC)}$$

This incentive compatibility constraint states that consumers must get a higher utility from paying than from forgetting to pay, which entails consuming their entire income in period 2 and getting the same consumption as those who forgot to pay in period 3.

Second, because of one-sided commitment, consumers cannot have a lower utility than what they would get if they dropped their policy and recontracted with a new firm. The resulting ("renegotiation proofness") constraints will be described in detail below. Third, by a standard Bertrand argument, any contract accepted with positive probability must give zero expected profits to the company offering it:

$$c_{1,j} + (1-l)\left[c_{2,j}^R + \alpha c_{3,j}^{R,D} + (1-\alpha)\,c_{3,j}^{R,A}\right] + l\left[I_2 + \alpha c_{3,j}^{F,D} + (1-\alpha)\,c_{3,j}^{F,A}\right] = I_1 + I_2 + (1-\alpha)\,I_3. \qquad \text{(Zero profits)}$$

The renegotiation proofness constraints are specified recursively. We start with period 2, after the consumer has either remembered (R) or forgotten (F) to pay. Let

$$U(w) \equiv \max_{c_D, c_A} \alpha u_D(c_D) + (1-\alpha)\,u_A(c_A) \text{ s.t. } \alpha c_D + (1-\alpha)\,c_A = w \qquad (1)$$

denote the highest expected utility in period 3 given expected continuation income w. Renegotiation proofness in period 2 requires that the consumer does not obtain a lower expected utility than his outside option:

$$\alpha u_D(c_{3,j}^{R,D}) + (1-\alpha)\,u_A(c_{3,j}^{R,A}) \geq U\,(I_3) \qquad (RP_R)$$

and

$$\alpha u_D(c_{3,j}^{F,D}) + (1-\alpha)\,u_A(c_{3,j}^{F,A}) \geq U\,(I_3)\,. \qquad (RP_F)$$

We now move to period 1. Since the consumer believes that he will remember to make the payment in period 2, the renegotiation proofness constraint in period 1 is

$$u_A(c_{2,j}^R) + \alpha u_D(c_{3,j}^{R,D}) + (1-\alpha)\,u_A(c_{3,j}^{R,A}) \geq V, \qquad (RP_1)$$

where V is his outside option at the end of period 1, that is,

$$V \equiv \max_{\tilde{c}_2^R, \tilde{c}_D^{3,R}, \tilde{c}_A^{3,R}} u_A(\tilde{c}_2^R) + \alpha u_D(\tilde{c}_3^{R,D}) + (1-\alpha)\, u_A(\tilde{c}_3^{R,A})$$

subject to

$$(1-l)\left[\tilde{c}_2^R + \alpha \tilde{c}_3^{R,D} + (1-\alpha)\, \tilde{c}_3^{R,A}\right] + l\left[I_2 + \alpha \tilde{c}_3^{F,A} + (1-\alpha)\, \tilde{c}_3^{F,A}\right] = I_2 + I_3 \tag{Zero profits$_R$}$$

$$\alpha u_D(\tilde{c}_3^{R,D}) + (1-\alpha)\, u_A(\tilde{c}_3^{R,A}) \geq U(I_3) \tag{RP_R}$$

$$\alpha u_D(\tilde{c}_3^{F,D}) + (1-\alpha)\, u_A(\tilde{c}_3^{F,A}) \geq U(I_3) \tag{RP_F}$$

As usual, equilibrium pins down the consumption along the equilibrium path but not the consumer's strategy. Whenever a renegotiation proofness constraint binds in some state, there are both equilibria in which the consumer lapses in that state and buys a new policy and equilibria in which the policy leaves the consumer indifferent between lapsing or not (who may or may not lapse). For concreteness, we say that the policy lapses in a state if the renegotiation proofness constraint binds in that state.

To summarize, any policy accepted by the consumer must solve the following *equilibrium program*:

$$\max_{\left(c_{1,j}, c_{2,j}^R, c_{3,j}^{R,D}, c_{3,j}^{R,A}, c_{3,j}^{F,D}, c_{3,j}^{F,A}\right)} u_A(c_{1,j}) + u_A(c_{2,j}^R) + \alpha u_D(c_{3,j}^{R,D}) + (1-\alpha)\, u_A(c_{3,j}^{R,A}) \tag{2}$$

subject to (IC), (RP_1), (RP_R), (RP_F), and *(Zero Profits)*.

In Online Appendix A, we formally show the equivalence between the equilibrium consumption and the solution of program (2). When there are more than two firms, there exist equilibria in which some firms offer contracts that are never accepted. However, because the equilibrium program has a unique solution, the set of contracts accepted with positive probability in any equilibrium is unique. We therefore omit the index j from contracts that are accepted with positive probability and refer to the solution of (2) as *the* equilibrium consumption.

Before presenting the main result, it is helpful to consider the following auxiliary program to help guide intuition:

$$\max_{c_1, c_2^R, c_3^R} u_A(c_1) + u_A(c_2^R) + U(c_3^R)$$

subject to

$$\frac{c_1}{1-l} + c_2^R + c_3^R = \frac{I_1}{1-l} + I_2 + I_3 \tag{BC}$$

$$c_1 \leq I_1 \tag{LC_1}$$

$$c_3^R \geq I_3 \tag{LC_2}$$

This auxiliary program corresponds to a standard consumption-savings problem, where (BC) is the budget constraint, $1-l$ is the interest rate between periods 1 and 2, (LC_1) and (LC_2) are liquidity constraints that prevent the consumer from borrowing in periods 1 and 2, and $U(\cdot)$ is the expected last-period utility as defined in equation (1).

The following lemma relates the auxiliary program to the equilibrium program:

Lemma 1. *Let $(c_1, c_2^R, c_3^{R,A}, c_3^{R,D}, c_3^{F,A}, c_3^{F,D})$ be the equilibrium consumption and let $c_3^R \equiv \alpha c_3^{R,A} + (1-\alpha) c_3^{R,D}$ denote the expected last-period consumption conditionally on remembering to pay. Then, (c_1, c_2^R, c_3^R) solves the auxiliary program. Moreover, $u_A'(c_3^{F,A}) = u_D'(c_3^{F,D})$ and $c_3^{F,A} + c_3^{F,D} = I_3$.*

Because a consumer would walk away from any contract that promised less future consumption than the consumer's future income, renegotiation proofness implies that the liquidity constraints (LC_1) and (LC_2) must be satisfied. The previous lemma shows that individuals who remember to make the payment get the smoothest consumption path consistent with the the ability to drop a policy.

If an individual forgets to make the payment, the policy endogenously lapses and the policyholder loses the front load $I_1 - c_1 \geq 0$. The probability of remembering to pay, $(1-l)$, plays the role of an interest rate because the policyholder loses the entire front load if the policy lapses, which happens with probability l. Therefore, firms are willing to trade a dollar of first-period consumption for $\frac{1}{1-l}$ dollars of future consumption. If the front load is strictly positive, the equilibrium policy is lapse-based, meaning that the insurance company makes money on those who lapse and loses money on those who do not lapse. If the front load is zero, the equilibrium policy can be interpreted as a sequence of annual policies, since the individual pays the actuarially-fair premium in each period. Since the front load is positive whenever the first-period liquidity constraint does not bind, the equilibrium policy is lapse-based if the individual has enough initial income or if the probability of remembering to pay (the "interest rate"), $1-l$, is high enough:

Proposition 1. *The cross-subsidy between consumers who lapse and those who do not lapse is always weakly positive ($t_1 \equiv I_1 - c_1 \geq 0$). It is strictly positive if $(1-l)\, u'(I_1) < u'(I_2)$.*

Therefore, when consumers have enough initial income and the probability of forgetting to pay is high enough, the equilibrium features a long-term policy that endogenously lapses whenever the individual forgets to make a payment. The policy makes a positive profit from those who forget to pay and a negative profit from those who make the payment. Otherwise, the equilibrium is analogous to purchasing a sequence of annual policies, which are actuarially fair in each period.[30]

4.1.4 Sophistication

What would happen in this model if consumers were aware that they may forget to pay their premiums? In Online Appendix A, we consider the same model as before, except that we assume that consumers have

[30]In Online Appendix A, we present comparative statics results involving the lapse fee. In particular, we show that the lapse fee is increasing (decreasing) in the probability of forgetting to pay if the elasticity of intertemporal substitution of savings is greater (smaller) than 1. Empirically, in Online Appendix D, we find suggestive evidence that lapse fees tend to increase in the lapse probability.

correct expectations about the probability that they forget to make the payment in period 2. We show that there are either no lapses in equilibrium, or that firms make money on consumers who remember to pay their premiums and lose money on those who forget. Moreover, if firms could commit to reminding consumers to pay their bill, they would choose to do so.

The intuition for these counterfactual results is that sophisticated individuals value consumption after they forget to pay the premium just as much as in states after they pay the premium. If we ignored the incentive constraint, individuals would want to strategically forget to pay the premium in period 2, allowing them to consume their entire income in that period while getting the same consumption in period 3. The equilibrium policy will then pay the smoothest consumption profile that is consistent with incentive compatibility, which insures the consumer against the risk of a lower consumption after to forgetting to pay. This means that the equilibrium policy transfers resources in the opposite direction than when consumers are naive (and what is observed in practice). That is, the cross subsidy goes from consumers who remember to pay to those who forget to pay the premium. Reminding consumers to pay their bill entirely removes this risk, leading to a smooth consumption path.

To summarize, the assumption that consumers may forget to pay their premiums can only account for the structure of life insurance policies observed in practice if they are not fully aware of this possibility.

4.2 Income Shocks

We now consider a model in which consumers underestimate some other "background" (or "income") shocks when purchasing life insurance. To highlight the underlying mechanism in a transparent way, we focus on the simple case where consumers entirely disregard these shocks, and we contrast it with the other polar (rational expectations) case where background shocks are perfectly taken into account. To build intuition, we initially assume a single possible loss, which we relax later by allowing for a continuum of (possibly heterogeneous) losses. As before, there are $N \geq 2$ insurance firms indexed by $j = 1, ..., N$ and a continuum of consumers.

4.2.1 Timing

There are three periods: 0, 1, and 2. Period 0 is a pure contracting stage (no consumption) in which each firm offers an insurance policy and consumers decide which one, if any, to purchase. Consumption occurs in periods 1 and 2. In period 1, consumers suffer a loss with probability l that is not observable to firms. In period 2, each consumer dies with probability $\alpha \in (0,1)$.[31] To examine the role that lapses play in providing liquidity, we assume, as before, that the consumer's other assets are fully illiquid, so they cannot be rebalanced after a loss.

[31] None of our results would change if we allowed for consumption in period 0. Similarly, the assumption that individuals are not subject to mortality risk in period 1 is only made to simplify notation. Ruling out income shocks in period 2 also helps the analysis, but can be substantially generalized.

An insurance contract is a vector of (possibly negative) state-contingent payments

$$\mathbf{T}_j \equiv \left(t_{1,j}^S, t_{1,j}^{NS}, t_{A,j}^S, t_{D,j}^S, t_{A,j}^{NS}, t_{D,j}^{NS}\right) \in \mathbb{R}^6,$$

where $t_{1,j}^S$ and $t_{1,j}^{NS}$ are payments in period 1 when the consumer does and does not suffer the income shock, respectively. The variables $t_{A,j}^S$, $t_{D,j}^S$, $t_{A,j}^{NS}$, and $t_{D,j}^{NS}$ denote the payments in period 2 when the consumer is alive (A) or dead (D) conditional on whether he suffered an income shock (S) or not (NS) in period 1. A natural interpretation of these state-contingent payments is as follows. Consumers pay a premium $t_{1,j}^{NS}$ for insurance when they buy a policy in period 0. In period 1, they choose whether or not to borrow from their policy. If they do not borrow, the insurance company pays $-t_{A,j}^{NS}$ if they survive and $-t_{D,j}^{NS}$ if they die at $t = 2$. If they borrow from their policy, the insurance company pays $-t_{A,j}^S$ if they survive and $-t_{D,j}^S$ if they die.

To be more concrete, the timing of the game is as follows:

t=0: Each firm $j = 1, ..., N$ offers an insurance contract T_j. Consumers start with initial wealth $W > 0$ and decide which contract, if any, to accept, randomizing with strictly positive probabilities between contracts when indifferent.

t=1: Firms offer new contracts. Consumers make income $I_1 > 0$, lose L dollars with probability l and decide whether to switch to a new contract. If they do not switch, they choose whether to report an income shock to the insurance company (i.e., they "borrow from the policy"). Consumers pay $t_{1,j}^{NS}$ if they do not report a shock and $t_{1,j}^S$ if they do.

t=2: Each consumer dies with probability $\alpha \in (0, 1)$. Households of consumers who survive have income $I_2^A > 0$, and households of those who die have income $I_2^D \geq 0$. Therefore, their expected income is $I_2 \equiv \alpha I_2^D + (1-\alpha) I_2^A$. Households of consumers who reported a loss $t = 1$ are paid either $-t_{A,j}^S$ or $-t_{D,j}^S$, depending on whether the head of the household is alive (A) or dead (D). Those who did not report a loss at $t = 1$ are paid either $-t_{A,j}^{NS}$ and $-t_{D,j}^{NS}$.

As before, we do not impose any exogenous restrictions on the space of contracts. Our goal is to understand when the equilibrium policies generate the patterns observed in practice, such as being lapse-based and allowing for policy loans. It is therefore important that lapses and policy loans emerge endogenously in equilibrium, rather than through exogenous restrictions on the contract space. We follow the mechanism design terminology and say that a consumer "reports a loss to the insurance firm" with the understanding that such a direct mechanism is equivalent to more realistic indirect mechanisms, in which consumers borrow from their policies after an unobserved need for money. As in the model of Subsection 4.1, we say that the policy lapses in a state if the renegotiation proofness constraint binds in that state.

The utility of household consumption is $u_A(c)$ when the consumer is alive and $u_D(c)$ when the consumer is dead (i.e., utility from bequests). It is convenient to work with state-contingent consumption

$\mathbf{C}_j \equiv \left(c^S_{1,j}, c^{NS}_{1,j}, c^S_{A,j}, c^S_{D,j}, c^{NS}_{A,j}, c^S_{D,j}\right)$ instead of state-contingent payments $\mathbf{T}_j$.[32] Our key assumption is that consumers do not take background risk into account when buying life insurance in period 0, so they attribute probability zero to suffering an income shock at the contracting stage. Consumers, therefore, only consider states in which background shocks do not occur when they evaluate contracts in period 0:

$$u_A(c^{NS}_{1,j}) + \alpha u_D\left(c^{NS}_{D,j}\right) + (1-\alpha)\, u_A\left(c^{NS}_{A,j}\right).$$

We refer to this expression as the consumer's "perceived utility."[33]

4.2.2 Equilibrium

As in the model of Subsection 4.1, an *equilibrium* of the game is a vector of policies offered by each firm in periods 0 and 1, a consumer acceptance decision in periods 0 and 1, and a report by the consumer conditional on the policy and on the liquidity shock with the following properties:

1. Each firm's policy maximizes the firm's expected profits in each period;
2. Each consumer chooses a policy that maximizes his/her perceived utility, randomizing with strictly positive probabilities when multiple policies give the same perceived utility;
3. Each consumer chooses whether or not to report an income shock to maximize his/her period-1 continuation utility.

As before, price competition forces firms to offer policies that maximize the consumer's perceived utility subject to zero profits, incentive compatibility, and renegotiation proofness constraints.

The zero profits condition is:

$$l\left[c^S_{1,j} + \alpha c^S_{D,j} + (1-\alpha) c^S_{A,j}\right] + (1-l)\left[c^{NS}_{1,j} + \alpha c^{NS}_{D,j} + (1-\alpha) c^{NS}_{A,j}\right] = W + I_1 + I_2 - lL. \quad \text{(Zero Profits)}$$

The incentive compatibility constraints require consumers to report the income shocks truthfully in period 1. Those who experience the shock report it truthfully if the following constraint holds:

$$u_A(c^S_{1,j}) + \alpha u_D(c^S_{D,j}) + (1-\alpha) u_A(c^S_{A,j}) \geq u_A(c^{NS}_{1,j} - L) + \alpha u_D(c^{NS}_{D,j}) + (1-\alpha) u_A(c^{NS}_{A,j}). \qquad (IC_S)$$

In words: borrowing must give consumers a higher utility than absorbing the loss. Similarly, those who

[32]The vector of state-contingent consumption is determined by $c^{NS}_{1,j} \equiv W - t^{NS}_{1,j}$, $c^{NS}_{A,j} \equiv I - t^{NS}_{A,j}$, $c^{NS}_{D,j} \equiv -t^{NS}_{A,j}$, $c^S_{1,j} \equiv W - L - t^S_{1,j}$, $c^S_{A,j} \equiv I - t^S_{A,j}$, and $c^S_{D,j} \equiv -t^S_{D,j}$.

[33]Our results would not change if we had both positive and negative shocks as long as we assume that companies cannot prevent consumers from buying additional coverage. Thus, our assumption is also consistent with consumers who decide how much life insurance to buy according to their expected future incomes, rather than taking the whole distribution into account (c.f., Eyster and Weizsäcker, 2011).

do not experience the income shock do not report one if the following constraint holds:

$$u_A(c_{1,j}^{NS})+\alpha u_D(c_{D,j}^{NS})+(1-\alpha)u_A(c_{A,j}^{NS})\geq u_A(c_{1,j}^{S}+L)+\alpha u_D(c_{D,j}^{S})+(1-\alpha)u_A(c_{A,j}^{S}). \qquad (IC_{NS})$$

Let

$$\mathscr{V}(w)\equiv \max_{c_1,c_D,c_A}\left\{\begin{array}{l} u_A(c_1)+\alpha u_D(c_D)+(1-\alpha)u_A(c_A) \\ \text{subject to } c_1+\alpha c_D+(1-\alpha)c_A=w \end{array}\right\} \qquad (3)$$

denote the highest period-1 continuation utility that can be obtained with an expected continuation income of w. The renegotiation proofness constraints state that consumers cannot obtain a utility below what they could obtain by lapsing and buying another policy:

$$u_A(c_{1,j}^{NS})+\alpha u_D(c_{D,j}^{NS})+(1-\alpha)u_A(c_{A,j}^{NS})\geq \mathscr{V}(I_1+I_2) \qquad (RP_{NS})$$

$$u_A(c_{1,j}^{S})+\alpha u_D(c_{D,j}^{S})+(1-\alpha)u_A(c_{A,j}^{S})\geq \mathscr{V}(I_1+I_2-L) \qquad (RP_{S})$$

Any equilibrium policy accepted by the consumer must maximize the consumer's perceived utility subject to the constraints described above:

$$\max_{\left(c_{1,j}^{NS},c_{D,j}^{NS},c_{A,j}^{NS},c_{1,j}^{S},c_{D,j}^{S},c_{A,j}^{S}\right)} u_A(c_{1,j}^{NS})+\alpha u_D\left(c_{D,j}^{NS}\right)+(1-\alpha)u_A\left(c_{A,j}^{NS}\right)$$

subject to (*Zero Profits*), (RP_{NS}), (RP_S), and (IC_S) and (IC_{NS}). We refer to this program as the *equilibrium program*.[34] As in Subsection 4.1, the equilibrium program has a unique solution, so we can drop the subscript j from the unique contract accepted in any equilibrium.

Let (c_A^*,c_D^*) denote the consumption vector that provides full insurance and costs the household's expected lifetime wealth:[35]

$$u_A'(c_A^*)=u_D'(c_D^*),$$
$$(2-\alpha)c_A^*+\alpha c_D^*=\frac{W}{1-l}+I_1+I_2.$$

As we will show below, a policy lapses after an income shock if the following inequality holds:

$$u_A(c_A^*)-u_A(c_A^*-L)\geq \mathscr{V}\left(\frac{W}{1-l}+I_1+I_2\right)-\mathscr{V}\left(I_1+I_2-L\right), \qquad (4)$$

where $\mathscr{V}$ was defined in (3). If it does not hold, the individual borrows from the policy after an income shock. Since consumption while alive is a normal good ($\frac{dc_A^*}{dW}>0$) and utility is concave, the expression on the left is decreasing in wealth. The expression on the right is increasing in wealth. Therefore, wealthier individuals are less likely to lapse after an income shock.

[34] See the online appendix for a formal discussion of the equivalence between the equilibrium consumption and the solution of the equilibrium program.

[35] By the strict concavity of the utility function and the Inada conditions, (c_A^*,c_D^*) exists and is unique. Note also that (c_A^*,c_D^*) only depends on exogenous variables.

To understand condition (4), suppose the consumer has a state-independent utility $u_A = u_D = u$, so that the condition becomes:

$$u\left(\frac{I_1+I_2-L}{2}\right) \geq \frac{u\left(\frac{\frac{W}{1-l}+I_1+I_2}{2}\right)+u\left(\frac{\frac{W}{1-l}+I_1+I_2}{2}-L\right)}{2}.$$

This condition holds if $\frac{W}{1-l}$ is small enough and, for $\frac{W}{1-l} < L$, when u is concave enough. It always fails if $\frac{W}{1-l}$ is large enough. This is because lapsing allows the individual to smooth consumption at the cost losing the front load (which, as we will show, equals $\frac{W}{1-l}$). When the consumer does not have enough initial income W or has a sufficiently concave utility function, his front-load payments are not enough to prevent him from lapsing after an income shock.

Given an equilibrium consumption, let

$$\pi^{NS} \equiv W + I_1 + I_2 - \left[c_1^{NS} + \alpha c_D^{NS} + (1-\alpha)c_A^{NS}\right]$$

denote the insurer's expected profits from a consumer who does not experience a shock, and let

$$\pi^{S} \equiv W + I_1 - L + I_2 - \left[c_1^{S} + \alpha c_D^{S} + (1-\alpha)c_A^{S}\right]$$

denote the insurer's expected profits from a consumer who experiences a shock.

The following proposition presents the main properties of the equilibrium contract:

Proposition 2. *In the essentially unique equilibrium, policyholders never lapse if they do not have an income shock – i.e.,* (RP_{NS}) *does not bind. Moreover, if equation (4) holds, then*

1. *Policyholders lapse after an income shock:* (RP_S) *binds;*
2. *Perfect smoothing conditional on each shock:* $u_A'(c_1^s) = u_D'(c_D^s) = u_A'(c_A^s)$ *for* $s = S, NS$*;*
3. *Lapse-supported pricing:* $\pi^S = W > 0$ *and* $\pi^{NS} = -\frac{l}{1-l}W < 0$.

If equation (4) does not hold, then:

4. *Policyholders do not lapse:* (RP_S) *does not bind;*
5. *Perfect smoothing after a loss:* $u_A'(c_1^S) = u_D'(c_D^S) = u_A'(c_A^S)$*;*
6. *Insufficient saving if no loss:* $u_D'(c_D^{NS}) = u_A'(c_A^{NS}) < u_A'(c_1^{NS})$*; and*
7. *Policy loans after an income shock:* $c_1^{NS} - c_1^S < L$, $c_D^{NS} > c_D^S$*, and* $c_A^{NS} > c_A^S$*; and*
8. *Policy loans at above-market rates:* $\pi^S > 0 > \pi^{NS}$.

We first interpret Proposition 2 and then present a sketch of its proof. Recall that consumers lapse after an income shock if (4) holds and borrow from their policies if (4) fails.

Consider the case in which consumers lapse after an income shock first, so that (RP_S) binds. Property 2 states that the equilibrium features full insurance with respect to mortality and perfect smoothing between periods (but not with respect to the income shock). Property 3 states that the insurer gets a positive profit of W on consumers who get an income shock (and therefore lapse), and a negative profit of $-\frac{l}{1-l}W$ on those who do not get an income shock (and therefore do not lapse). That is, not only does the policy fail to insure against the income loss, it exacerbates the loss since consumers also lose the front load W when they lapse. This amount is used to cross subsidize those who do not have an income shock.

Now, consider the case in which (RP_S) does not bind, so that policyholders do not lapse. Property 7 states that individuals who suffer a loss borrow from the policy. That is, consumption in period 1 falls by less than the income lost in exchange for a reduction in future consumption. Conditionally on the loss, they are fully insured (property 5). The equality in property 6 states that consumers are fully insured against mortality risk conditional on not suffering an income shock, whereas the inequality states that policy loans charge an interest rate higher than the market rate (normalized to zero). As a result, the insurance policy discourages borrowing relative to efficient consumption smoothing, which would equate the marginal utility of consumption across periods. Property 8 states that firms obtain positive profits on consumers that borrow from their policies and negative profits from those that do not. Since, in expectation, insurance companies break even, all profits from policy loans are competed away by charging a lower price from policyholders who do not experience an income shock and, therefore, do not borrow from their policies. Therefore, the policy allows consumers to borrow at above-market interest rates.

4.2.3 Sketch of Proof to Proposition 2

We now provide a sketch of the proof to Proposition 2 (see Online Appendix A for a formal proof). Since the equilibrium policy maximizes the consumer's perceived utility, which assigns probability zero to an income shock, it will transfer as much resources from states after a shock occurs to states in which the shock does not occur. Therefore, the relevant incentive problem consists of inducing consumers to report a shock truthfully, not preventing those who did not suffer a shock from misreporting one. We will, therefore, ignore constraint (IC_{NS}) for now and verify that it holds later. That is, we consider the following relaxed program:

$$\max_{\left(c_1^{NS},c_D^{NS},c_A^{NS},c_1^S,c_D^S,c_A^S\right)} u_A\left(c_1^{NS}\right)+\alpha u_D(c_D^{NS})+(1-\alpha)u_A(c_A^{NS})$$

subject to

$$u_A(c_1^S)+\alpha u_D(c_D^S)+(1-\alpha)u_A(c_A^S)\geq u_A(c_1^{NS}-L)+\alpha u_D(c_D^{NS})+(1-\alpha)u_A(c_A^{NS}) \qquad (IC_S)$$

$$u_A(c_1^S) + \alpha u_D(c_D^S) + (1-\alpha) u_A(c_A^S) \geq \mathscr{V}(I_1 + I_2 - L) \tag{RP_S}$$

$$u_A(c_1^{NS}) + \alpha u_D(c_D^{NS}) + (1-\alpha) u_A(c_A^{NS}) \geq \mathscr{V}(I_1 + I_2) \tag{RP_{NS}}$$

$$l\left[c_1^S + \alpha c_D^S + (1-\alpha) c_A^S\right] + (1-l)\left[c_1^{NS} + \alpha c_D^{NS} + (1-\alpha) c_A^{NS}\right] \leq W + I_1 + I_2 - lL. \tag{Zero Profits}$$

Note that, generically, either (IC_S) or (RP_S) do not bind, depending on whether absorbing a loss while keeping the contract or signing a new contract is more attractive:

$$u_A(c_1^{NS} - L) + \alpha u_D(c_D^{NS}) + (1-\alpha) u_A(c_A^{NS}) \left\{ \begin{array}{c} > \\ < \end{array} \right\} \mathscr{V}(I_1 + I_2 - L).$$

We start with the case where (IC_S) does not bind, so that there is full insurance conditional on each realization of the shock:

$$u_A'(c_1^s) = u_D'(c_D^s) = u_A'(c_A^s)$$

for $s = S, NS$. If this were not the case, it would be possible to save resources by smoothing consumption between these states, allowing us to raise consumption in the first period, thereby increasing the consumer's perceived utility. Since there is also full insurance after the consumer lapses, we can rewrite the renegotiation proofness constraints as:

$$c_1^S + \alpha c_D^S + (1-\alpha) c_A^S \geq I_1 + I_2 - L. \tag{RP_S}$$

and

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha) c_A^{NS} \geq I_1 + I_2 \tag{RP_{NS}}$$

We claim that (RP_S) must bind, so consumers lapse after experiencing an income loss. If this were not the case, we would be able to increase the agent's perceived utility by shifting consumption from a state after the shock to a state in which the shock does not occur. Substituting (RP_S) into the zero profits constraint, we obtain:

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha) c_A^{NS} = \frac{W}{1-l} + I_1 + I_2. \tag{5}$$

Substituting (5) into (RP_{NS}) verifies that consumers who do not experience a loss do not lapse:

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha) c_A^{NS} = \frac{W}{1-l} + I_1 + I_2 > I_1 + I_2.$$

We now verify that the omitted constraint (IC_{NS}) is satisfied. Notice that because there is full insurance conditional on the realization of the shock, (IC_{NS}) becomes:

$$\mathscr{V}\left(c_1^{NS} + \alpha c_D^{NS} + (1-\alpha) c_A^{NS}\right) \geq u_A(c_1^S + L) - u_A(c_1^S) + \mathscr{V}\left(c_1^S + \alpha c_D^S + (1-\alpha) c_A^S\right).$$

Substituting (5) and the binding renegotiation proofness constraint (RP_S), we obtain:

$$\mathscr{V}\left(\frac{W}{1-l}+I_1+I_2\right) \geq u_A(c_1^S+L)-u_A(c_1^S)+\mathscr{V}(I_1+I_2-L). \tag{6}$$

Since (RP_S) binds and c_1^S, c_A^S, and c_D^S maximize the agent's expected utility conditional on the loss, we must have

$$\mathscr{V}(I_1+I_2-L)=\alpha u_D(c_D^S)+(2-\alpha)u_A(c_1^S),$$

where we used the fact that $c_1^S=c_A^S$. Use this expression to rewrite the terms on the RHS of (6) as

$$u_A(c_1^S+L)+\mathscr{V}(I_1+I_2-L)-u_A(c_1^S)=u_A(c_1^S+L)+\alpha u_D(c_D^S)+(1-\alpha)u_A(c_1^S). \tag{7}$$

Since $c_1^S+L+\alpha c_D^S+(1-\alpha)c_1^S=I_1+I_2$, it follows that

$$u_A(c_1^S+L)+\alpha u_D(c_D^S)+(1-\alpha)u_A(c_1^S)\leq \mathscr{V}(I_1+I_2)<\mathscr{V}\left(\frac{W}{1-l}+I_1+I_2\right),$$

where the last inequality uses the fact that $\mathscr{V}$ is strictly increasing. Using (7) again verifies that the (6) holds.

We now move to the case where (RP_S) does not bind and, for the moment, ignore (RP_{NS}), which can be verified afterwards. The relaxed program becomes:

$$\max_{\left(c_1^{NS},c_D^{NS},c_A^{NS},c_1^S,c_D^S,c_A^S\right)} u_A\left(c_1^{NS}\right)+\alpha u_D(c_D^{NS})+(1-\alpha)u_A(c_A^{NS})$$

subject to

$$u_A(c_1^S)+\alpha u_D(c_D^S)+(1-\alpha)u_A(c_A^S)\geq u_A(c_1^{NS}-L)+\alpha u_D(c_D^{NS})+(1-\alpha)u_A(c_A^{NS}) \tag{IC_S}$$

$$l\left[c_1^S+\alpha c_D^S+(1-\alpha)c_A^S\right]+(1-l)\left[c_1^{NS}+\alpha c_D^{NS}+(1-\alpha)c_A^{NS}\right]\leq W+I_1+I_2-lL. \tag{Zero Profits}$$

The results then follow from the FOCs of this program. To understand why the equilibrium policy induces excessive saving relative to full insurance (property 6), start from an allocation that provides full insurance for consumers who do not experience a shock, $u_A'(c_1^{NS})=u_A'(c_A^{NS})=u_D'(c_D^{NS})$, and consider the effect of shifting consumption from period 1 to period 2 by a small amount $\delta>0$. This perturbation does not affect the total consumption, leaving (*Zero Profits*) unchanged. Moreover, it has second-order effects on the objective. However, this perturbation relaxes (IC_S) by

$$\left[-u_A'(c_1^{NS}-L)+\alpha u_D'(c_D^{NS})+(1-\alpha)u_A'(c_A^{NS})\right]\delta=\left[u_A'(c_A^{NS})-u_A'(c_1^{NS}-L)\right]\delta,$$

which is positive since $c_A^{NS}=c_1^{NS}>c_1^{NS}-L$. Intuitively, shifting consumption away from period 1 increases the cost not to borrowing after an income shock. Consumers are fully aware of the intertemporal wedge induced by the equilibrium policy. Nevertheless, without this wedge, it is costlier to induce them to borrow after an income shock, so any firm that attempts to offer a contract that smooths inter-temporal consumption would be unable to price it competitively.

To conclude, Proposition 2 established that when all consumers are subject a single loss L, the equilibrium either involves policy loans at above market rates or the policy lapses after the loss. We now turn to the more realistic case where consumers are subject to a continuum of (possibly heterogeneous) losses and show that, not only do the insights from the single-loss model generalize, but the equilibrium contract becomes more realistic, providing policy loans up to a cap at above-market rates. Consumers with small losses borrow from their policies, whereas those with large enough losses lapse.

4.2.4 Continuum of Losses

We now consider the more realistic case in which consumers are subject to multiple losses. Losses are continuously distributed according to a CDF F with full support in the interval $[0,\bar{L}]$ where $0<\bar{L}<I_1+I_2$. To simplify notation, we assume that consumers have a state-independent utility function: $u(c)=u_A(c)=u_D(c)$. As before, we assume that consumers believe they will not lapse, meaning that their beliefs assign full mass to $L=0$. Since all consumers have the same beliefs in period 0, the model allows them to have heterogeneous loss distributions (not observed by firms). In that case, F denotes the distribution of losses in the population.

Let $c_1(L)$ denote the consumption in period 1 conditional on loss $L\in[0,\bar{L}]$. Similarly, let $c_A(L)$ and $c_D(L)$ denote the consumption in period 2 if the individual is alive and the bequests upon death, conditional on loss L. The following theorem shows the main properties of the equilibrium:

Theorem 1. *There exits $\tau\in[0,\bar{L}]$ such that the consumer does not lapse if $L<\tau$ and lapses if $L>\tau$. Moreover:*

1. $c_1(L)=c_D(L)=c_A(L)=\frac{I_1+I_2-L}{2}$ *for all $L>\tau$.*
2. $c_1(L)<c_D(L)=c_A(L)$ *for all $L<\tau$.*
3. $c_1(L)-L$ *is non-decreasing.*
4. *Profits are non-decreasing in L.*
5. *When the lapse region is non-empty ($\tau<L$), insurance companies make a profit of W on each consumer who lapses and lose an average of $\frac{W}{1-F(\tau)}$ among those who do not lapse.*

Theorem 1 shows how the properties from the model with a single loss generalize to the case of a continuum of possible losses. Consumers who suffer large losses lapse and buy a new policy, obtaining full insurance (property 1). Those with small losses borrow from their policies, so their period-1 consumption $c_1(L)$ falls by less than the loss itself L (property 3). Policy loans are provided at above-market rates, generating insufficient savings among those who do not lapse (property 2) and causing profits to be increasing in the size of the loan (property 4). Moreover, insurance is lapse-based (property 5).

The proof of Theorem 1 is interesting in itself, since renegotiation proofness introduces a pure state constraint, which makes the resulting optimal control problem non-standard. We provide a formal treatment and proof in Online Appendix A along a sketch in the Appendix. Our approach can be useful to other researchers interested in solving mechanism design problems with both incentive and renegotiation proofness constraints, even without behavioral components (as in our analysis of the rational expectations case below).

In general, the equilibrium policy does not not have a closed-form solution. However, in the special case of log utility and a uniform distribution, the equilibrium is remarkably simple:

Proposition 3. *Let $L \sim U[0,\bar{L}]$ and let $u(c) = \ln c$. There exists $\tau \in [0,\bar{L}]$ such that*

$$c_1(L) = \begin{cases} \frac{I-\tau}{2} & \text{if } L < \tau \\ \frac{I-L}{2} & \text{if } L > \tau \end{cases},$$

and

$$c_A(L) = c_D(L) = \begin{cases} \frac{I-\tau}{2} \exp\left(2\frac{\tau-L}{I-\tau}\right) & \text{if } L < \tau \\ \frac{I-L}{2} & \text{if } L > \tau \end{cases}.$$

In this case, the equilibrium policy allows consumers to borrow up to an amount τ, which fully compensates the income shock, keeping current consumption constant in exchange for a reduction in future consumption. Loans are offered at above-market interest rates (which have been normalized to zero), so that future consumption falls by more than the loss itself: $c'_A(L) = c'_D(L) < -1$ for $L < \tau$. The loan is capped at τ, and policyholders with losses greater than τ lapse, losing the front load W.

4.2.5 Rational Expectations

We conclude our analysis of the model of income shocks with the case of consumers with rational expectations. The model is the same as in Subsection 4.2.2, except that, in period 1, consumers anticipate that they will incur a loss L according to the continuous distribution with a CDF F described previously.[36] The following proposition establishes the main properties of the equilibrium contract:

Proposition 4. *When consumers have rational expectations, there exits $\tau \in [0,\bar{L}]$ such that the consumer lapses if $L < \tau$ and does not lapse if $L > \tau$. Moreover:*

1. $c_1(L) = c_D(L) = c_A(L) = \frac{I_1+I_2-L}{2}$ *for all $L < \tau$;*
2. $c_1(L) > c_D(L) = c_A(L)$ *for all $L > \tau$;*
3. $c_1(L) - L$ *is non-decreasing;*

[36]In the rational expectations case, heterogeneity in losses could lead to adverse selection. We rule out this channel by assuming that consumers are homogeneous (or, analogously, that firms know each consumer's loss distribution).

4. *When the lapse region is non-empty ($\tau > 0$), insurance companies make a profit of W for each consumer that lapses and lose an average of $\frac{W}{F(L)}$ among consumers who do not lapse.*

The proposition shows that lapses and policy loans have the opposite pattern when consumers have rational expectations. The equilibrium policy sets a *floor*---rather than a cap---on the amount that can be borrowed from the policy, so that individuals with small losses lapse, whereas those with larger losses borrow from their policies. Policy loans are *subsidized*, so that individuals have excessive consumption in period 1 relative to efficient consumption smoothing. These predictions are inconsistent with the pattern of life insurance policies observed in practice.

To understand the intuition behind Proposition 4, suppose losses are observable and consumers can commit not to drop their policies, so that the equilibrium program does not have to satisfy incentive compatibility and renegotiation proofness constraints. With rational expectations, the equilibrium contract provides full insurance against all shocks---i.e., consumers are fully compensated for each loss without having to repay in the future. Of course, these policies are not incentive compatible when losses are not observed since consumers would pretend to have incurred a large loss so as to increase their current consumption. It also violates renegotiation proofness for consumers with small enough losses, who have to cross subsidize those with large losses. Thus, when we incorporate these two constraints, equilibrium policies offer subsidized loans: consumers have to repay part of it to maintain incentive compatibility but are given below-market interest rates to insure against losses. With below-market interest rates, consumers who take up smaller loans subsidize those with large loans, causing renegotiation proofness to bind at the bottom.

To summarize, the key incentive aspect in the rational expectation model is the provision of insurance against unobservable liquidity shocks, which causes policies to offer loans against large shocks at subsidized rates and causes those with the lowest shocks to lapse. In contrast, in the naive model considered earlier, consumers believe that they will not have any losses. So, the insurer would like to charge as much as possible from everyone who gets a positive loss (since consumers think that such states would not happen). This would correspond to charging an infinitely high interest rate for everyone who experiences a loss. But an infinitely high interest rate would induce consumers to pretend that they did not experience a loss. Thus, to restore incentive compatibility, the firm charges an above market, but still finite, interest rate. This disproportionately hurts consumers with larger losses, who now cross subsidize others by paying above-market interest rates. Thus, renegotiation proofness binds at the top (i.e., those with large enough losses).[37]

[37] Assuming that some consumers are aware of possible losses and some are not does not affect our previous results. In this case, firms screen consumers by offering two policies: consumers who are unaware of their income shocks would pick the policies characterized in Subsection 4.2.4, while consumers with rational expectations would take the policies characterized in this subsection (the incentive compatibility constraints between rational and behavioral consumers would not bind).

4.3 Inefficiency and Secondary Markets

We believe that the appropriate efficiency criterion in the models considered previously takes into account the true probability of forgetting to pay a premium and of getting an income shock. Therefore, we say that an allocation is *efficient* if there is no other allocation that increases the expected utility of consumers---evaluated according to the true distributions---without decreasing the expected profit of any firm.

Because consumers are risk averse and insurance companies are risk neutral, any efficient allocation equalizes the marginal utility of consumption across all states. In the forgetfulness model (Section 4.1), firms do not observe whether a consumer remembered to pay, only if he/she paid. Therefore, allocations must incentivize consumers to pay when they remember to do so. Similarly, in the model of differential attention with income shocks (Section 4.2), firms do not observe the loss of each household, so allocations must induce consumers not to misrepresent their losses. An allocation is *constrained efficient* if there is no other incentive compatible allocation that increases the expected utility of consumers without decreasing the expected profit of any firm.

The equilibrium in the (empirically-relevant) naive version of the forgetfulness model is inefficient since consumption falls after the consumer fails to pay. Similarly, the equilibrium in the (empirically-relevant) naive version of the model with income shocks is inefficient, since consumption is decreasing in the loss. While the inefficiency in these models is not surprising given the presence of private information, the equilibria are also constrained inefficient.[38]

It is natural to ask which regulation could improve efficiency in these markets. In a previous version of this paper (Gottlieb and Smetters, 2012), we describe the effects from introducing a secondary market for life insurance policies. In a secondary market, individuals may resell their policies to risk-neutral firms, who then become the beneficiaries of such policies. In the rational model with reclassification risk, a secondary market reduces efficiency by undermining dynamic risk pooling (Daily, Hendel, and Lizzeri, 2008). With unanticipated income shocks, however, consumers are able to obtain a smoother consumption stream by renegotiating on the secondary market. In the long run, firms anticipate that lapse-based profits will be arbitraged away in the secondary market.[39] As a result, they offer policies that are neither front-loaded nor lapse-based. Nevertheless, because consumers do not anticipate background shocks, there is still imperfect consumption smoothing as consumption falls after the shock. Still, in the

[38]The sophisticated versions of these models (Subsections 4.1.4 and 4.2.5) obtain the point on the Pareto frontier associated with zero profits. As discussed in Subsection 4.1.4, the constrained efficient allocation in the forgetfulness model insures consumers against the risk of forgetting to pay. In contrast, the equilibrium with naive consumer maximally exacerbates the loss after forgetting to pay, inducing consumers to lapse. Similarly, as discussed in Subsection 4.2.5, the equilibrium of the rational expectations model of income shocks provides subsidized loans to reduce the risk of income shocks. In contrast, in the differential attention model, the equilibrium exacerbates the effect of of income shocks by providing above-market rates, effectively transferring consumption from the states with large losses (where marginal utility is high) to those with small losses (where marginal utility is low).

[39]On February 2, 2010, the American Council of Life Insurance, representing 300 large life insurance companies, released a statement asking policymakers to ban the securitization of life settlement contracts. Life insurance industry organizations have also organized media campaigns warning the public and investors about life settlements. The opposition to life settlements contrasts with some other markets, where firms encourage the development of secondary markets. The market for initial public offerings, for example, would be substantially smaller without the ability to resell securities.

long run, firms earn zero profits in both primary and secondary markets, while consumers are better off. However, in the short run, the introduction of a secondary market makes consumers better off and primary market insurers worse off. It is, therefore, not surprising that the life insurance industry has waged an intense lobbying effort aimed at state legislatures, where life insurance is regulated in the United States, to try to ban life settlement contracts. Of course, when lapses are due to forgetting to pay, a secondary market does not help. Instead, to deal with forgetfulness, regulators could impose other rules on lapses, including mandatory reminders and a penalty for policies with "excessive" lapses.

A potential political-economy issue with interventions in these markets is that consumers who do not think they will not lapse may not favor regulations that appear to undermine cross-subsidies that are favorable to these consumers. Therefore, the same behavioral trait that introduces inefficiency in the competitive equilibrium also prevents consumers from supporting the implementation of efficiency-enhancing regulations.[40] More specifically, whether consumers would support such a regulation depends on why they think firms offer these contracts. Consumers who are overconfident about their rationality, thinking that others will lapse but they will not, would be against these regulations. Consumers who do not update their beliefs based on the policies they are offered will also not see a benefit from these regulations. If consumers can be fully educated about their lapse probabilities, there is a scope for stronger efficiency-enhancing regulations. However, if it is easier to educate consumers who are less likely to lapse then, again, there might be limited support for such policies.

5 Other Potential Explanations

Our goal in this paper is to study mechanisms that can simultaneously explain both the demand and the supply side of the life insurance market. There are several possible reasons for a consumer to prefer a lapse-based life insurance policy, holding the design of policies fixed, or for a life insurer to offer a lapse-based policy, taking consumers' decisions as fixed. It is considerably harder to provide a unified account of both consumers' and the life insurers' decisions. Section 2 already discussed shortcomings of rational models in detail. This section discusses other potential explanations.

5.1 Present Bias

A large literature in behavioral economics has established that illiquid assets may be valuable to time-inconsistent individuals because they serve as commitment devices (c.f., DellaVigna and Malmendier, 2004, and Heidhues and Kőszegi, 2010). Since front-loaded premiums reduce the incentive to drop the policy, time inconsistency may, at first glance, explain why insurance policies are front loaded. Moreover, one might think that consumer naiveté could explain why individuals lapse on their policies.

[40] See, for example, Warren and Wood (2014) and Bisin, Lizzeri, and Yariv (2015) for interesting analyses of political economy based on behavioral economics models.

Gottlieb and Zhang (2019) study a very general contracting model with partially naive present-biased individuals and show that the equilibrium consumption converges to the one that maximizes their long-term preferences.[41] As they show, equilibrium contracts offer two repayment options in each period to exploit consumer naiveté. Consumers think they will make most of the payments in the immediate future, but end up delaying until the last period, when they consume less than the efficient amount. Since periods in quasi-hyperbolic models are thought to be very short, typically no more than a day (O'Donoghue and Rabin, 2015), and life-insurance contracts usually last for many years, their model predicts essentially the same consumption path as in the model with no present bias. Therefore, present bias by itself cannot generate more lapses than with time-consistent consumers.

Moreover, controlling for impatience, lapses happen *more* often for time-consistent agents than for naive time-inconsistent agents. The reason is that a naive time-inconsistent agent mispredicts his future choices, overestimating the value from keeping the contract. As a result, that agent is actually less tempted to lapse.[42] In sum, present-biased preferences by themselves cannot explain why life insurance policies are front loaded or why individuals lapse.

5.2 Fixed Costs

Insurance companies may also charge surrender fees in order to recover sales commissions paid to brokers. But there are two problems with this explanation for the life insurance pricing observed in practice.

First, commissions are endogenous. An explanation for front-loaded premiums that is based on the fact that sales commissions are front loaded needs to justify why commissions are front loaded in the first place. For example, commissions paid to wealth managers are fairly proportional to the actual fee revenue collected from clients.[43] With life insurance, the bulk of the commission payment is typically made in the first couple years of the policy, except if the policy is surrendered in the first year.[44] In fact, consistently with our model, paying front-loaded commissions may be an effective way to encourage insurance brokers to sell to consumers with short-term horizons.

Second, since the bulk of commissions are paid early into the policy, lapse fees paid by consumers should be constant after the first few years, when commissions are no longer paid. That is, according

[41] With sophisticated consumers and a delay between signing the contract and consuming (as with the contracting periods in the models above), the equilibrium maximizes the individual's long-term preferences (see DellaVigna and Malmendier, 2004 and Heidhues and Kőszegi, 2010).

[42] Controlling for impatience, sophisticated agents are also less likely to lapse than time-consistent agents because they value commitment.

[43] With broker-dealers, the client typically pays an initial fee along with a trailer fee that is proportional to ongoing assets under management. With fiduciary financial advisors, clients typically pay just a fee that is proportional to their assets being managed. In both cases, the wealth advisor collects a proportion of the revenue collected from clients, and so the product provider does not actually lose money if the client leaves. Moreover, all wealth managers are incentivized to keep clients active because of the potential to collect ongoing fees.

[44] For example, Genworth Life's (2011) commission schedule reads: "In the event a withdrawal or partial surrender (above any applicable penalty-free amount) is granted or a policy or contract is surrendered or canceled within the first twelve (12) months after the date specified in paragraph (c) of this Section 2, compensations will be charged back to you as follows: 100% of compensations paid during that twelve (12) month period."

to the commissions explanation, insurance firms should not obtain different actuarial profits if rational consumers lapse after 5, 10, or 20 years since they do not have to pay any additional commissions after the first few years. Empirically, however, actuarial profits are substantially different if policies lapse after 5, 10, or 20 years (see Figure 3).

Another potential reason for insurance companies to charge large surrender fees is to compensate for the cost of liquidating their investments. An insurer with more predictable lapses can obtain higher returns on its portfolio by making more illiquid investments. However, consumers with rational expectations about their liquidity needs value the flexibility of being able to recover some of their front-loaded payments when needed. Therefore, the optimal surrender fee should balance the gain to the insurance company's portfolio against the cost of preventing policyholders from adjusting their consumption after liquidity shocks.[45]

To test this prediction, we collected detailed whole life insurance policy data from two national insurance companies, MetLife and SBLI.[46] It is widely documented that younger individuals are more likely to be liquidity constrained, and age is a frequently-used proxy for the presence of liquidity constraints.[47] Online Appendix D outlines the data collection and results in detail. The results are inconsistent with the rational model with the cost of liquidating investments. Younger individuals face higher surrender fees (Online Appendix D, Figure 6). The differences by age are statistically significant and economically large. While a 20-year-old policyholder who surrenders after 5 years would not collect any cash value, a 70-year-old who surrenders after 5 years collects about 30% of the amount paid in excess of actuarially-fair prices. Moreover, because larger policies require more liquid assets to be held by insurers in order to repay those who surrender, and because larger policies are typically purchased by wealthier individuals with lower liquidity needs, we should expect surrender fees to *increase* with policy size. In practice, surrender fees weakly decrease with policy size (Online Appendix D, Figure 7).[48]

[45] A similar argument can be made about rational models based on reclassification risk. As Online Appendix C documents, younger people are quite likely to remain healthy; health shocks only become material at older ages. Because younger individuals tend to be more liquidity constrained, the cost of preventing them from adjusting their consumption after a liquidity shock is relatively high. In Online Appendix C, we extend the models of Hendel and Lizzeri (2003), Daily, Hendel, and Lizzeri (2008), and Fang and Kung (2010) by adding an initial period in which consumers are subject to an unobservable liquidity shock. Consumers are then subject to liquidity shocks in the first period, health shocks in the second period, and mortality risk in the third period—a stylized representation of the fact that health shocks are more important later in life. We show that the reclassification risk model predicts that policies should not charge a positive lapse fee if the individual decides to lapse early on. The reason is that lapse fees exist to penalize agents who lapse due to favorable health shocks, thereby ensuring that the pool remains balanced. Charging a lapse fee for non-health related shocks is inefficient, as it exacerbates the consumer's demand for money and undermines the amount of insurance provision. So, consumers who are more likely to suffer liquidity shocks—e.g., younger consumers and those who buy smaller policies—should face *lower* surrender fees.

[46] The choice of these two firms was dictated by data availability.

[47] See, for example, Jappelli (1990), Jappelli, Pischke, and Souleles (1998), Besley, Meads, and Surico (2010), and Zhang (2017).

[48] The decreasing relationship between surrender fees and coverage can be explained if costs are constant rather than proportional to the size of the policy. This explanation, however, cannot account for the strong decreasing relationship between age and surrender fees.

6 Conclusion

This paper documents that there is substantial lapsing in the life insurance market and lapse-supported pricing. We argued, both empirically and theoretically, that the rational model cannot explain key observable features of the life insurance market: direct survey evidence of anticipated lapse rates by new buyers as well as reasons for lapses by recent lapsers; the lack of *positive* correlation between lapses and health; the sheer amount of lapses relative to a realistically calibrated rational models; and, the counterfactual predictions about the life insurance policy *structure* including the pattern of surrender fees, the structure of policy loans, and the direction of cross-subsidies between consumers. We study several mechanisms that can potentially account for these salient features. Two mechanisms jointly account for a small majority of lapses: (i) consumers forgetting to pay their premiums or (ii) underestimating the likelihood of needing money during the policy period. They can also account for the structure of policies, theoretically.

References

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." American Economic Review, 101(4): 1180-1210.

American Council of Life Insurers, 2011, 2012, 2015, 2017, 2018. *Life Insurers Fact Book.* Washington, DC.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein, 2015. "Behavioral Hazard in Health Insurance," Quarterly Journal of Economics, 130 (4): 1623-1667.

Barberis, Nicholas, Ming Huang, and Richard Thaler, 2006. "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing," American Economic Review 96: 1069-1090.

Basu, Rashmita, 2016. "Lapse of Long-Term Care Insurance Coverage in the US." Working Paper, Texas A&M Health Science Center.

Bénabou, Roland and Jean Tirole, 2002. "Self-Confidence and Personal Motivation." Quarterly Journal of Economics, 117(3): 871-915.

Bénabou, Roland and Jean Tirole, 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs," Journal of Economic Perspectives, 30 (3), 141-163.

Besley, Timothy, Neil Meads, and Paolo Surico, 2008. "Household external finance and consumption." Working Paper, London School of Economics, Bank of England, and London Business School.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor, 2015. "Do Individuals Make Sensible Health Insurance Decisions? Evidence from a Menu with Dominated Options," NBER Working Paper No. 21160.

Bisin, Alberto, Alessandro Lizzeri and Leeat Yariv, 2015. "Government Policy with Time Inconsistent Voters," American Economic Review, 105(6): 1711–1737

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2016. "Competition for Attention." Review of Economic Studies, 83(2): 481-513.
Compulife, 2013. Individual Agent Version. https://compulife.com/product-pricing/individual-agent-version/ [URL last checked: 2021/01/09].
Conlon, John J., Laura Pilossoph, Matthew Wiswall, and Basit Zafar, 2018. "Labor Market Search With Imperfect Information and Learning," Working Paper.
Canadian Institute of Actuaries, 2007. "Lapse Experience under Universal Life Level Cost of Insurance Policies." Research Committee Individual Life Subcommittee Report.
Conning Research & Consulting, 2009. "Life Settlements, It's A Buyer's Market for Now." Report.
Daily, Glen, Igal Hendel, and Alessandro Lizzeri, 2008. "Does the Secondary Life Insurance Market Threaten Dynamic Insurance?" American Economic Review Papers and Proceedings, 98: 2, 151-156.
DellaVigna, Stefano, and Ulrike Malmendier, 2004. "Contract Design and Self-Control: Theory and Evidence," Quarterly Journal of Economics 119: 353-402.
Diamond, Rebecca, Michael J. Dickstein, Timothy McQuade and Petra Persson, 2018. "Take-Up, Drop-Out, and Spending in ACA Marketplaces." NBER Working Paper 24668.
Doherty, Neil and Georges Dionne, 1994. "Adverse Selection, Commitment, and Renegotiation: Extension to and Evidence from Insurance Markets," Journal of Political Economy, 102(2): 209-235.
Doherty, Neil and Hal J. Singer, 2002. "The Benefits of a Secondary Market For Life Insurance Policies." The Wharton School, Financial Institutions Center, WP 02-41.
Eggleston, Jonathan and Robert Munk. "New Worth of Households: 2014." Census Department, 2018.
Ellison, Glenn, 2005. "Bounded Rationality in Industrial Organization," in Whitney Newey, Torsten Persson, and Richard Blundell eds., Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress (Cambridge, UK: Cambridge University Press, 2005).
Eliaz, Kfir and Ran Spiegler, 2006. "Contracting with Diversely Naive Agents," Review of Economic Studies, 73(3): 689-714.
—–, 2008. "Consumer Optimism and Price Discrimination," Theoretical Economics, 3: 459-497.
—–, 2011. "On the strategic use of attention grabbers," Theoretical Economics, 6: 127–155.
Ellison, Glenn, 2005. "A Model of Add-On Pricing." Quarterly Journal of Economics, 120(2): 585 - 637.
Ericson, Keith M. 2011. "Forgetting We Forget: Overconfidence and Prospective Memory." Journal of the European Economic Association, 9(1), 43-60.
Ericson, Keith M. and Amanda Starc, 2012. "Heuristics and Heterogeneity in Health Insurance Exchanges: Evidence from the Massachusetts Connector," American Economic Review, 102(3): 493-97.
Eyster, Erik and Georg Weizsäcker, 2011. "Correlation Neglect in Financial Decision-Making." Working Paper, London School of Economics and University College London.
Fang, Hanming and Edward Kung, 2010. "The Welfare Effect of Life Settlement Market: The Case of Income Shocks," NBER Working Paper 15761.

——, 2012. "Why Do Life Insurance Policyholders Lapse? The Roles of Income, Health and Bequest Motive Shocks," NBER Working Paper 17899.

Fier, Stephen G. and Andre P. Liebenberg, 2013. "Life Insurance Lapse Behavior," North American Actuarial Journal, 17(2): 153-167.

Friedberg, Leora, Wenliang Hou, Wei Sun and Anthony Webb. "Lapses in Long-Term Care Insurance." Working Paper # 2017-08, Schwartz Center for Economic Policy Analysis (SCEPA), The New School.

Gabaix, Xavier and David Laibson, 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," Quarterly Journal of Economics 121: 505-540.

Genworth Life Insurance Company, 2011. "Commission Schedule." Dated May 19, 2011.

Gigante, Shelly, 2018. "Life insurance: Treat cash value with care." MassMutual, Blog. https://blog.massmutual.com/post/life-insurance-treat-cash-value-with-care. [Last checked: 9/22/2018]

Grubb, Michael, 2009. "Selling to Overconfident Consumers," American Economic Review, 99(5): 1770-1807.

——, 2015. "Overconfident Consumers in the Marketplace," Journal of Economic Perspectives, 29(4): 9-36.

Hambel, Christopher, Holger Kraft, Lorenz Schendel, and Mogens Steffensen, 2016. "Life Insurance Demand under Health Shock Risk." Journal of Risk and Insurance, 84 (4): 1171 - 1202.

Handel, Benjamin R. and Jonathan T. Kolstad, 2015. "Health Insurance for 'Humans': Information Frictions, Plan Choice, and Consumer Welfare," American Economic Review, vol. 105(8): 2449-2500.

Handel, Benjamin, Igal Hendel, and Michael D. Whinston, 2015. "Equilibria in Health Exchanges: Adverse Selection vs. Reclassification Risk" Econometrica, 83(4): 1261-1313.

Health and Retirement Study (various years), RAND Biennial Data for years 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010 and 2012, [https://hrsdata.isr.umich.edu/data-products/rand], RAND HRS Family Data 2014 [https://hrsdata.isr.umich.edu/data-products/rand-hrs-family-data-2014], and RAND HRS Detailed Imputations File 2016 [https://hrsdata.isr.umich.edu/data-products/rand-hrs-detailed-imputations-file-2016]. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). [last checked: 01/09/2021]

Hendel, Igal and Alessandro Lizzeri, 2003. "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance." Quarterly Journal of Economics, 118(1): 299-327.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka. 2016. "Inferior Products and Profitable Deception," Review of Economic Studies, 84(1): 323-356.

Heidhues, Paul and Botond Kőszegi, 2010. "Exploiting NaÃ¯vete about Self-Control in the Credit Market." American Economic Review, 100(5): 2279 - 2303.

Howard, Kevin, 2006. "Pricing Lapse Supported Products: Secondary Guarantee Universal Life," Presentation to the October 2006 Annual Meeting of the Society of Actuaries, session titled "Pricing Lapse Supported / Lapse-Sensitive Products."

Keeton, R.E., 1970. “Insurance law rights at variance with policy provisions.” Harvard Law Review, 83(5): 961 - 985.

Konetzka, R. Tamara, and Ye Luo, 2011. “Explaining lapse in long-term care insurance markets.” Health Economics 20(10): 1169-1183.

Insurance Studies Institute, 2009. “Portrayal of Life Settlements in Consumer-Focused Publications,” Report, September 10, Keystone, Colorado.

Insurance Information Institute, 2018. “2018 Insurance Fact Book.” https://www.iii.org/publications/2018-insurance-fact-book. [Last checked: 2-25-2019]

InvestmentNews, 2011. “Court to insurer: you can’t triple life premiums,” January 23, 2011.

—–, 2012. “Long-term-care insurance suddenly short on sellers,” March 8, 2012.

Huynh, Kim and Juergen Jung, 2014. “Subjective Health Expectations,” Working Paper, Bank of Canada.

Kőszegi, Botond, 2014. “Behavioral Contract Theory,” Journal of Economic Literature, 52(4): 1075-1118.

Krebs, Tom, Moritz Kuhn and Mark L.J. Wright, 2015. “Human Capital Risk, Contract Enforcement, and the Macroeconomy.” American Economic Review, November, 105(11): 3223-3272

LeBel, Dominique, 2006. “Pricing Pricing Lapse Supported / Lapse-Sensitive Products.” Presentation to the October 2006 Annual Meeting of the Society of Actuaries, session titled “Pricing Lapse Supported / Lapse-Sensitive Products.”

Lifequotes, 2013. URL: www.Lifequotes.com. [Data downloaded on 9/22/2013].

LIMRA, 2011A. “U.S. Individual Life Insurance Persistency: A Joint Study Sponsored by the Society of Actuaries and LIMRA.” Windsor, CT.

—–, 2011B. “Person-Level Trends in U.S. Life Insurance Ownership.” Windsor, CT.

—–, 2012. “U.S. Individual Life Insurance Persistency: A Joint Study Sponsored by Society of Actuaries and LIMRA.” Windsor, CT. https://www.soa.org/resources/experience-studies/2012/2007-09-us-individual-life-persistency-update/ [Last checked: 2-6-2021]

—–, 2014. “The Facts of Life and Annuities.” Windsor, CT.

—–, 2016. “Trends in Life Insurance Ownership.” Windsor, CT.

LLI, 2018. URL: https://github.com/smetterspa/LLI_Survey. [Last checked: 2/6/2021].

Mahony, Mark, 1998. “Current Issues in Product Pricing,” Record of the Society of Actuaries, 24(3): 1 - 20.

Milliam USA, 2004. Correspondence to Coventry, Dated February 19, 2004.

Mueller, Andreas, Johannes Spinnewijn, and Giorgio Topa, 2019. “Job Seekers’ Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias,” Working Paper.

National Underwriter Company, 2008. *Tools and Techniques of Life Settlements Planning*. October, 2008. Erlanger, KY.

Jappelli, Tullio, 1990. “Who is Credit-Constrained in the US Economy?” Quarterly Journal of Economics, 105(1): 219–34.

Jappelli, Tullio, Jorn-Steffen Pischke, and Nicholas S. Souleles, 1998. “Testing for Liquidity Constraints in Euler Equations with Complementary Data Sources,” Review of Economics and Statistics, 80(2): 251-62.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman, 2016. "Getting to the top of mind: How reminders increase saving." Management Science 62(12): 3393-3411.

Ortoleva, Pietro and Erik Snowberg, 2015. “Overconfidence in Political Behavior.” American Economic Review, 105(2): 504–535.

Rabin, Matthew and Georg Weizsäcker, 2009. “Narrow Bracketing and Dominated Choices.” American Economic Review, 99(4): 1508 - 1543.

Read, Daniel, George Loewenstein, and Matthew Rabin, 1999. “Choice bracketing,” In: Fischoff, Baruch and Charles Manski, “Elicitation of Preferences,” Spinger Verlag: 171 - 197.

Robinson, Jim, 1996. “A Long-Term-Care Status Transition Model.” Society of Actuaries, 1996 Bowles Symposium: 72 - 79. Data tables located at https://demonstrations.wolfram.com/LifeTransitions/ [last checked: 01/19/2021]

Sandroni, Alvaro and Francesco Squintani, 2007. “Overconfidence, Insurance, and Paternalism,” American Economic Review, 97(5): 1994-2004.

Society of Actuaries, 1998. “Session 133PD: Current Issues in Product Pricing.” RECORD, (24), 3: 1 - 20.

—–, 2008. “2008 Valuation Basic Tables [VBT] Report.” https://www.soa.org/globalassets/assets/files/zip/research-2008-vbt-report-app-f.zip [last checked: 01/09/2021].

—–, 2015. “Life Insurance Regulatory Structures and Strategy: EU Compared with US.” .

Spinnewijn, Johannes, 2015. “Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs,” Journal of the European Economic Association, 13(1): 130-167.

Sussman, Abigail B, and Adam L Alter, 2012. “The Exception Is the Rule: Underestimating and Overspending on Exceptional Expenses.” Journal of Consumer Research 39(4): 800–814.

Tversky, Amos, and Daniel Kahneman, 1983. “Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment.” Psychological Review 90(4): 293-315.

U.S. Census, 2011. “Wealth and Asset Ownership.” www.census.gov/people/wealth/data/dtables.html [last checked: 6-1-2016]

Wall Street Journal, 2000. “Unexpected Rate Rises Jolt Elders Insured for Long-Term Care.” June 22.

Warren, Patrick L. and Daniel H. Wood, 2014. “The Political Economy of Regulation in Markets with NaÃ¯ve Consumers.” Journal of the European Economic Association 12(6): 1617-1642.

Zhang, C. Yiwei, 2017. “Consumption Responses to Pay Frequency: Evidence from ‘Extra’ Paychecks.” ACR North American Advances.

Appendix: Outline of Proof to Theorem 1

Let $I \equiv I_1 + I_2$ denote the individual's total future income at the time of contracting. To keep the notation close to the optimal control literature, we associate each possible *loss* with a "type" t. Types are distributed according to a differentiable PDF f with full support in the interval of possible losses $[0,T]$, where $0 < T < I$. For each possible loss $t \in [0,T]$, let $c(t)$ denote the consumption in period 1, let $c_A(t)$ denote the consumption in period 2 if the individual is alive, and let $c_D(t)$ denote the bequests in period 2 (i.e., the resources repaid if the consumer dies). Let $V(t) \equiv \alpha u(c_D(t)) + (1-\alpha)u(c_A(t))$ denote the continuation payoff of a consumer who gets a loss of t.

Since firms do not observe the loss, contracts must satisfy the incentive-compatibility (IC) constraint:

$$u(c(t)) + V(t) \geq u(c(\hat{t}) - t + \hat{t}) + V(\hat{t}) \ \forall t, \hat{t}.$$

It is helpful to work with each type's indirect utility $\mathscr{U}(t) \equiv u(c(t)) + V(t)$. As usual in mechanism design, we can characterize the consumer's incentive compatibility constraint in terms of an envelope condition and a monotonicity condition. Formally, IC holds if and only if

$$\dot{\mathscr{U}}(t) = -u'(c(t)) \text{ and } c(t) + t \text{ is non-decreasing in } t.$$

Therefore, incentive compatibility alone has the following implications. The amount paid in period 1 ($I_1 - c(t) - t$) is decreasing in the loss, meaning that people with larger income shocks pay less in period 1. This lower payment is associated with a lower future consumption: $\dot{V}(t) = -u'(c(t))[1 + \dot{c}(t)] \leq 0$. Intuitively, since types with larger shocks borrow more in period 1, they have to repay what they borrowed in period 2.

To obtain the renegotiation proofness constraint, note that the best contract that an individual can obtain if he lapses perfectly smooths his remaining income: $c_A = c_D = \frac{I-t}{2}$. Therefore, the renegotiation proofness constraint is:

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right) \ \forall t.$$

The equilibrium contract must provide full insurance in the second period: $c_D(t) = c_A(t) =: c_2(t)$ (otherwise, it is possible be possible to keep the same promised continuation utility at a lower cost). Using the definition of the indirect utility, we can write the zero profits constraint as:

$$\int \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) + t\right] f(t) dt \leq W + I.$$

As in Lemma 2, the equilibrium must maximize the utility conditional on no loss subject to the constraints above:

$$\max_{c, \mathscr{U}} \mathscr{U}(0)$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)), \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right), \tag{RP}$$

$$\int \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) + t\right] f(t)dt \leq W + I, \tag{Zero Profits}$$

and $c(t)+t$ non-decreasing.

We follow the standard approach of ignoring the monotonicity constraint, which can be verified ex-post. It is convenient to work with the dual program that minimizes the cost of providing perceived utility $\mathscr{U}(0)$:

$$\min_{c,\mathscr{U}} \int_0^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) + t\right]$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)), \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right), \tag{RP}$$

$$\mathscr{U}(0) \geq \bar{u}. \tag{8}$$

This is an optimal control problem with state variable $\mathscr{U}$ and control variable c. Equation (RP) is a pure state inequality constraint, since it involves the state variable $\mathscr{U}$ and "time" t but not the control variable c. Pure state constraints pose some challenges to the analysis of the solution, because they impose indirect constraints to the path of the control. Here, in any interval in which (RP) binds, the individual lapses and consumes half his future income in each period. Intuitively, at any point immediately before (RP) binds, the individual's period-1 consumption must be less than half his income (his utility must be growing at a lower rate than his outside option). Similarly, at any point immediately after (RP) binds, his period-1 consumption must exceed half his income.

In the Online Appendix A, we solve this program in detail. The first key result shows that period-1 consumption never exceeds the consumption that perfectly smooths income:

$$u(c(t)) \leq \frac{\mathscr{U}(t)}{2}. \tag{9}$$

Second, we show that if (RP) binds for one type, it must also bind for all higher types. Along with the characterization of incentive compatibility, this means that policy loans are provided up to a cap. Individuals with small losses borrow from their policies whereas those with large losses lapse. The third key result shows that (9) holds with equality if and only if (RP) binds. That is, policy loans are provided at above-market rates so that individuals under-consume in period 1 relative to perfect smoothing.

ONLINE APPENDICES

Appendix A: Proofs

Characterization of Equilibrium

In Subsections 4.1.3 and 4.2.2, we stated and informally described the equivalence between the equilibrium consumption and the solution of the equilibrium program.

Lemma 2. *Consider either the model of forgetting to repay (Subsection 4.1.3) or the model of income shocks (Subsection 4.2.2). A set of state-dependent consumption* $\{\mathbf{C}_j\}_{j=1,\dots,N}$ *and a set of acceptance and payment decisions is an equilibrium of the game if and only if:*

1. *At least two offers are accepted with positive probability,*
2. *All offers accepted with positive probability solve the equilibrium program, and*
3. *All offers that are not accepted give consumers a perceived utility lower than the solutions of the equilibrium program.*

Since firms get zero profits in equilibrium, when there are more than two firms, there exist equilibria in which some firms offer contracts that are never accepted. An equilibrium of the game is *essentially unique* if consumption in any contract accepted with positive probability is the same in all equilibria. An equilibrium of the game is *symmetric* if consumption in all contracts accepted with positive probability is the same: if $\mathbf{C}_j$ and $\mathbf{C}'_{j'}$ are accepted with positive probability, then $\mathbf{C}_j = \mathbf{C}'_{j'}$. The next lemma establishes existence, uniqueness, and symmetry of the equilibrium:

Lemma 3. *Consider either the model of forgetting to repay (Subsection 4.1.3) or the model of income shocks (Subsection 4.2.2). There exists an equilibrium. Moreover, the equilibrium is essentially unique and symmetric.*

Lemma 3 justifies the approach in the paper of omitting the index j from contracts that are accepted with positive probability.

Proof of Lemma 2

We establish the result for the model in which individuals forget to repay. The proof for the model with income shocks is analogous and therefore omitted.

The proof follows a standard Bertrand argument. Necessity:

1. First, suppose no offer is accepted in equilibrium. Then, a firm can get positive profits by offering full insurance conditional on remembering to pay and no insurance if the individual forgets to pay at a price slightly above actuarially fair. Since the perceived utility function is concave (consumers are risk averse), we can ensure that consumers buy the policy by taking prices to be close enough to actuarially fair. Next, suppose only one offer is accepted. If this offer yields strictly positive profits, another firm can profit by offering a policy with a slightly higher consumption, thereby attracting all customers. If the only offer that is accepted in equilibrium yields zero profits, there are two possibilities. If the policy gives a strictly positive perceived utility to consumers, the firm offering it can obtain strictly positive profits by charging a slightly higher price (say, at period 1). If, instead, the policy gives a zero perceived utility to consumers, because consumers are risk averse, the firm offering it can obtain strictly positive profits by shifting to a policy that offers full insurance conditional for those who remember to pay and no insurance for those that forget to pay. Therefore, at least two offers must be accepted with positive probability in equilibrium.

2. Because consumers put zero weight on forgetting to pay when they are choosing which policy to buy, they do not take into account any states that happen after they forget to pay. Therefore, policies must maximize the consumers' perceived utility (which attributes probability zero to forgetting to pay) subject to the incentive constraint.

Firms are willing to provide insurance policies as long as they obtain non-negative profits. If an offer with strictly positive profits is accepted in equilibrium, another firm can obtain a discrete gain by slightly undercutting the price of this policy. Moreover, if the policy does not maximize the consumer's perceived utility subject to the zero-profits constraint, another firm can offer a policy that yields a higher perceived utility and extract a positive profit.

3. If a consumer is accepting an offer with a lower perceived utility, either a policy that solves Program (2) is being rejected (which is not optimal for the consumer) or it is not being offered (which is not optimal for the firms).

To establish sufficiency, note that whenever these conditions are satisfied, any other offer by another firm must either not be accepted or yield non-positive profits.

Proof of Lemma 3

We establish the result for the model in which individuals forget to repay. The proof for the model with income shocks is analogous and therefore omitted.

From Lemma 2, the equilibrium must solve the following program:

$$\max_{\left(c_{1,j},c_{2,j}^{R},c_{3,j}^{R,D},c_{3,j}^{R,A},c_{3,j}^{F,D},c_{3,j}^{F,A}\right)} u_A(c_{1,j}) + u_A(c_{2,j}^{R}) + \alpha u_D(c_{3,j}^{R,D}) + (1-\alpha)\, u_A(c_{3,j}^{R,A}) \tag{10}$$

subject to:

$$c_{1,j}+(1-l)\left[c_{2,j}^{R}+\alpha c_{3,j}^{R,D}+(1-\alpha)\,c_{3,j}^{R,A}\right]+l\left[I_2+\alpha c_{3,j}^{F,D}+(1-\alpha)\,c_{3,j}^{F,A}\right]\leq I_1+I_2+(1-\alpha)\,I_3. \tag{Zero profits}$$

$$u_A(c_{2,j}^{R})+\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq u_A(I_2)+\alpha u_D(c_{3,j}^{F,D})+(1-\alpha)\,u_A(c_{3,j}^{F,A}). \tag{IC}$$

$$u_A(c_{2,j}^{R})+\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq V, \tag{RP_1}$$

$$\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq U\,(I_3) \tag{RP_R}$$

and

$$\alpha u_D(c_{3,j}^{F,D})+(1-\alpha)\,u_A(c_{3,j}^{F,A})\geq U\,(I_3)\,. \tag{RP_F}$$

By a duality argument, the solution of this program solves:

$$\min_{(c_{1,j},c_{2,j}^{R},c_{3,j}^{R,D},c_{3,j}^{R,A},c_{3,j}^{F,D},c_{3,j}^{F,A})} c_{1,j}+(1-l)\left[c_{2,j}^{R}+\alpha c_{3,j}^{R,D}+(1-\alpha)\,c_{3,j}^{R,A}\right]+l\left[I_2+\alpha c_{3,j}^{F,D}+(1-\alpha)\,c_{3,j}^{F,A}\right]$$

subject to

$$u_A(c_{1,j})+u_A(c_{2,j}^{R})+\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq \bar{u},$$

$$u_A(c_{2,j}^{R})+\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq u_A(I_2)+\alpha u_D(c_{3,j}^{F,D})+(1-\alpha)\,u_A(c_{3,j}^{F,A}), \tag{IC}$$

$$u_A(c_{2,j}^{R})+\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq V, \tag{RP_1}$$

$$\alpha u_D(c_{3,j}^{R,D})+(1-\alpha)\,u_A(c_{3,j}^{R,A})\geq U\,(I_3) \tag{RP_R}$$

and

$$\alpha u_D(c_{3,j}^{F,D})+(1-\alpha)\,u_A(c_{3,j}^{F,A})\geq U\,(I_3)\,. \tag{RP_F}$$

for some $\bar{u}\in\mathbb{R}$. Letting $v_A\equiv u_A^{-1}$ and $v_D\equiv u_D^{-1}$, we can rewrite this program in terms of utils rather than consumption:

$$\min_{(u_{1,j},u_{2,j}^{R},u_{3,j}^{R,D},u_{3,j}^{R,A},u_{3,j}^{F,D},u_{3,j}^{F,A})} \begin{array}{l} v_A\left(u_{1,j}\right)+(1-l)\left[v_A\left(u_{2,j}^{R}\right)+\alpha v_D\left(u_{3,j}^{R,D}\right)+(1-\alpha)\,v_A\left(u_{3,j}^{R,A}\right)\right] \\ +l\left[I_2+\alpha v_D\left(u_{D,j}^{F}\right)+(1-\alpha)\,v_A\left(u_{A,j}^{F}\right)\right] \end{array}$$

subject to

$$u_{1,j}+u_{2,j}^{R}+\alpha u_{3,j}^{R,D}+(1-\alpha)\,u_{3,j}^{R,A}\geq \bar{u},$$

$$u_{2,j}^{R}+\alpha u_{3,j}^{R,D}+(1-\alpha)\,u_{3,j}^{R,A}\geq u_A(I_2)+\alpha u_{3,j}^{F,D}+(1-\alpha)\,u_{3,j}^{F,A}, \tag{IC}$$

$$u_{2,j}^{R}+\alpha u_{3,j}^{R,D}+(1-\alpha)\,u_{3,j}^{R,A}\geq V, \tag{RP_1}$$

$$\alpha u_{3,j}^{R,D}+(1-\alpha)\,u_{3,j}^{R,A}\geq U\,(I_3) \tag{RP_R}$$

and

$$\alpha u_{3,j}^{F,D} + (1-\alpha)\, u_{3,j}^{F,A} \geq U(I_3)\,. \tag{RP_F}$$

It is straightforward to verify that the set of vectors satisfying the linear constraints is non-empty (in fact, we will construct one such vector below). Therefore, this program corresponds to the minimization of a continuous function in a non-empty and compact set, so a solution exists. Moreover, because the objective function is strictly convex and the feasibility set is convex, the solution is unique.

Proof of Lemma 1

We start with the renegotiation proofness constraint in period 1:

$$u_A(c_2^R) + \alpha u_D(c_D^{3,R}) + (1-\alpha)\, u_A(c_A^{3,R}) \geq V, \tag{RP_1}$$

where the outside option V solves the subprogram:

$$V \equiv \max_{\tilde{c}_2^R, \tilde{c}_D^{3,R}, \tilde{c}_A^{3,R}} u_A(\tilde{c}_2^R) + \alpha u_D(\tilde{c}_3^{R,D}) + (1-\alpha)\, u_A(\tilde{c}_3^{R,A})$$

subject to

$$(1-l)\left[\tilde{c}_2^R + \alpha \tilde{c}_3^{R,D} + (1-\alpha)\, \tilde{c}_3^{R,A}\right] + l\left[I_2 + \alpha \tilde{c}_3^{F,D} + (1-\alpha)\, \tilde{c}_3^{F,A}\right] \leq I_2 + I_3 \tag{Zero profits}$$

$$\alpha u_D(\tilde{c}_3^{R,D}) + (1-\alpha)\, u_A(\tilde{c}_3^{R,A}) \geq U(I_3) \tag{RP_R}$$

$$\alpha u_D(\tilde{c}_3^{F,D}) + (1-\alpha)\, u_A(\tilde{c}_3^{F,A}) \geq U(I_3) \tag{RP_F}$$

Note that (RP_F) must bind in the program above (otherwise, it would be possible to increase the objective by transferring consumption from $\tilde{c}_3^{F,D}$ or $\tilde{c}_3^{F,A}$ to $\tilde{c}_3^{R,D}$ or $\tilde{c}_3^{R,A}$).

There must be perfect consumption smoothing after forgetting to pay—i.e., $u_D'(\tilde{c}_3^{F,D}) = u_A'(\tilde{c}_3^{F,A})$—, which minimizes the cost of providing utility $U(I_3)$ in constraint (RP_F). There must also be perfect consumption smoothing after remembering to pay, which, not only minimizes the cost of providing utility $U(I_3)$ in (RP_S) but also maximizes the objective function.

Let $c_3^R \equiv \alpha c_3^{R,D} + (1-\alpha)\, c_3^{R,A}$ denote the average consumption in period 3 when the individual remembers to pay (which he believes will happens with probability 1). Using the previous observations, we can rewrite the outside option as

$$V \equiv \max_{\tilde{c}_2^R, \tilde{c}_3^R} u_A(\tilde{c}_2^R) + U(\tilde{c}_3^R) \tag{11}$$

subject to

$$\tilde{c}_2^R + \tilde{c}_3^R \leq I_2 + I_3 \tag{12}$$

$$\tilde{c}_3^R \geq I_3 \tag{13}$$

where (12) is the zero profits condition and (13) is the renegotiation proofness constraint (RP_R).

Recall that the equilibrium program is:

$$\max_{\left(c_1, c_2^R, c_3^{R,D}, c_3^{R,A}, c_3^{F,D}, c_3^{F,A}\right)} u_A(c_1) + u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A})$$

subject to (*Zero Profits*), (RP_R), (RP_F), (RP_1), and (*IC*).

As in the renegotiation program considered above, (RP_F) must bind. Otherwise, it would be possible to relax both (IC) and (*Zero Profits*) by reducing $c_3^{F,D}$ or $c_3^{F,A}$, while not affecting any other constraint or the objective.

Note also that solution must provide full insurance against mortality conditional on forgetting to pay: $u_D'(c_3^{F,D}) = u_A'(c_3^{F,A})$ with $\alpha c_3^{F,D} + (1-\alpha)\, c_3^{F,A} = I_3$. This follows from the fact that the solution must minimize the cost of providing utility $U(I_3)$ in case of forgetting (so as to satisfy RP_F), and this is obtained by providing full insurance. Substituting these conditions in the equilibrium program, we obtain:

$$\max_{\mathbf{c}} u_A(c_1) + u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A})$$

subject to

$$\frac{c_1}{1-l} + c_2^R + \alpha c_3^{R,D} + (1-\alpha)\, c_3^{R,A} \leq \frac{I_1}{1-l} + I_2 + I_3 \tag{Zero profits}$$

$$\alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A}) \geq U(I_3) \tag{RP_R}$$

$$u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A}) \geq V \tag{RP_1}$$

$$u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A}) \geq u_A(I_2) + U(I_3) \tag{IC}$$

By the same argument as in the renegotiation program, there must be full insurance against mortality conditionally on remembering to pay:

$$u_D'(c_3^{R,D}) = u_A'(c_3^{R,A}).$$

Then, (RP_R) becomes

$$\alpha c_3^{R,D} + (1-\alpha)\, c_3^{R,A} \geq I_3.$$

Letting, $c_3^R \equiv \alpha c_3^{R,D} + (1-\alpha)\, c_3^{R,A}$, we can rewrite the program as

$$\max_{c_1, c_2^R, c_3^R} u_A(c_1) + u_A(c_2^R) + U(c_3^R) \tag{14}$$

subject to

$$\frac{c_1}{1-l} + c_2^R + c_3^R \leq \frac{I_1}{1-l} + I_2 + I_3 \tag{Zero profits}$$

$$c_3^R \geq I_3 \qquad (RP_R)$$

$$u_A(c_2^R) + U(c_3^R) \geq V \qquad (RP_1)$$

$$u_A(c_2^R) + U(c_3^R) \geq u_A(I_2) + U(I_3) \qquad \text{(IC)}$$

We claim that any allocation satisfying (RP_1) also satisfies (IC), so that (IC) can be omitted from this program. To see this, recall that V solves Program (11)-(13). But, since $\tilde{c}_2^R = I_2$ and $\tilde{c}_3^R = I_3$ satisfies the constraints (12) and (13), revealed preference gives

$$V \geq u_A(I_2) + U(I_3),$$

showing that (RP_1) is tighter than (IC).

Let $\left\{\hat{c}_1, \hat{c}_2^R, \hat{c}_3^R\right\}$ be a solution of the program above. We claim that we must have

$$\left(\hat{c}_2^R, \hat{c}_3^R\right) = \arg\max_{c_2^R, c_3^R} u_A(c_2^R) + U(c_3^R)$$

subject to

$$c_3^R \geq I_3 \qquad (RP_R)$$

and

$$c_2^R + c_3^R = \hat{c}_2^R + \hat{c}_3^R.$$

First note that this program corresponds to the maximization of a strictly concave function in a convex set, so the solution must be unique. Suppose $\left(\hat{c}_2^R, \hat{c}_3^R\right)$ is not the solution, so there exists (c_2^R, c_3^R) satisfying the constraints which has

$$u_A(c_2^R) + U(c_3^R) > u_A(\hat{c}_2^R) + U(\hat{c}_3^R).$$

But this implies that $\left\{\hat{c}_1, c_2^R, c_3^R\right\}$ is also feasible---i.e., it satisfies the (Zero Profits), (RP_R), (RP_1), and (IC)---and attains a higher objective, contradicting the optimality of $\left\{\hat{c}_1, \hat{c}_2^R, \hat{c}_3^R\right\}$.

Let the indirect utility of future consumption be:

$$W(C) \equiv \max u_A(c_2^R) + U(c_3^R)$$

subject to

$$c_3^R \geq I_3 \qquad (RP_R)$$

and

$$c_2^R + c_3^R = C.$$

Note that $V = W(I_2 + I_3)$, so RP2 can be rewritten as

$$W(C) \geq W(I_2 + I_3),$$

which, by the monotonicity of W, is equivalent to

$$C \geq I_2 + I_3. \tag{RP_1}$$

By zero profits (which must bind at the optimum), we can rewrite this constraint as

$$c_1 \leq I_1. \tag{RP_2}$$

Therefore, the equilibrium program can be rewritten as

$$\max_{c_1, C} u_A(c_1) + W(C)$$

subject to

$$\frac{c_1}{1-l} + C = \frac{I_1}{1-l} + I_2 + I_3 \tag{Zero profits}$$

$$c_1 \leq I_1. \tag{RP_1}$$

Proof of Proposition 1

The equilibrium contract solves:

$$\max_{c_1, c_2^R, c_3^R} u_A(c_1) + u_A(c_2^R) + U(c_3^R)$$

subject to

$$\frac{c_1}{1-l} + c_2^R + c_3^R \leq \frac{I_1}{1-l} + I_2 + I_3 \tag{Zero profits}$$

$$c_1 \leq I_1 \tag{RP_1}$$

$$c_3^R \geq I_3 \tag{RP_R}$$

The necessary FOCs for the solution to entail $c_1 = I_1$ are:

$$(1-l)\, u_A'(I_1) \geq u_A'(c_2^R) \geq U'(c_3^R), \tag{15}$$

with $u_A'(c_2^R) = U'(c_3^R)$ if $c_3^R > I_3$, as well as the zero profits condition (which must bind):

$$c_2^R + c_3^R = I_2 + I_3. \tag{16}$$

By zero profits, we have:

$$c_3^R \geq I_3 \iff c_2^R \leq I_2 \iff u_A'(c_2^R) \geq u_A'(I_2).$$

Substituting in (15), we find that the following inequality is necessary for the solution to entail $c_1 = I_1$:

$$(1-l)\,u'_A(I_1) \geq u'_A(I_2).$$

Proof of Claim made in Footnote 30

Let $\mathscr{W}(s) \equiv U(s + I_2 + I_3)$ denote the utility from saving s dollars (in future value) to period 3. By the auxiliary program, the following liquidity constraint must hold: $s \geq 0$. The corollary below determines how the equilibrium consumption in the first period changes with the probability of forgetting to pay l:

Corollary 1. *c_1 is weakly decreasing (increasing) in the probability of forgetting to pay l if the elasticity of $\mathscr{W}$ is greater (smaller) than 1: $-\frac{\mathscr{W}'(s)}{s\mathscr{W}''(s)} \geq (\leq) 1$.*

Proof. There are two possible cases. If (RP_1) does not bind, c_1 is implicitly determined by the following Euler equation:

$$u'_A(c_1) = \frac{1}{1-l} W'\left(\frac{I_1 - c_1}{1-l} + I_2 + (1-\alpha)\,I_3\right).$$

Using the Implicit Function Theorem, we find that $\frac{dc_1}{dl} < 0$ if and only if

$$-\frac{\left(\frac{I_1-c_1}{1-l}\right) W''\left(\frac{I_1-c_1}{1-l} + I_2 + (1-\alpha)\,I_3\right)}{W'\left(\frac{I_1-c_1}{1-l} + I_2 + (1-\alpha)\,I_3\right)} < 1.$$

If (RP_1) binds, we must have $c_1 = I_1$, which is constant in l. □

Recall that the lapse fee equals $I_1 - c_1$. Therefore, the previous corollary implies that the lapse fee is weakly increasing (decreasing) in the probability of forgetting to pay l if the elasticity of $\mathscr{W}$ is greater (smaller) than 1. Since, in equilibrium, a lapse happens whenever a consumer forgets to pay, if individuals only differ with respect to their probability of forgetting to pay, the model predicts a positive (negative) relationship between lapses and lapse fees if if the elasticity of $\mathscr{W}$ is greater (smaller) than 1.

Forgetfulness Model with Sophistication (Subsection 4.1.4)

We now formally consider the model described in Subsection 4.1.4, in which consumers have rational expectations about their likelihood of forgetting to pay the fee. As before, forgetting to pay the fee corresponds to consuming the entire income in period 2: $c_2^F = I_2$.

We start by specifying the renegotiation proofness constraints, which, as before, arise from the fact that the consumer is allowed to drop the policy at the end of periods 1 and 2.

Renegotiation Proofness Constraints at $t = 2$

At the end period 2, the individual either remembered (R) or forgot (F) to make the payment. At that point, beliefs are the same as in the model consider previously, so the renegotiation proofness constraints remain the same:

$$\alpha u_D(c_3^{R,D}) + (1-\alpha)\,u_A(c_3^{R,A}) \geq U\,(I_3) \qquad (\,RP_R)$$

$$\alpha u_D(c_3^{F,D}) + (1-\alpha)\,u_A(c_3^{F,A}) \geq U\,(I_3) \qquad (\,RP_F)$$

Renegotiation Proofness Constraints at $t = 1$

The renegotiation proofness constraints in period 1 are different from before because the consumer now has rational expectations about the probability of forgetting to pay, and therefore takes the correct probability of forgetting to pay into account. The consumer's outside option at the end of period 1 is:

$$V \equiv \max_{\mathbf{c}}\,(1-l)\left[u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\,u_A(c_3^{R,A})\right] + l\left[u_A\,(I_2) + \alpha u_D\left(c_3^{F,D}\right) + (1-\alpha)\,u_A\left(c_3^{F,A}\right)\right]$$

subject to

$$(1-l)\left[c_2^R + \alpha c_3^{R,D} + (1-\alpha)\,c_3^{R,A}\right] + l\left[I_2 + \alpha c_3^{F,D} + (1-\alpha)\,c_3^{F,A}\right] \leq I_2 + I_3 \qquad \text{(Zero profits)}$$

$$\alpha u_D(c_3^{R,D}) + (1-\alpha)\,u_A(c_3^{R,A}) \geq U\,(I_3) \qquad (\,RP_R)$$

$$\alpha u_D(c_3^{F,D}) + (1-\alpha)\,u_A(c_3^{F,A}) \geq U\,(I_3) \qquad (\,RP_F)$$

$$u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\,u_A(c_3^{R,A}) \geq u_A(I_2) + \alpha u_D(c_3^{F,D}) + (1-\alpha)\,u_A(c_3^{F,A}) \qquad \text{(IC)}$$

It is straightforward to see that the solution of this program must provide full insurance in period 3:

$$u_D'(c_3^{j,D}) = u_A'(c_3^{j,A}) \;\; j = R, F.$$

Therefore, the outside option can be rewritten as

$$V \equiv \max_{c}\,(1-l)\left[u_A(c_2^R) + U\left(C_3^R\right)\right] + l\left[u_A\,(I_2) + U\left(c_3^F\right)\right]$$

subject to

$$(1-l)\left(c_2^R + c_3^R\right) + l\left(I_2 + c_3^F\right) \leq I_2 + I_3 \qquad \text{(Zero profits)}$$

$$u_A(c_2^R) + U\left(C_3^R\right) \geq u_A(I_2) + U\left(c_3^F\right) \qquad \text{(IC)}$$

$$c_3^j \geq I_3 \;\; j = R, F \qquad (RP_j)$$

Equilibrium Program

The equilibrium program is:

$$\max_{\mathbf{c}} u_A(c_1) + (1-l)\left[u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A})\right] + l\left[u_A(I_2) + \alpha u_D\left(c_3^{F,D}\right) + (1-\alpha)\, u_A\left(c_3^{F,A}\right)\right]$$

subject to

$$c_1 + (1-l)\left[c_2^R + \alpha c_3^{R,D} + (1-\alpha)\, c_3^{R,A}\right] + l\left[I_2 + \alpha c_3^{F,D} + (1-\alpha)\, c_3^{F,A}\right] \le I_1 + I_2 + I_3 \qquad \text{(Zero profits)}$$

$$\alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A}) \ge U(I_3) \qquad (RP_R)$$

$$\alpha u_D(c_3^{F,D}) + (1-\alpha)\, u_A(c_3^{F,A}) \ge U(I_3) \qquad (RP_F)$$

$$(1-l)\left[u_A(c_2^R) + \alpha u_D(c_D^{3,R}) + (1-\alpha)\, u_A(c_A^{3,R})\right] + l\left[u_A(I_2) + \alpha u_D\left(c_3^{F,D}\right) + (1-\alpha)\, u_A\left(c_3^{F,A}\right)\right] \ge V \qquad (RP_1)$$

$$u_A(c_2^R) + \alpha u_D(c_3^{R,D}) + (1-\alpha)\, u_A(c_3^{R,A}) \ge u_A(I_2) + \alpha u_D(c_3^{F,D}) + (1-\alpha)\, u_A(c_3^{F,A}) \qquad \text{(IC)}$$

As before, it is straightforward to show that the period-1 renegotiation proofness constraints (RP_j, $j = R, F$) can be substituted by liquidity constraints:

$$\alpha c_3^{j,D} + (1-\alpha)\, c_3^{j,A} \ge I_3, \quad j = R, F.$$

Using the function U, we can rewrite the equilibrium program as

$$\max_{\mathbf{c}} u_A(c_1) + (1-l)\left[u_A(c_2^R) + U\left(c_3^R\right)\right] + l\left[u_A(I_2) + U\left(c_3^D\right)\right]$$

subject to

$$c_1 + (1-l)\left(c_2^R + c_3^R\right) + l\left(I_2 + c_3^F\right) \le I_1 + I_2 + I_3 \qquad \text{(Zero profits)}$$

$$c_3^R \ge I_3 \qquad (RP_R)$$

$$c_3^F \ge I_3 \qquad (RP_F)$$

$$u_A(c_2^R) + U\left(c_3^R\right) \ge u_A(I_2) + U\left(c_3^F\right) \qquad \text{(IC)}$$

$$(1-l)\left[u_A(c_2^R) + U\left(c_3^R\right)\right] + l\left[u_A(I_2) + U\left(c_3^F\right)\right] \ge V \qquad (RP_1)$$

Let $\mathscr{V}(C)$ denote the highest continuation utility that can be obtained at period 2 with a total expected consumption of C:

$$\mathscr{V}(C) \equiv \max_{c} (1-l)\left[u_A(c_2^R) + U\left(C_3^R\right)\right] + l\left[u_A(I_2) + U\left(c_3^F\right)\right]$$

subject to

$$(1-l)\left(c_2^R+c_3^R\right)+l\left(I_2+c_3^F\right)\leq C$$

$$c_3^j \geq I_3 \quad j=R,F$$

$$u_A(c_2^R)+U\left(c_3^R\right)\geq u_A(I_2)+U\left(c_3^F\right) \tag{IC}$$

Note that $V=\mathscr{V}(I_2+I_3)$. It is straightforward to show that the solution of the equilibrium program must solve this continuation program for $C=(1-l)\left(c_2^R+c_3^R\right)+l\left(I_2+c_3^F\right)$. Thus, (RP_1) can be written as

$$\mathscr{V}\left((1-l)\left(c_2^R+c_3^R\right)+l\left(I_2+c_3^F\right)\right)\geq\mathscr{V}\left(I_2+I_3\right),$$

which, since $\mathscr{V}$ is strictly increasing, can be further simplified to:

$$(1-l)\left(c_2^R+c_3^R\right)+l\left(I_2+c_3^F\right)\geq I_2+I_3.$$

Then, using the zero profit constraint, which must bind at the solution, it follows that RP_2 is equivalent to the following liquidity constraint:

$$c_1\leq I_1.$$

The equilibrium program can thus be written as:

$$\max_{\mathbf{c}} u_A(c_1)+(1-l)\left[u_A(c_2^R)+U\left(c_3^R\right)\right]+l\left[u_A\left(I_2\right)+U\left(c_3^F\right)\right]$$

subject to

$$c_1+(1-l)\left(c_2^R+c_3^R\right)+l\left(I_2+c_3^F\right)\leq I_1+I_2+I_3 \tag{Zero profits}$$

$$u_A(c_2^R)+U\left(c_3^R\right)\geq u_A(I_2)+U\left(c_3^F\right) \tag{IC}$$

$$c_3^R\geq I_3 \tag{RP_R}$$

$$c_3^F\geq I_3 \tag{RP_F}$$

$$c_1\leq I_1 \tag{RP_1}$$

There are two possibilities depending on whether IC binds.

Case 1: IC does not bind

If the IC does not bind, the solution must have full insurance against the risk of forgetting to make the payment $c_3^R=c_3^D$. Substituting back in the IC, we find that it holds if and only if the equilibrium contract has a negative premium at period 2, so the individual does not have an incentive to forget to "pay": $c_2^R\geq I_2$. Then, the consumer collects a payment from the insurance company in period 2, and forgetting to do so does not affect future consumption.

This is not realistic for three reasons. First, because the premium is negative, “forgetting to pay” actually corresponds to forgetting to collect a benefit. Insurance policies do not typically make payments before they expire. Second, forgetting to pay does not cause the policyholder to lapse (the consumer either always lapses or he never lapses). Third, because either everyone lapses or no one lapses, policies are not lapse-based, meaning that there are no cross-subsidies between lapsers and non-lapsers.

Case 2: IC binds

Now, consider the case where the IC binds. The necessary FOCs are:

$$u_A'(c_1) = \lambda + \mu_{RP_2}$$

$$u_A'(c_2^R)\left(1+\frac{\mu_{IC}}{1-l}\right) = \lambda$$

$$U'(c_3^R)\left(1+\frac{\mu_{IC}}{1-l}\right) + \frac{\mu_{RP_R}}{1-l} = \lambda$$

$$U'(c_3^F)\left(1-\frac{\mu_{IC}}{l}\right) + \frac{\mu_{RP_F}}{l} = \lambda$$

We claim that the solution must entail $c_3^F \le c_3^R$. Suppose $c_3^F > c_3^R$. Then, we must have $\mu_{RP_F} = 0$, so that the previous conditions give:

$$U'(c_3^F)\left(1-\frac{\mu_{IC}}{l}\right) = U'(c_3^R)\left(1+\frac{\mu_{IC}}{1-l}\right) + \frac{\mu_{RP_R}}{1-l}.$$

Rearranging, we obtain

$$U'(c_3^F) \ge U'(c_3^F)\left(1-\frac{\mu_{IC}}{l}\right) \ge U'(c_3^R)\left(1+\frac{\mu_{IC}}{1-l}\right) \ge U'(c_3^R),$$

which, by the concavity of U, gives $c_3^F \le c_3^R$, a contradiction.

Since $c_3^F \le c_3^R$, we can omit (RP_R) from the program (i.e., policies cannot lapse only for individuals who remember to pay the premium). Removing RP_R from the program, we find that there must be perfect smoothing between periods 2 and 3 for those that remember to pay:

$$u_A'(c_2^R) = U'\left(c_3^R\right). \tag{17}$$

By the binding IC constraint, we must have

$$u_A(c_2^R) + U\left(c_3^R\right) = u_A(I_2) + U\left(c_3^F\right). \tag{18}$$

Then, since $c_3^F \le c_3^R$, it must be the case that $c_2^R \le I_2$.

By 18, c_2^R and c_3^R provide the same utility as I_2 and c_3^F more efficiently---since, by 17 there is perfect smoothing between c_2^R and c_3^R. Therefore, c_2^R and c_3^R must be cheaper than I_2 and c_3^F:

$$c_2^R + c_3^R \leq I_2 + c_3^F.$$

Substituting this inequality in the (binding) zero profits condition, we obtain:

$$c_1 + I_2 + c_3^F \geq I_1 + I_2 + I_3 \geq c_1 + c_2^R + c_3^R, \tag{19}$$

and both inequalities are strict if $c_3^F < c_3^R$. In words: firms make (weakly) positive profits on consumers who remember to pay and lose money on those who forget. That is, cross-subsidies go in the opposite direction from the one observed in practice.

Suppose there are lapses in period 2. Since $c_3^F \leq c_3^R$, the policyholder must lapse after forgetting to pay:

$$c_3^F = I_3.$$

Substituting in (19), we obtain

$$c_1 \geq I_1,$$

which, by (RP_1), gives $c_1 = I_1$. Therefore, the firm makes zero profits on those who forget to pay, so (by zero profits) it must also make zero profits on those who remember to pay.

To summarize, either there are no lapses after forgetting to pay, or there is no cross-subsidy from lapsers to non-lapsers. And, when there are no lapses, cross-subsidies go in the opposite direction from the one observed in practice.

Proof of Proposition 2

The equilibrium contract maximizes

$$u_A(c_1^{NS}) + \alpha u_D\left(c_D^{NS}\right) + (1-\alpha)\, u_A\left(c_A^{NS}\right) \tag{20}$$

subject to

$$l\left[c_1^S + \alpha c_D^S + (1-\alpha)c_A^S\right] + (1-l)\left[c_1^{NS} + \alpha c_D^{NS} + (1-\alpha)c_A^{NS}\right] = W + I_1 + I_2 - lL \qquad \text{(Zero Profits)}$$

$$u_A(c_1^S) + \alpha u_D(c_D^S) + (1-\alpha)u_A(c_A^S) \geq u_A(c_1^{NS} - L) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}) \qquad (IC_S)$$

$$u_A(c_1^{NS}) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}) \geq u_A(c_1^S + L) + \alpha u_D(c_D^S) + (1-\alpha)u_A(c_A^S) \qquad (IC_{NS})$$

$$u_A(c_1^{NS}) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}) \geq \mathscr{V}(I_1 + I_2) \qquad (RP_{NS})$$

$$u_A(c_1^S) + \alpha u_D(c_D^S) + (1-\alpha)u_A(c_A^S) \geq \mathscr{V}(I_1 + I_2 - L) \qquad (RP_S)$$

Ignore (IC_{NS}) and (RP_{NS}) for the moment (we will verify that these constraints hold later). Note that

(IC_S) does not bind if the solution satisfies

$$\mathscr{V}(I_1 + I_2 - L) > u_A(c_1^{NS} - L) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}),$$

and (RP_S) does not bind if the reverse inequality holds.

Case 1. (IC_S) **does not bind.**

Since (IC_S) does not bind, the equilibrium contract maximizes (20) subject to *(Zero Profits)* and (RP_S). By the same duality argument as in the proof of Lemma 3 this program has a a unique solution, and there is full insurance conditional on the income shock:

$$u_A'(c_1^S) = u_D'(c_D^S) = u_A'(c_A^S) \text{ and } u_A'(c_1^{NS}) = u_D'(c_D^{NS}) = u_A'(c_A^{NS}).$$

Substituting in (RP_S), which must bind, we find that the total expected consumption after the shock equals the expected income after the shock:

$$c_1^S + \alpha c_D^S + (1-\alpha)c_A^S = I_1 + I_2 - L. \tag{21}$$

This means that the insurance company makes a profit of the initial income W if the individual has an income shock. By *(Zero Profits)*, the individual's expected consumption when there is no shock equals

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha)c_A^{NS} = \frac{W}{1-l} + I_1 + I_2, \tag{22}$$

so the insurance company loses $-\frac{l}{1-l}W$ if there is no income shock.

Since there is perfect smoothing conditional on both S and NS, we can use the definition of $\mathscr{V}$ to rewrite (IC_S) as:

$$u_A(c_1^{NS}) - u_A(c_1^{NS} - L) \geq \mathscr{V}\left(\frac{W}{1-l} + I_1 + I_2\right) - \mathscr{V}(I_1 + I_2 - L).$$

We now verify that the omitted constraints (RP_{NS}) and (IC_{NS}) are satisfied. Because there is perfect smoothing conditional on NS, (RP_{NS}) becomes

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha)c_A^{NS} \geq I_1 + I_2.$$

Substituting (22), verifies that this inequality holds. To verify that (IC_{NS}) holds, note that

$$\begin{aligned} u_A(c_1^{NS}) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}) &= \mathscr{V}\left(\tfrac{W}{1-l} + I_1 + I_2\right) \\ &\geq \mathscr{V}\left(\tfrac{W}{1-l} + I_1 + I_2 - L\right) \\ &\geq u_A(c_1^{NS} - L) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}) \end{aligned},$$

where the first line uses full insurance conditional on NS and (22), the second line uses the fact that $\mathscr{V}$ is increasing, and the third line uses (22) and revealed preference.

Case 2. (RP_S) **does not bind.**

Since (RP_S) does not bind, the equilibrium contract maximizes (20) subject to *(Zero Profits)* and (IC_S). Again, we can use a duality argument to rewrite the program as the minimization of a continuous and strictly convex function subject to linear constraints, establishing existence and uniqueness of the solution. Calculating the first-order conditions, we find that there is full insurance conditional on the shock:

$$u'_A(c_1^S) = u'_D(c_D^S) = u'_A(c_A^S), \tag{23}$$

and imperfect intertemporal smoothing:

$$u'_A(c_1^{NS}) > u'_A(c_A^{NS}) = u'_D(c_D^{NS}). \tag{24}$$

We now verify that the omitted constraint (IC_{NS}) is satisfied. Use the binding (IC_S) to rewrite (IC_{NS}) as

$$u_A(c_1^{NS}) - u_A(c_1^{NS} - L) \geq u_A(c_1^S + L) - u_A(c_1^S),$$

which, by the concavity of u_A, holds if and only if $c_1^S \geq c_1^{NS} - L$. That is, (IC_{NS}) holds as long as reporting a loss would increase period-1 consumption relative to absorbing the income loss. In fact, this inequality is strict in the solution.

Suppose for the sake of contradiction that $c_1^S \leq c_1^{NS} - L$. Recall that there is full insurance against mortality conditional on both S and NS:

$$u'_A(c_A^S) = u'_D(c_D^S) \text{ and } u'_A(c_A^{NS}) = u'_D(c_D^{NS}).$$

Since (IC_S) holds with equality, it then follows that $c_A^S \geq c_A^{NS}$ and $c_D^S \geq c_D^{NS}$. Then, because u_A is concave, we have

$$u'_A(c_1^{NS}) \leq u'_A(c_1^{NS} - L) \leq u'_A(c_1^S)$$

and

$$u'_A(c_A^S) \leq u'_A(c_A^{NS}).$$

Moreover, perfect smoothing conditional on the shock (23) gives

$$u'_A(c_1^{NS}) \leq u'_A(c_1^S) = u'_A(c_A^S) \leq u'_A(c_A^{NS}),$$

which contradicts (24). Thus, $c_1^S > c_1^{NS} - L$, $c_A^S < c_A^{NS}$, $c_D^S < c_D^{NS}$, and (IC_{NS}) holds.

To conclude the proof, we need to show that $\pi^S > 0 > \pi^{NS}$ and verify that (RP_{NS}) does not bind. Let

$\bar{c}_1^{NS}$, $\bar{c}_A^{NS}$, and $\bar{c}_D^{NS}$ solve

$$\max u_A(c_1^{NS}) + \alpha u_D\left(c_D^{NS}\right) + (1-\alpha)u_A\left(c_A^{NS}\right)$$

subject to

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha)c_A^{NS} \leq W + I_1 + I_2.$$

Note that the profile $\left(c_1^{NS}, c_D^{NS}, c_A^{NS}, c_1^S, c_D^S, c_A^S\right) = \left(\bar{c}_1^{NS}, \bar{c}_A^{NS}, \bar{c}_D^{NS}, \bar{c}_1^{NS} - L, \bar{c}_A^{NS}, \bar{c}_D^{NS}\right)$ is feasible in the maximization of (20) subject to *(Zero Profits)* and (IC_S). By revealed preference, it cannot attain a higher objective:

$$\begin{array}{rcl} u_A(c_1^{NS}) + \alpha u_D\left(c_D^{NS}\right) + (1-\alpha)u_A\left(c_A^{NS}\right) & \geq & u_A(\bar{c}_1^{NS}) + \alpha u_D\left(\bar{c}_D^{NS}\right) + (1-\alpha)u_A\left(\bar{c}_A^{NS}\right) \\ & = & \mathscr{V}\left(W + I_1 + I_2\right) \end{array},$$

showing that (RP_{NS}) holds. Moreover, since $\mathscr{V}\left(W + I_1 + I_2\right)$ is the highest utility that can be provided at cost $W + I_1 + I_2$, it follows that:

$$c_1^{NS} + \alpha c_D^{NS} + (1-\alpha)c_A^{NS} \geq W + I_1 + I_2. \tag{25}$$

In fact, because there is imperfect smoothing (24), this inequality must be strict. Hence, the insurance company makes negative profits if there is no shock, so, by zero expected profits, it must make positive profits if there is a shock:

$$\pi^S > 0 > \pi^{NS}.$$

Model with a Continuum of Losses (Theorem 1)

Formulation of the Program

To keep the notation close to the optimal control literature, we associate each possible *loss* with a "type" t. Types are distributed according to a differentiable PDF f with full support in the interval of possible losses $[0, T]$, where $0 < T < I_1 + I_2$. For each possible loss $t \in [0, T]$, let $c(t)$ denote the consumption in period 1, let $c_A(t)$ denote the consumption in period 2 if the individual is alive, and let $c_D(t)$ denote the bequests in period 2 (i.e., the resources repaid if the consumer dies). Let $V(t) \equiv \alpha u(c_D(t)) + (1-\alpha)u(c_A(t))$ denote the continuation payoff of a consumer who gets a loss of t.

Since firms do not observe the loss, contracts must be incentive compatible. The incentive-compatibility (IC) constraints are:

$$u(c(t)) + V(t) \geq u(c(\hat{t}) - t + \hat{t}) + V(\hat{t}) \ \forall t, \hat{t}.$$

Following standard nomenclature from mechanism design, let

$$\mathscr{U}(t) \equiv u(c(t)) + V(t)$$

denote the indirect utility of type t. The following lemma provides a standard characterization of incentive compatibility.

Lemma 4. *IC is satisfied if and only if* $\dot{\mathscr{U}}(t) = -u'(c(t))$ *and* $c(t)+t$ *is non-decreasing in* t.

Proof. Let $X \equiv c+t$. Using the taxation principle, IC can be written as

$$(X(t), V(t)) \in \arg\max_{X,V} u(X-t)+V.$$

Note that the objective function satisfies single crossing:

$$\frac{d^2}{dXdt}[u(X-t)+V] = -u''(X-t) > 0.$$

Therefore, $X(t)$ must be non-decreasing, meaning that $c(t)+t$ is non-decreasing in t. By the Envelope Theorem, $\dot{\mathscr{U}}(t) = -u'(c_1(t)) < 0$. The argument for sufficiency is standard given the validity of the single-crossing condition. □

The previous lemma shows that incentive compatibility alone has the following implications:

- The amount paid in period 1, $I_1 - c(t) - t$ is decreasing in the size of the shock: people with larger shocks pay a lower premium at period 1 if they do not lapse (that is, they borrow from their policies).
- Conversely, $\dot{V} = -u'(c(t))[1+\dot{c}(t)] \leq 0$, meaning that a lower premium at period 1 is associated with less consumption in the future. Since types with larger shocks borrow more in period 1, they have to repay what they borrowed in period 2.

We now turn to constraints imposed by the possibility of lapsing. The best contract that an individual can obtain if he lapses solves:

$$\max_{c,c_A,c_D} u(c)+\alpha u(c_D)+(1-\alpha)u(c_A)$$

subject to

$$c+\alpha c_D+(1-\alpha)c_A \leq I_1+I_2-t.$$

The solution features perfect smoothing conditional on the income shock: $c = c_A = c_D = \frac{I_1+I_2-t}{2}$. Therefore, the renegotiation proofness constraints are:

$$\mathscr{U}(t) \geq 2u\left(\frac{I_1+I_2-t}{2}\right) \forall t.$$

The zero profits constraint is

$$\int [c(t)+\alpha c_D(t)+(1-\alpha)c_A(t)+t] f(t)dt \leq W+I_1+I_2.$$

The solution of the program must entail full insurance in the second period: $c_D(t) = c_A(t) =: c_2(t)$ (otherwise, it would be possible to keep the same promised continuation utility at a lower cost). As before, it is helpful to rewrite the zero profits constraint in terms of the indirect utility $\mathscr{U}$ instead of period-2 consumption c_2:

$$\mathscr{U}(t) = u(c(t)) + u(c_2(t)) \therefore c_2(t) = u^{-1}(\mathscr{U}(t) - u(c(t))).$$

Substituting in the zero profits constraint, we obtain

$$\int \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) + t\right] f(t)dt \leq W + I_1 + I_2.$$

By the exact argument as in Lemma 2, the equilibrium must solve:

$$\max_{c,\mathscr{U}} \mathscr{U}(0)$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)), \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I_1 + I_2 - t}{2}\right), \tag{RP}$$

$$\int \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) + t\right] f(t)dt \leq W + I_1 + I_2,$$

and $c(t) + t$ non-decreasing. In what follows, we will follow the standard approach of ignoring the monotonicity constraint, which can be verified ex-post.

To simplify notation, let $I \equiv I_1 + I_2$. It is convenient to work with the dual program:

$$\min_{c,\mathscr{U}} \int_0^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) + t\right]$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)), \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I_1 + I_2 - t}{2}\right), \tag{RP}$$

$$\mathscr{U}(0) \geq \bar{u}, \tag{26}$$

where $\bar{u} > 2u\left(\frac{I}{2}\right)$ so the feasible set is non-empty.[49]

This is an optimal control problem, where $\mathscr{U}$ is a state variable and c is a control variable. Equation (26), which must hold as an equality in any solution, gives the initial condition for $\mathscr{U}$. The terminal condition $\mathscr{U}(T)$ is free. Equation (IC) is a standard constraint, which gives the law of motion for the

[49]To see that $\bar{u} > 2u\left(\frac{I}{2}\right)$ implies that the set of functions $(c, \mathscr{U})$ that satisfy the constraints is non-empty, note that $\mathscr{U}(t) = 2u\left(u^{-1}\left(\frac{\bar{u}}{2}\right) - \frac{t}{2}\right)$ and $c(t) = u^{-1}\left(\frac{\bar{u}}{2}\right) - \frac{t}{2}$ satisfies all the constraints.

state variable as a function of the control variable c.

Equation (RP) is a first-order pure state inequality constraint — see Hartl, Sethi, and Vickson (1995, Section 5) and Grass et al. (2008, Section 3.6). It is a pure state inequality constraint because it involves the state variable $\mathscr{U}$ and "time" t but not the control variable c. To see why it has order one, rewrite it in its canonical form

$$h(\mathscr{U},t) \equiv \mathscr{U} - 2u\left(\frac{I-t}{2}\right),$$

so (RP) becomes

$$h(\mathscr{U}(t),t) \geq 0.$$

Total differentiation, gives:

$$\frac{d}{dt}\left[h(\mathscr{U}(t),t)\right] = \dot{\mathscr{U}}(t) + u'\left(\frac{I-t}{2}\right) = u'\left(\frac{I-t}{2}\right) - u'(c(t)),$$

where the last equality used the law of motion (IC). Therefore, differentiating h once allows us to express h as a constraint involving the control variable.

Notation

Let $V(\bar{u}_a, a)$ denote the "continuation cost" at time a with initial state $\bar{u}_a$:

$$V(\bar{u}_a, a) := \min_{c,\mathscr{U}} \int_a^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t)))\right] f(t)dt$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)) \ \forall t, \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right) \ \forall t, \tag{RP}$$

$$\mathscr{U}(a) = \bar{u}_a,$$

where $\bar{u}_a \geq 2u\left(\frac{I-a}{2}\right)$ so the feasible set is non-empty. It is straightforward to verify that $V(\cdot, a)$ is a non-decreasing function of $\bar{u}_a$.

Below, we will use the fact that, by the Principle of Optimality, if $(\mathscr{U}, c)$ solves the original program then, for any $t_L, t_H \in (0,T)$ with $0 \leq t_L \leq t_H < T$, the restriction of $(\mathscr{U}, c)$ to $[t_L, t_H]$ must solve:

$$\min_{\tilde{c},\tilde{\mathscr{U}}} \int_{t_L}^{t_H} \left[\tilde{c}(t) + u^{-1}\left(\tilde{\mathscr{U}}(t) - u(\tilde{c}(t))\right)\right] f(t)dt + V(\tilde{\mathscr{U}}(t_H), t_H)$$

subject to

$$\dot{\tilde{\mathscr{U}}}(t) = -u'(\tilde{c}(t)) \ \forall t, \tag{IC}$$

$$\tilde{\mathscr{U}}(t) \geq 2u\left(\frac{I-t}{2}\right)\ \forall t, \tag{RP}$$

$$\tilde{\mathscr{U}}(t_L) = \mathscr{U}(t_L). \tag{B}$$

As usual, for a given $\mathscr{U}$, we say that RP *binds* at t if (RP) holds with equality at t. It is helpful to introduce some standard notation from optimal control:

Definition 1. The point τ is called an *entry point* if there exists $\varepsilon > 0$ such that RP binds for $t \in (\tau, \tau+\varepsilon)$ but not for $t \in (\tau-\varepsilon, \tau)$; τ is called an *exit point* if there exists $\varepsilon > 0$ such that RP binds for $t \in (\tau-\varepsilon, \tau)$ but not for $t \in (\tau, \tau+\varepsilon)$; and τ is called a *contact point* if there exists $\varepsilon > 0$ such that RP binds for $t = \tau$ but not for $t \in (\tau-\varepsilon, \tau+\varepsilon)\backslash\{\tau\}$. A point is called a *junction point* if it either an entry point, an exit point, or a contact point.

Let $S \subset [0,T]$ denote the subset of points where RP binds. A contact point is an isolated point of S. Therefore, if all junction points in $[0,T]$ are contact points, then S must be a countable set, which has Lesbegue measure zero (i.e., RP does not bind for almost all t).

In the proofs below, we will use the following result:

Lemma 5. *Let $\phi(\cdot)$ be a Lipschitz continuous function. Suppose $f'(t) < \phi(f(t))$, $g'(t) \geq \phi(g(t))$, and $f(a) = g(a) = \alpha$. Then, $f(t) < g(t)$ for all $t > a$.*

Proof. Suppose not. Then, there is $b > a$ such that $f(b) \geq g(b)$. Let

$$c \equiv \inf\{x > a:\ f(x) \geq g(x)\}.$$

If $c > a$, we must have $f'(c) \geq g'(c)$, which contradicts the fact that

$$f'(c) < \phi(f(c)) = \phi(g(c)) \leq g'(c).$$

If $c = a$, then we must have $f'(a) = g'(a)$, contradicting the fact that

$$f'(a) < \phi(f(a)) = \phi(\alpha) = \phi(g(a)) \leq g'(a).$$

□

Results

As a *benchmark*, consider a version of the original program without the IC constraint:

$$\min_{c,\mathscr{U}} \int_0^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t)))\right] f(t)dt$$

subject to

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right)\ \forall t, \tag{RP}$$

$$\mathscr{U}(0) = \bar{u}. \tag{B}$$

The solution can be obtained by minimizing the objective pointwise, which gives:

$$1 - \frac{u'(c(t))}{u'(u^{-1}(\mathscr{U}(t) - u(c(t))))} = 0 \therefore \mathscr{U}(t) = 2u(c(t))$$

for almost all t.

We are now ready to present the proof of Theorem 1 using a series of lemmas. The first lemma shows that introducing the IC constraint can only distort the control $c(t)$ downwards relative to the benchmark:

Lemma 6. *Suppose $(\mathscr{U}, c)$ solves the program. Then, $2u(c(t)) \leq \mathscr{U}(t)$ for all t.*

The proof will be given in four separate claims.

Claim 1. Let $(\mathscr{U}, c)$ satisfy IC and RP and suppose $2u(c(t)) > \mathscr{U}(t)$ for $t \in (t_L, t_H)$. Then, RP does not bind for any $t \in (t_L, t_H)$.

Proof. Suppose, for the sake of contradiction, that RP binds for some $t \in (t_L, t_H)$. Let

$$t_- = \inf\{t \in (t_L, t_H) : \mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right) = 0\}.$$

There are two cases: $t_- > t_L$ and $t_- = t_L$. If $t_- > t_L$, then we must have

$$\frac{d}{dt}\left[\mathscr{U}(t_-) - 2u\left(\frac{I-t_-}{2}\right)\right] \leq 0.$$

Use IC to rewrite this condition as

$$\dot{\mathscr{U}}(t_-) + u'\left(\frac{I-t_-}{2}\right) \leq 0 \therefore \frac{I-t_-}{2} \geq c(t_-). \tag{27}$$

Therefore, we have

$$\mathscr{U}(t_-) < 2u(c(t_-)) \leq 2u\left(\frac{I-t_L}{2}\right),$$

where the first inequality follows from $t_- > t_L$ and the second condition follows from 27. But this contradicts the hypothesis that $(\mathscr{U}, c)$ satisfies RP.

If $t_- = t_L^*$, then there must exist $\varepsilon > 0$ such that, for all $t \in (t_L^*, t_L^* + \varepsilon)$,

$$\mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right) = 0.$$

Differentiate this condition and use IC to write

$$\frac{d}{dt}\left[\mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right)\right] = \dot{\mathscr{U}}(t) + u'\left(\frac{I-t}{2}\right) = -u'(c(t)) + u'\left(\frac{I-t}{2}\right) = 0$$

$$\therefore c(t) = \frac{I-t}{2}$$

for $t \in (t_L^*, t_L^* + \varepsilon)$, so that $2u(c^*(t)) = \mathscr{U}^*(t)$ in this interval, a contradiction, since $2u(c^*(t)) > \mathscr{U}^*(t)$ for all $t \in (t_L^*, t_H^*)$. ☐

Claim 1 shows that t_H cannot be an exit point. There are three possibilities: (i) RP does not bind at t_H, (ii) t_H is an entry point, or (iii) t_H is a contact point. We address each of them separately:

Claim 2. Let $(\mathscr{U}, c)$ satisfy IC and RP and suppose $2u(c(t)) > \mathscr{U}(t)$ for $t \in (t_L, t_H)$. If RP does not bind at t_H, then $(\mathscr{U}, c)$ is not optimal.

Proof. Let $(\mathscr{U}, c)$ satisfy IC and RP. Suppose $2u(c(t)) > \mathscr{U}(t)$ for $t \in (t_L, t_H)$ and $\mathscr{U}(t_H) > 2u\left(\frac{I-t_H}{2}\right)$. Then, the restriction of $(\mathscr{U}, c)$ to $[t_L, t_H]$ must solve the following continuation program:

$$\min_{c,\mathscr{U}} \int_{t_L}^{t_H} \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t)))\right] f(t)dt + V(\mathscr{U}(t_H), t_H)$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)) \; \forall t, \tag{IC}$$

$$\mathscr{U}(t_L) = \bar{u}_L,$$

where $\mathscr{U}^*(t_L) = \bar{u}_L$. The optimality conditions are:

$$\lambda(t) = -\left[1 - \frac{u'(c(t))}{u'(u^{-1}(\mathscr{U}(t) - u(c(t))))}\right] \frac{f(t)}{u''(c(t))}, \tag{28}$$

$$\dot{\lambda}(t) = \frac{f(t)}{u'(u^{-1}(\mathscr{U}(t) - u(c(t))))} > 0, \tag{29}$$

and the transversality condition

$$\lambda(t_H) = -\frac{\partial V}{\partial \mathscr{U}}(\mathscr{U}(t_H), t_H) \leq 0.$$

Transversality and (29) imply $\lambda(t) < 0$ for $t < t_H$, which, by (28), implies

$$\mathscr{U}(t) > 2u(c(t)),$$

contradicting the fact that $2u(c(t)) > \mathscr{U}(t)$ for all $t \in (t_L, t_H)$. ☐

Claim 3. Let $(\mathscr{U}, c)$ satisfy IC and RP and suppose $2u(c(t)) > \mathscr{U}(t)$ for $t \in (t_L, t_H)$. If t_H is an entry point, then $(\mathscr{U}, c)$ is not optimal.

Proof. Let $(\mathscr{U}^*, c^*)$ be a solution. Suppose first that RP binds for all $t \in [t_H^*, L)$, that is $2u(c^*(t)) = \mathscr{U}^*(t)$

for (t_H^*, T). Construct $(\mathscr{U}, c)$ as follows. Let $\mathscr{U}(t) = \mathscr{U}^*(t)$ and $c(t) = c^*(t)$ for $t < t_L^*$. For $t \geq t_L^*$, let

$$\dot{\mathscr{U}}(t) = -u'\left(u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)\right),$$

with $\mathscr{U}(t_L^*) = \mathscr{U}^*(t_L^*)$ and let $c(t) = u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)$. Note that $c(t) = u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)$ and $c^*(t) > u^{-1}\left(\frac{\mathscr{U}^*(t)}{2}\right)$. Then, since $-u'(c)$ is a strictly increasing function of c, IC and the initial conditions give:

$$\dot{\mathscr{U}}(t) < -u'\left(u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)\right),$$

$$\dot{\mathscr{U}}^*(t) > -u'\left(u^{-1}\left(\frac{\mathscr{U}^*(t)}{2}\right)\right),$$

and

$$\mathscr{U}(t_L^*) = \mathscr{U}^*(t_L^*).$$

It then follows from Lemma 5 that $\mathscr{U}(t) < \mathscr{U}^*(t)$ for all $t \in (t_L^*, t_H^*]$. Moreover, for $t \geq t_H^*$, we have

$$\dot{\mathscr{U}}(t) = -u'\left(u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)\right) \text{ and } \dot{\mathscr{U}}^*(t) = -u'\left(u^{-1}\left(\frac{\mathscr{U}^*(t)}{2}\right)\right),$$

with

$$\mathscr{U}(t_H^*) < \mathscr{U}^*(t_H^*),$$

which implies $\mathscr{U}(t) < \mathscr{U}^*(t)$ for all $t > t_H^*$. Therefore, we have shown that $\mathscr{U}(t) < \mathscr{U}^*(t)$ for all $t > t_L^*$ and $2u(c(t)) = \mathscr{U}(t)$. But this implies that

$$c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) = u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right) < u^{-1}\left(\frac{\mathscr{U}^*(t)}{2}\right) \leq c^*(t) + u^{-1}(\mathscr{U}^*(t) - u(c^*(t)))$$

for all $t > t_L^*$, contradicting the optimality of $(\mathscr{U}^*, c^*)$.

Next, suppose that RP binds for $t \in [t_H^*, \tau)$ but RP does not bind on $\tau < L$. Construct the restriction of $(\mathscr{U}, c)$ to $[0, \tau)$ as in the previous case, and let the restriction of $(\mathscr{U}, c)$ to $[\tau, L]$ be given by the solution to the continuation program

$$\min_{c,\mathscr{U}} \int_\tau^L \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t)))\right] f(t)dt$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)) \; \forall t, \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right) \; \forall t, \tag{RP}$$

$$\mathscr{U}(\tau) = \mathscr{U}(\tau^-), \tag{B}$$

where $\mathscr{U}(\tau^-) \equiv \lim_{t \nearrow \tau} \mathscr{U}(t)$. By the same argument as in the previous case, we find that $(\mathscr{U}, c)$ yields the same cost for all $t < t_L^*$, has a strictly lower cost for all points in (t_L^*, τ). Moreover, $\mathscr{U}(\tau) < \mathscr{U}^*(\tau)$, which implies that the cost that solves the continuation program above is weakly lower than the cost under $(\mathscr{U}^*, c^*)$. Therefore, the total cost under $(\mathscr{U}, c)$ is strictly lower than under $(\mathscr{U}^*, c^*)$, contradicting the optimality of $(\mathscr{U}^*, c^*)$. □

Claim 4. Let $(\mathscr{U}, c)$ satisfy IC and RP and suppose $2u(c(t)) > \mathscr{U}(t)$ for $t \in (t_L, t_H)$. If t_H is a contact point, then $(\mathscr{U}, c)$ is not optimal.

Proof. Since τ is a contact point, we must have

$$\lim_{t \nearrow \tau} \frac{d}{dt}\left[\mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right)\right] \leq 0.$$

Use IC to rewrite this as

$$c(\tau^-) \leq \frac{I-\tau}{2}. \tag{30}$$

Therefore,

$$\mathscr{U}(\tau) < 2u\left(c\left(\tau^-\right)\right) \leq 2u\left(\frac{I-\tau}{2}\right) = \mathscr{U}(\tau),$$

where the first inequality follows from the fact that $\mathscr{U}(t) < 2u(c(t))$ for all $t < \tau$ and $\mathscr{U}$ is continuous, the second inequality uses (30), and the equality at the end uses the fact that τ is a contact point. But this is a contradiction. □

We follow the indirect adjoining approach and refer the reader to Hartl, Sethi, and Vickson (1995) for a description of the method. The Hamiltonian and Lagrangian functions are defined as follows:

$$H(c, \mathscr{U}, \lambda, t) = -\left[c + u^{-1}(\mathscr{U} - u(c)) + t\right] f(t) - \lambda u'(c)$$

$$L(c, \mathscr{U}, \lambda, \eta, t) = H(c, \mathscr{U}, \lambda, t) + \nu\left[u'\left(\frac{I-t}{2}\right) - u'(c)\right].$$

The necessary optimality conditions are:[50]

$$\lambda(t) + \nu(t) = -\left[1 - \frac{u'(c(t))}{u'(u^{-1}(\mathscr{U}(t) - u(c(t))))}\right] \frac{f(t)}{u''(c(t))}, \tag{31}$$

$$\dot{\lambda}(t) = \frac{f(t)}{u'(u^{-1}(\mathscr{U}(t) - u(c(t))))}, \tag{32}$$

$$\nu(t) \geq 0 \text{ with } = \text{if either } c(t) > \frac{I-t}{2} \text{ or } \mathscr{U}(t) > 2u\left(\frac{I-t}{2}\right), \tag{33}$$

$$\dot{\nu}(t) \leq 0, \tag{34}$$

[50] In terms of the conditions in Grass et al. (2008), it is easy to verify that the problem is normal (i.e., $\lambda_0 = 1$ using their notation).

and

$$\lambda(T^-) \geq 0 \text{ with } = \text{ if } \mathscr{U}(T^-) > 2u\left(\frac{I-T}{2}\right). \tag{35}$$

If τ is an entry or contact point, then

$$\lambda(\tau^-) = \lambda(\tau^+) + \eta(\tau), \tag{36}$$

$$\begin{aligned} &\left[c(\tau^-) + u^{-1}\left(\mathscr{U}(\tau) - u\left(c(\tau^-)\right)\right)\right] f(t) + \lambda(\tau^-)u'(c(\tau^-)) = \\ &= \left[c(\tau^+) + u^{-1}\left(\mathscr{U}(\tau) - u\left(c(\tau^+)\right)\right)\right] f(t) + \lambda(\tau^+)u'(c(\tau^+)) + \eta(\tau)u'\left(\tfrac{I-t}{2}\right) \end{aligned}, \tag{37}$$

and

$$\eta(\tau) \geq 0 \text{ with } = \text{ if } \mathscr{U}(t) > 2u\left(\frac{I-t}{2}\right). \tag{38}$$

Moreover, at any entry point τ_1,

$$\eta(\tau_1) \geq \nu(\tau_1^+), \tag{39}$$

and $\lambda(\tau_2)$ is continuous at any exit point τ_2.

We first show that there is "no distortion at the top":

Lemma 7. *$\mathscr{U}(T) = 2u\left(\frac{I-T}{2}\right)$ and $\lambda(T^-) = \nu(T^-) = 0$.*

Proof. From (35), there are two possible cases. If RP does not bind in a neighborhood of T, then $\lambda(T^-) = \nu(T^-) = 0$ and the result follows from (31). If RP binds in a neighborhood of T, then

$$\mathscr{U}(T) = 2u\left(\frac{I-T}{2}\right).$$

Differentiate of RP and use IC to establish that $c(T^-) = \frac{I-T}{2}$. Substituting in (31), gives

$$\lambda(T^-) + \nu(T^-) = 0,$$

and, since $\lambda(T^-) \geq 0$ (35) and $\nu(t) \geq 0$ for all t (33), we find $\lambda(T^-) = \nu(T^-) = 0$. □

Lemma 8. *Let τ_2 be an exit point. Then $c(\cdot)$ is continuous at τ_2.*

Proof. Let τ_2 be an exit point, so that RP binds in $(\tau_2 - \varepsilon, \tau_2]$ and does not bind in $(\tau_2, \tau_2 + \varepsilon)$. Differentiate RP at $t \in (\tau_2 - \varepsilon, \tau_2]$ and use IC to find that $2u(c(t)) = \mathscr{U}(t)$ for all $t \in (\tau_2 - \varepsilon, \tau_2)$. In particular, $2u\left(c\left(\tau_2^-\right)\right) = \mathscr{U}(\tau_2)$.

Recall that λ must be continuous at any exit point. Since RP does not bind at $t > \tau_2$ but binds at $t < \tau_2$, condition (33) gives:

$$\nu(\tau_2^-) \geq 0 = \nu(\tau_2^+).$$

From (31), we have

$$\lambda(\tau_2)+\nu(\tau_2^-) = -\left[1-\frac{u'\left(c\left(\tau_2^-\right)\right)}{u'\left(u^{-1}\left(\mathscr{U}\left(\tau_2\right)-u\left(c\left(\tau_2^-\right)\right)\right)\right)}\right]\frac{f(\tau_2)}{u''\left(c\left(\tau_2^-\right)\right)} = 0,$$

so $\lambda(\tau_2) \leq 0$, and

$$\lambda(\tau_2) = -\left[1-\frac{u'\left(c\left(\tau_2^+\right)\right)}{u'\left(u^{-1}\left(\mathscr{U}\left(\tau_2\right)-u\left(c\left(\tau_2^+\right)\right)\right)\right)}\right]\frac{f(\tau_2)}{u''\left(c\left(\tau_2^+\right)\right)}.$$

Combining both, we find that

$$-\left[1-\frac{u'\left(c\left(\tau_2^+\right)\right)}{u'\left(u^{-1}\left(\mathscr{U}\left(\tau_2\right)-u\left(c\left(\tau_2^+\right)\right)\right)\right)}\right]\frac{f(\tau_2)}{u''\left(c\left(\tau_2^+\right)\right)} \leq 0.$$

But this condition is equivalent to $\mathscr{U}\left(\tau_2\right) \geq 2u\left(c\left(\tau_2^+\right)\right)$. Substituting $\mathscr{U}\left(\tau_2\right) = 2u\left(\frac{I-\tau_2}{2}\right)$ and using the monotonicity of u, we obtain

$$\frac{I-\tau_2}{2} \geq c\left(\tau_2^+\right).$$

Therefore, if $c(\cdot)$ is discontinuous at the exit point τ_2, it must jump downwards. But jumping downwards is not possible at an exit point since, to make RP no longer bind at a neighborhood to the right of τ_2, we must have

$$0 \leq \frac{d}{dt}\left[\mathscr{U}(t)-2u\left(\frac{I-t}{2}\right)\right]_{t\searrow\tau_2} = -u'(c(\tau_2^+))+u'\left(\frac{I-\tau_2}{2}\right),$$

$$\therefore \frac{I-\tau_2}{2} \leq c(\tau_2^+).$$

So $c(\cdot)$ must be continuous at any exit point. $\square$

We now show that there are no exit points, so that if τ is an entry point, RP must bind for all $t > \tau$ (i.e., there can be at most one entry point).

Lemma 9. *There are no exit points.*

Proof. Suppose τ_2 is an exit point, so RP binds in $(\tau_2-\varepsilon,\tau_2]$ but not in $(\tau_2,\tau_2+\varepsilon)$. From condition (31), for $t \in (\tau_2,\tau_2+\varepsilon)$, we must have

$$\lambda(t) = -\left[1-\frac{u'\left(c\left(t\right)\right)}{u'\left(u^{-1}\left(\mathscr{U}\left(t\right)-u\left(c\left(t\right)\right)\right)\right)}\right]\frac{f(t)}{u''(c)}.$$

Since $c(t)$ is continuous at τ_2 (by the previous lemma) and $\mathscr{U}(\tau_2) = 2u\left(\frac{I-\tau_2}{2}\right)$ (since τ_2 is a junction point), this condition yields $\lambda\left(\tau_2^+\right) = 0$. Then, because $\dot{\lambda}(t) > 0$ for all $t \in (\tau_2,\tau_2+\varepsilon)$ by condition (32), it follows that $\lambda(t) > 0$ for $t \in (\tau_2, \tau_2 + \varepsilon)$, implying that

$$2u(c(t)) > \mathscr{U}(t)$$

in this interval, a contradiction by Lemma 6. □

Lemma 10. *Let τ be an entry point. Then, $c(\cdot)$ is continuous at τ.*

Proof. From the previous lemma, there are no exit points. Therefore, if τ is an entry point, RP must bind for all $t \in [\tau, T]$. From (35), $\lambda(T^-) = 0$. Moreover, from (32), $\dot{\lambda}(t) = \frac{f(t)}{u'\left(\frac{I-t}{2}\right)}$ for all $t > \tau$. Therefore, we must have

$$\lambda(\tau^+) < 0. \tag{40}$$

As shown in Grass et al. (2008, pp. 151), the following condition must hold at any junction point, so it must also hold at the entry point τ:

$$\eta(\tau)\left[u'\left(\frac{I-\tau}{2}\right) - u'\left(c\left(\tau^-\right)\right)\right] = 0. \tag{41}$$

By (39), we must have

$$\eta(\tau) \geq \nu(\tau^+). \tag{42}$$

As in the proof of Lemma 8, differentiate RP and use IC to find that $2u(c(t)) = \mathscr{U}(t)$ for $t > \tau$. Then, by (31), we must have

$$\lambda(t) + \nu(t) = 0 \therefore \nu(t) = -\lambda(t)$$

for all $t > \tau$.

Taking the limit as $t \searrow \tau$ and using the fact that $\lambda(\tau^+) < 0$ (equation 40), we obtain:

$$\nu(\tau^+) = -\lambda(\tau^+) > 0. \tag{43}$$

Substituting in (42), we find that $\eta(\tau) \geq \nu(\tau^+) > 0$. Therefore, (41) implies that

$$u'\left(\frac{I-\tau}{2}\right) - u'\left(c\left(\tau^-\right)\right) = 0 \therefore c\left(\tau^-\right) = \frac{I-\tau}{2},$$

showing that c is continuous at τ. □

Lemma 11. *Let τ be an entry point. Then, $2u(c(t)) < \mathscr{U}(t)$ for all $t < \tau$ and $2u(c(t)) = \mathscr{U}(t)$ for $t > \tau$.*

The proof will use the following claim:

Claim 5. Let $\hat{\tau}$ be a contact point. Then, $\lambda(\cdot)$ is continuous at $\hat{\tau}$ and $c(\cdot)$ jumps upwards at $\hat{\tau}$ ($c(\hat{\tau}^-) < c(\hat{\tau}^+)$).

Proof. This proof uses results Grass et al. (2008, page 151). Let $\hat{\tau}$ be a contact point, so that

$$\lim_{t\nearrow\tau}\frac{d}{dt}\left[\mathscr{U}(t)-2u\left(\frac{I-t}{2}\right)\right]\leq 0\leq \lim_{t\searrow\tau}\frac{d}{dt}\left[\mathscr{U}(t)-2u\left(\frac{I-t}{2}\right)\right].$$

Use IC to rewrite this as

$$c(\tau^-)\leq\frac{I-\tau}{2}\leq c(\tau^+). \tag{44}$$

At any contact time, we must have

$$\lambda(\hat{\tau}^-)=\lambda(\hat{\tau}^+)+\eta(\hat{\tau}),$$

where $\eta(\hat{\tau})\geq 0$ and

$$\eta(\hat{\tau})\left[u'\left(\frac{I-\hat{\tau}}{2}\right)-u'\left(c\left(\hat{\tau}^-\right)\right)\right]=0.$$

Suppose, in order to obtain a contradiction, that $\eta(\hat{\tau})>0$, so that $c(\hat{\tau}^-)=\frac{I-\hat{\tau}}{2}$. Since the Hamiltonian evaluated at contact times must be continuous, we must have:

$$\left\{\left[c(\tau^+)+u^{-1}\left(\mathscr{U}-u\left(c(\tau^+)\right)\right)+t\right]-(I-\hat{\tau}+t)\right\}f(t)=\lambda^+\left[u'\left(\frac{I-\hat{\tau}}{2}\right)-u'\left(c(\tau^+)\right)\right]+\eta(\hat{\tau})u'\left(\frac{I-\hat{\tau}}{2}\right).$$

Note that if $c(\cdot)$ were continuous at τ, so that $c(\tau^+)=c(\tau^-)=\frac{I-\hat{\tau}}{2}$, this condition would become:

$$\eta(\hat{\tau})u'\left(\frac{I-\hat{\tau}}{2}\right)=0,$$

contradicting our assumption that $\eta(\hat{\tau})>0$. Thus, $c(\cdot)$ must be discontinuous at $\hat{\tau}$, which, by (44), requires

$$c(\tau^-)=\frac{I-\tau}{2}<c(\tau^+)$$

$$\therefore\mathscr{U}(\hat{\tau})=2u\left(\frac{I-\hat{\tau}}{2}\right)<2u\left(c\left(\tau^+\right)\right),$$

where the second line uses the fact that $\hat{\tau}$ is a contact point. But this contradicts Lemma 6. Thus, the solution entails $\eta(\hat{\tau})=0$, so $\lambda(\cdot)$ is continuous at $\hat{\tau}$. □

We are now ready to present the proof of the lemma:

Proof of Lemma 11. Let τ be an entry point. Because RP binds at $t>\tau$, we have:

$$\mathscr{U}(\tau)=2u\left(\frac{I-\tau}{2}\right)$$

and

$$\dot{\mathscr{U}}(\tau)=-u'(c(\tau^+))=-u'\left(\frac{I-\tau}{2}\right)\therefore c(\tau^+)=\frac{I-\tau}{2}.$$

Then, by the continuity of $c(\cdot)$ at τ (Lemma 10), we have $c(\tau^-)=\frac{I-\tau}{2}$. By Lemma 9, any $t<\tau$ is either a point in which RP does not bind or a contact point, so $\nu(t)=0$ for $t\in(\tau-\varepsilon,\tau)$. It then follows from (31) that

$$\lambda(\tau^-)=-\left[1-\frac{u'\left(\frac{I-\tau}{2}\right)}{u'\left(u^{-1}\left(2u\left(\frac{I-\tau}{2}\right)-u\left(\frac{I-\tau}{2}\right)\right)\right)}\right]\frac{f(\tau)}{u''\left(\frac{I-\tau}{2}\right)}=0.$$

By our previous claim, $\lambda(\cdot)$ is continuous at $[0,\tau)$. Since $\dot{\lambda}(t)>0$ at all points of differentiability of $\lambda(\cdot)$ and $\lambda(\tau^-)=0$, it follows that $\lambda(t)<0$ for all $t<\tau$. But this implies that $2u(c(t))<\mathscr{U}(t)$ for all $t<\tau$. □

Interpretation of Results

The equilibrium must be such that insurance is lapse based. That is, if the lapsing region is non-empty $(\tau<T)$, insurance companies must make a profit of W when individuals lapse and lose $\frac{W}{\int_\tau^T f(t)dt}$ when individuals do not lapse.

The fact that insurance is lapse based is not surprising since consumers put zero weight on the chance they will get any shock $t>0$, so they do not think they will ever lapse. The distortion effect is more interesting. If there were no ICs (i.e., if shocks were observable), the most profitable way to exploit different beliefs would be to extract the entire surplus when there is a loss while always providing full insurance. With unobservable losses, the contract must ensure that consumers report losses truthfully. Then, distorting their consumption profile by shifting consumption to the future reduces their incentives to misreport losses. This is equivalent to offering policy loans at above market rates. So, the high interest rates on policy loans are a consequence of the unobservability of income shocks.

We conclude by showing that profits increase with types. Let π denote the firm's profits:

$$\pi(t)\equiv W+I_1+I_2-t-c(t)-u^{-1}\left(\mathscr{U}(t)-u(c(t))\right).$$

Differentiation gives, at all points of differentiability,

$$\dot{\pi}(t)=-[1+\dot{c}(t)]-\frac{\dot{\mathscr{U}}(t)-u'(c(t))\,\dot{c}(t)}{u'\left(u^{-1}\left(\mathscr{U}(t)-u(c(t))\right)\right)}.$$

Using IC, we obtain

$$\begin{aligned}\dot{\pi}(t)&=-[1+\dot{c}(t)]+\frac{u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t)-u(c(t))\right)\right)}[1+\dot{c}(t)]\\&=-[1+\dot{c}(t)]\left[1-\frac{u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t)-u(c(t))\right)\right)}\right].\end{aligned}$$

By monotonicity, $1+\dot{c}(t)\geq 0$. Moreover,

$$1-\frac{u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t)-u(c(t))\right)\right)}\leq 0\iff\mathscr{U}(t)\geq 2u(c(t)).$$

Then, using Lemma 6, we find that $\dot{\pi}(t) \geq 0$. Note also that $\pi(t) = W$ for all $t > \tau$.

Log Utility and Uniform Distribution (Proposition 3)

Let $u(c) = \ln c$, and suppose losses are uniformly distributed in $[0, \bar{L}]$, so that $f(t) = \frac{1}{\bar{L}}$. Let τ denote the entry point. Consider the relaxed program in which we ignore RP for $t < \tau$, so we can ignore contact points for now. We will verify that RP does not bind in this interval later.

The optimality conditions (31) and (32) become:

$$\lambda(t) = \frac{c(t)^2 - \exp \mathscr{U}(t)}{\bar{L}} \text{ for all } t < \tau \tag{45}$$

where we used the fact that $\nu(t) = 0$ if (RP) does not bind, and

$$\dot{\lambda}(t) = \frac{1}{\bar{L}} \frac{\exp \mathscr{U}(t)}{c(t)} \text{ for all } t. \tag{46}$$

Differentiate (45) to obtain:

$$\dot{\lambda}(t) = \frac{2c(t) \cdot \dot{c}(t) - \exp \mathscr{U}(t) \cdot \dot{\mathscr{U}}(t)}{\bar{L}}.$$

Recall that, from IC, we have $\dot{\mathscr{U}}(t) = -\frac{1}{c(t)}$. Therefore, the previous condition becomes:

$$\dot{\lambda}(t) = \frac{2c(t) \cdot \dot{c}(t)}{\bar{L}} + \frac{1}{\bar{L}} \frac{\exp \mathscr{U}(t)}{c(t)}.$$

Substituting in (46), gives $\dot{c}(t) = 0$. Therefore, $c(t) = \bar{c}$ constant for $t < \tau$. Moreover, since $c(\cdot)$ is continuous at the entry point τ (Lemma 10) and $c(t) = \frac{I-t}{2}$ for all $t > \tau$, we must have $\bar{c} = \frac{I-\tau}{2}$.

We now verify that RP does not bind at $t < \tau$, so there are no contact points. Using the logarithmic utility, (RP) becomes:

$$\mathscr{U}(t) - 2 \ln \left(\frac{I-t}{2} \right) \geq 0.$$

Since, by construction, this inequality binds at the entry point τ, to show that it does not bind at any $t < \tau$, it suffices to verify that

$$\frac{d}{dt} \{ \mathscr{U}(t) - 2 [\ln(I-t) - \ln 2] \} \leq 0.$$

Differentiate and use IC to rewrite this inequality as:

$$-\frac{1}{\bar{c}} \leq -\frac{2}{I-t} \iff \frac{I-t}{2} \geq \bar{c}.$$

Using $\bar{c} = \frac{I-\tau}{2}$, this condition becomes:

$$\frac{I-t}{2} \geq \frac{I-\tau}{2} \iff t \leq \tau.$$

Therefore, RP does not bind at any $t < \tau$.

We now obtain the expressions for the equilibrium consumption. As shown previously, first-period consumption equals

$$c_1(t) = \begin{cases} \frac{I-\tau}{2} & \text{if } t < \tau \\ \frac{I-t}{2} & \text{if } t \geq \tau \end{cases}.$$

Since $\mathscr{U}(0) = \bar{u}$ and, by IC, $\dot{\mathscr{U}}(t) = -\frac{1}{c(t)}$, we can recover the indirect utility:

$$\mathscr{U}(t) = \bar{u} - \int_0^t \frac{1}{c(s)} ds = \begin{cases} \bar{u} - \frac{2t}{I-\tau} & \text{for } t \leq \tau \\ \bar{u} - \frac{2\tau}{I-\tau} + 2\ln\left(\frac{I-t}{I-\tau}\right) & \text{for } t > \tau \end{cases}.$$

Since RP binds at all $t \geq \tau$, we must have

$$\mathscr{U}(t) = 2\ln\left(\frac{I-t}{2}\right) \text{ for all } t \geq \tau$$

Thus,

$$\bar{u} - \frac{2\tau}{I-\tau} + 2\ln\left(\frac{I-t}{I-\tau}\right) = 2\ln\left(\frac{I-t}{2}\right) \therefore \bar{u} = 2\left[\frac{\tau}{I-\tau} + \ln\left(\frac{I-\tau}{2}\right)\right].$$

Substituting back in the indirect utility, gives

$$\mathscr{U}(t) = \begin{cases} 2\left[\ln\left(\frac{I-\tau}{2}\right) + \frac{\tau-t}{I-\tau}\right] & \text{for } t \leq \tau \\ 2\ln\left(\frac{I-t}{2}\right) & \text{for } t > \tau \end{cases}.$$

Using the definition of the indirect utility, $\mathscr{U}(t) \equiv \ln(c(t)) + \ln(c_2(t))$, we can recover the second-period consumption for $t < \tau$:

$$c_2(t) = \frac{I-\tau}{2} \exp\left(2\frac{\tau-t}{I-\tau}\right).$$

For $t \geq \tau$, the individual lapses and gets perfect smoothing $c_1(t) = c_2(t) = \frac{I-t}{2}$. Finally, note that $c(t)+t$ is non-decreasing in t, so the omitted monotonicity constraint does not bind.

Rational Expectations (Subsection 4.2.5)

We now show that the equilibrium contract has the opposite pattern when consumers have rational expectations about the distribution of losses. Formally:

Proposition 5. *With rational expectations about the income shock, there exists $\tau \in [0,T]$ such that RP binds for $t \in [0,\tau)$ and does not bind for almost all $t \in (\tau,T]$. Moreover, $\mathscr{U}(t) < 2u(c(t))$ for almost all $t \in (\tau,T]$.*

The intuition for why the conclusions flip when consumers have rational expectations is as follows:

- Without IC and RP, the optimal contract when consumers have rational expectations provides full insurance (they are fully compensated for a loss without having to repay at all). This violates IC because consumers would always report a large loss, consuming the additional reimbursement. If RP fails, it fails for consumers with small losses who have to subsidize those with large losses. Thus, in equilibrium, policies offer subsidized loans (individuals have to repay part of it to maintain IC, but they are given below-market interest rates to smooth consumption), causing RP to bind at the bottom (with below-market interest rates, those with small loans subsidize those with large loans). That is, the key incentive issue with rational expectations is to provide insurance against unobservable liquidity shocks while ensuring that individuals do not inflate their losses.

- In contrast, when consumers think that they will not have any losses, the insurance company would like to charge as much as possible from those with positive losses, since consumers think that those states would not happen. This corresponds to charging an infinitely high interest rate for everyone with a loss, which gives consumers an incentive to pretend that they not have any income loss. To restore IC, the company cannot charge infinite rates, so it charges an above market but still finite interest rate. This disproportionately hurts consumers with larger losses, who now cross subsidize others by paying above-market interest rates. Thus, RP binds at for those with large enough losses.

The equilibrium program is:

$$\max_{c,\mathscr{U}} \int \mathscr{U}(t) f(t) dt$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)) \ \forall t, \tag{IC}$$

$$\mathscr{U}(t) \geq 2u\left(\frac{I-t}{2}\right) \ \forall t, \tag{RP}$$

$$\int_0^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) - t\right] f(t) dt \leq W$$

As a benchmark, ignore RP:

$$\max_{c,\mathscr{U}} \int \mathscr{U}(t) f(t) dt$$

subject to

$$\dot{\mathscr{U}}(t) = -u'(c(t)) \ \forall t, \tag{IC}$$

$$W - \int_0^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) - t\right] f(t) dt \geq 0$$

Note that without IC, we have

$$\max_{c,\mathscr{U}} \int \mathscr{U}(t) f(t) dt$$

subject to

$$W - \int_0^T \left[c(t) + u^{-1}(\mathscr{U}(t) - u(c(t))) - t\right] f(t) dt \geq 0.$$

Pointwise maximization gives:

$$f(t)-\frac{\lambda}{u'\left(u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)\right)}f(t)=0\therefore\mathscr{U}(t)-u\left(c(t)\right)=\text{constant}$$

$$1=\frac{u'(c)}{u'\left(u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)\right)}\therefore\mathscr{U}(t)=2u\left(c(t)\right).$$

Thus, the solution entails $c(t)=\bar{c}$ constant and $\mathscr{U}(t)=2u(\bar{c})$ (also constant) for all t. But this violates IC, which requires $\dot{\mathscr{U}}(t)=-u'(\bar{c})>0$.

We now incorporate IC. Introduce the auxiliary variable:

$$X(s)\equiv-\int_0^s\left[c(t)+u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)-t\right]f(t)dt,$$

so zero profits can be written as:

$$\dot{X}(t)=-\left[c(t)+u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)-t\right]f(t),$$

with $X(0)=0,\;\;X(T)=W$.

The equilibrium program becomes:

$$\max_{c,\mathscr{U}}\int\mathscr{U}(t)f(t)dt$$

subject to

$$\dot{\mathscr{U}}(t)=-u'\left(c(t)\right)\;\forall t, \tag{IC}$$

$$\dot{X}(t)=-\left[c(t)+u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)-t\right]f(t),$$

$X(0)=0,\;\;X(T)=W$, $\mathscr{U}(0)$ and $\mathscr{U}(T)$ free.

Set up the Hamiltonian:

$$H(\mathscr{U},c,\lambda,\mu,t)=\mathscr{U}f(t)-\lambda u'(c)-\mu\left[c(t)+u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)-t\right]f(t).$$

The optimality conditions are:

$$\lambda(t)=-\mu\left[1-\frac{u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)\right)}\right]\frac{f(t)}{u''(c(t))}$$

$$\dot{\lambda}(t)=\left[\frac{\mu}{u'\left(u^{-1}\left(\mathscr{U}(t)-u\left(c(t)\right)\right)\right)}-1\right]f(t)$$

$$\lambda(0)=\lambda(T)=0,\;\;\mu(t)=\mu\text{ constant}$$

From the transversality condition, it follows that $\mathscr{U}\left(c(0)\right)=2u\left(c(0)\right)$ and $\mathscr{U}\left(c(T)\right)=2u\left(c(T)\right)$. This implies that $\mu > 0$ (otherwise, we would have $\dot{\lambda} < 0$ for all t, which would violate transversality). This is consistent with the fact that, by the envelope theorem, μ is the shadow cost of wealth W, which has to be positive.

We now verify that we cannot have $\lambda(t) = 0$ for all t. For this to be the case, we would need $\mathscr{U}(t) = 2u(c(t))$ for all t, so that

$$\mathscr{U}(t) - u(c(t)) = u(c(t)) \ \forall t$$

Moreover, since $\dot{\lambda}(t) = 0$ for all t, we would have

$$u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) = \mu \ \ \forall t \therefore \mathscr{U}(t) - u(c(t)) \text{ is constant.}$$

Therefore, this is only possible if $c(t) = \bar{c}$ constant in t, which implies $\mathscr{U}(t) = 2u(\bar{c})$ constant in t. But this would give $\dot{\mathscr{U}}(t) = 0$, which violates IC.

Thus, we must have $\lambda(t) \neq 0$ in some interval of types (the interval requirement comes from the continuity of λ), which requires that the sign of $\dot{\lambda}(t)$ cannot be constant. Let

$$\xi(t) \equiv \mu - u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right),$$

and note that

$$\dot{\lambda}(t) > 0 \iff \mu > u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \iff \xi(t) > 0.$$

Differentiation gives

$$\begin{aligned}
\dot{\xi}(t) &= -u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \times \frac{d}{dt}\left[\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)\right] \\
&= -u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \times \frac{\dot{\mathscr{U}}(t) - u'(c(t))\,c'(t)}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)} \\
&= -u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \times \frac{-u'(c(t)) - u'(c(t))\,c'(t)}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)} \\
&= \frac{u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)}\left[1 + c'(t)\right].
\end{aligned}$$

By monotonicity, we must have $1 + c'(t) \geq 0$. Thus, we have $\dot{\xi}(t) < 0$, so that there exists $\tau \in (0, L)$ such that

$$\dot{\lambda}(t)\left\{\begin{array}{c} > \\ < \end{array}\right\} 0 \iff t \left\{\begin{array}{c} < \\ > \end{array}\right\} \tau.$$

It then follows that $\lambda(t) > 0$ for all $t \in (0, L)$:

$$\lambda(t) = -\mu\left[1 - \frac{u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)}\right]\frac{f(t)}{u''(c(t))} > 0 \iff u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) > u'(c(t))$$

$$\iff u^{-1}\left(\mathscr{U}(t)-u(c(t))\right)<c(t) \iff \mathscr{U}(t)<2u(c(t))$$

for all $t\in(0,L)$. That is, with rational liquidity shocks, loans are subsidized (below-market interest). This is because the IC binds in the opposite direction (absent IC, we would want to fully insure against losses, but then everyone would like to pretend to have had a higher shock).

To understand where RP may bind once we introduce it into the model, note that:

$$h(\mathscr{U}(t),t)=\mathscr{U}(t)-2u\left(\frac{I-t}{2}\right),$$

so that

$$\frac{d}{dt}\left[h(\mathscr{U}(t),t)\right]=\dot{\mathscr{U}}(t)+u'\left(\frac{I-t}{2}\right)=u'\left(\frac{I-t}{2}\right)-u'(c(t))\leq 0 \iff \frac{I-t}{2}\geq c(t).$$

We have seen that the solution of the relaxed program where we ignore RP entails $\mathscr{U}(t)<2u(c(t))$ for all $t\in(0,T)$. Thus, $u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)<c(t)$.

Suppose $\frac{d}{dt}\left[h(\mathscr{U}(t),t)\right]\leq 0$ for some t so that $\frac{I-t}{2}\geq c(t)$. Then, we must have

$$u^{-1}\left(\frac{\mathscr{U}(t)}{2}\right)<c(t)\leq\frac{I-t}{2}\therefore\mathscr{U}(t)<2u\left(\frac{I-t}{2}\right),$$

meaning that RP fails. Thus, if RP holds, we must have $\frac{d}{dt}\left[h(\mathscr{U}(t),t)\right]>0$. But this means that RP holds if and only if it holds at $t=0$. More generally, this shows that if the solution entails $\mathscr{U}(t)<2u(c(t))$, then any junction point must be an exit point. This is intuitive because the incentive issue here is to give insurance against large shocks so the type most willing to leave the mechanism is the one with the lowest shock.

We now formally introduce the RP constraint. Define the Hamiltonian and Lagrangian functions as:

$$H(\mathscr{U},c,\lambda,\mu,t)=\mathscr{U}f(t)-\lambda u'(c)-\mu\left[c(t)+u^{-1}\left(\mathscr{U}(t)-u(c(t))\right)-t\right]f(t)$$

$$L(c,\mathscr{U},\lambda,\mu,\eta,t)=H(c,\mathscr{U},\lambda,t)+\nu\left[u'\left(\frac{I-t}{2}\right)-u'(c)\right].$$

The optimality conditions are:

- $\lambda(t)+\nu(t)=-\mu\left[1-\frac{u'(c(t))}{u'\left(u^{-1}(\mathscr{U}(t)-u(c(t)))\right)}\right]\frac{f(t)}{u''(c(t))}$,
- $\dot{\lambda}(t)=\left[\frac{\mu}{u'\left(u^{-1}(\mathscr{U}(t)-u(c(t)))\right)}-1\right]f(t)$,
- $\mu(t)=\mu$ constant,
- $\nu(t)\geq 0$ with $=$ if either $c(t)>\frac{I-t}{2}$ or $\mathscr{U}(t)>2u\left(\frac{I-t}{2}\right)$,

- $\dot{\nu}(t) \leq 0$,
- $\lambda(T^-) \geq 0$, with $=$ if $\mathscr{U}(T^-) > 2u\left(\frac{I-T}{2}\right)$,
- $\lambda(0^+) \geq 0$, with $=$ if $\mathscr{U}(0^+) > 2u\left(\frac{I}{2}\right)$.

Moreover, at entry or contact points,

- $\lambda(\tau^-) = \lambda(\tau^+) + \eta(\tau)$
- $\left[c(\tau^-) + u^{-1}\left(\mathscr{U}(\tau) - u(c(\tau^-))\right)\right] f(t) + \lambda(\tau^-) u'(c(\tau^-)) =$

$$= \left[c(\tau^+) + u^{-1}\left(\mathscr{U}(\tau) - u\left(c(\tau^+)\right)\right)\right] f(t) + \lambda(\tau^+) u'(c(\tau^+)) + \eta(\tau) u'\left(\frac{I-t}{2}\right)$$

- $\eta(\tau) \geq 0$ with $=$ if $\mathscr{U}(t) > 2u\left(\frac{I-t}{2}\right)$.

In addition, $\eta(\tau_1) \geq \nu(\tau_1^+)$ at any entry time τ_1 and $\lambda(\tau_2)$ is continuous at any exit time τ_2.

We now show that there cannot be an entry point.

Lemma 12. *There are no entry points.*

Proof. Let τ be the **first entry point**. Then, we must have:

$$\lim_{t \nearrow \tau} \frac{d}{dt}\left[\mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right)\right] = \lim_{t \nearrow \tau}\left[\dot{\mathscr{U}}(t) + u'\left(\frac{I-t}{2}\right)\right] \leq 0$$

$$\iff u'\left(\frac{I-\tau}{2}\right) \leq u'(c(\tau^-) \iff c(\tau^-) \leq \frac{I-\tau}{2}. \tag{47}$$

Moreover, since *RP* binds at $(\tau, \tau+\varepsilon)$, we have $c(\tau^+) = \frac{I-\tau}{2}$. That is, if c jumps, it must jump upwards.

By the first optimality condition for $t < \tau$ (where RP doesn't bind), we have

$$\lambda(\tau^-) = -\mu\left[1 - \frac{u'(c(\tau^-))}{u'\left(u^{-1}\left(2u\left(\frac{I-\tau}{2}\right) - u(c(\tau^-))\right)\right)}\right]\frac{f(\tau)}{u''(c(\tau^-))},$$

and, because $c(\tau^-) \leq \frac{I-\tau}{2}$, we must have

$$\lambda(\tau^-) \leq 0. \tag{48}$$

By the first optimality condition for $t > \tau$ (where RP binds), we have $\lambda(\tau^+) + \nu(\tau^+) = 0$ for all such t. Recall that $\eta(\tau) \geq \nu(\tau^+)$ at any entry time. Thus,

$$\lambda(\tau^+) + \eta(\tau) \geq \lambda(\tau^+) + \nu(\tau^+) = 0.$$

Then, using $\lambda(\tau^-) = \lambda(\tau^+) + \eta(\tau)$, we obtain:

$$\lambda(\tau^-) \geq 0. \tag{49}$$

Combining (48) and (49), we find that

$$\lambda(\tau^-) = 0,$$

so that $c(\tau^-) = \frac{I-\tau}{2}$. Therefore, $c(\cdot)$ is continuous at τ.

There are two possible cases. Suppose first that there are no contact points between 0 and τ so that λ is continuous at $[0,\tau]$. Then, we have:

$$\lambda(0) = \lambda(\tau^-) = 0,$$

and

$$\dot{\lambda}(t) = \left[\frac{\mu}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)} - 1\right] f(t).$$

Let

$$\xi(t) \equiv \mu - u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right),$$

and note that

$$\dot{\lambda}(t) > 0 \iff \mu > u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \iff \xi(t) > 0.$$

Differentiation gives

$$\begin{aligned}
\dot{\xi}(t) &= -u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \times \frac{d}{dt}\left[\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)\right] \\
&= -u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \times \frac{\dot{\mathscr{U}}(t) - u'(c(t))\,c'(t)}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)} \\
&= -u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) \times \frac{-u'(c(t)) - u'(c(t))\,c'(t)}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)} \\
&= \frac{u''\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right) u'(c(t))}{u'\left(u^{-1}\left(\mathscr{U}(t) - u(c(t))\right)\right)}\left[1 + c'(t)\right].
\end{aligned}$$

By monotonicity, we must have $1 + \dot{c}(t) \geq 0$. Thus, we have $\dot{\xi}(t) < 0$. Therefore, there exists $\bar{t} \in (0,\tau)$ such that $\dot{\lambda}(t) > 0$ if and only if $t < \bar{t}$. Because $\lambda(0) = \lambda(\tau) = 0$, this implies that $\lambda(t) > 0$ for all $t \in (0,\tau)$, so that

$$c(\tau^-) > \frac{I-\tau}{2}$$

for all $t \in (0,\tau)$, a contradiction to (47).

Turning to the second possible case, suppose there exists a contact point $\hat{\tau}$ between 0 and τ. Since $\lambda(\tau^-) = \lambda(\tau^+) + \eta(\tau) \geq \lambda(\tau^+)$, it follows that λ can only jump downwards. In principle, we could then have an intermediate region with $\lambda(t) < 0$ (immediately after a contact point). This would not happen if either $\lambda(\tau^+) \geq 0$ or if $\lambda(\tau^-) \leq 0$ (so both sides of the contact point are either positive or negative).

Thus, the only possible case is when[51]

$$\lambda(\tau^-) \geq 0 \geq \lambda(\tau^+)$$

with at least one strict inequality, which happens if and only if

$$c(\tau^+) \leq \frac{I-\tau}{2} \leq c(\tau^-) \tag{50}$$

with at least one strict inequality.

But any contact point τ must satisfy

$$\lim_{t \nearrow \tau} \frac{d}{dt}\left[\mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right)\right] \leq 0 \leq \lim_{t \searrow \tau} \frac{d}{dt}\left[\mathscr{U}(t) - 2u\left(\frac{I-t}{2}\right)\right].$$

Use IC to rewrite this as

$$c(\tau^-) \leq \frac{I-\tau}{2} \leq c(\tau^+), \tag{51}$$

contradicting (50). □

To conclude the proof, note that $\lambda(\cdot)$ is continuous at any exit point, so that $\lambda(\tau) = 0$. Then, by the same argument as in the model without RP, it follows from $\lambda(0) = \lambda(\tau) = 0$ that $\lambda(\tau) > 0$ for all intermediate points.

Appendix B: LLI Survey Questions

This Appendix describes the LLI (2018) survey methodology, lists all questions and along with the responses to each of them. Using administrative data provided by LLI, it also compares responders ("Complete") and non-responders ("Incomplete") by observable characteristics. The input files (survey results) and processing files can be found in the online repository. Both surveys were emailed on February 7, 2018, specifically, at 11:28 AM EST for the New Buyers survey and at 10:52 AM EST for the Lapsers

[51] Recall that

$$\lambda(t) > 0 \iff \mathscr{U}(t) < 2u(c(t)).$$

Since RP binds at τ, we have $\mathscr{U}(\tau) = 2u\left(\frac{I-\tau}{2}\right)$. Thus,

$$\lambda(\tau^+) < 0 \iff \frac{I-\tau}{2} > c(\tau^+)$$

$$\lambda(\tau^-) > 0 \iff \frac{I-\tau}{2} < c(\tau^-)$$

Thus, the possible case is when:

$$\lambda(\tau^-) > 0 > \lambda(\tau^+) \iff c(\tau^+) < \frac{I-\tau}{2} < c(\tau^-).$$

Survey. A reminder email message was sent to people in both surveys who had not completed the survey (both non-responders and people who started but did not complete) on February 14th at 12:00 PM EST. A second reminder email was sent on February 21st at 12:00 PM EST. Both surveys were closed at 9AM EST on March 9, 2018. Following standard IRB protocols, the first page of each survey informed subjects of the purpose of the survey and gave them our contact information in case they had any concerns. Neither author received any emails.

New Buyers Survey

Below are the questions for the New Buyers Survey along with percent responses for each question shown in parentheses. This survey was sent to all LLI customers who purchased life insurance between October 2013 and end of November 2017. Subjects were asked either 7 or 8 questions. The response rate was 13.0%, producing 1,689 respondents.

1. Your term life insurance policy has about *N* years left on it. What is the chance that you might stop your policy (sometimes called lapsing) before then?

1.1. I have not given it much thought (12.0%, 202 responders)
1.2. I do not currently anticipate stopping my policy (80.4%. 1357)
1.3. I currently anticipate stopping my policy with a 10 percent or lower chance (1.4%, 24)
1.4. I currently anticipate stopping my policy with a chance of 10 – 25 percent (1.7%, 28)
1.5. I currently anticipate stopping my policy with a chance of 25 – 50 percent (2.1%, 35)
1.6. I anticipate stopping my policy with a chance greater than 50 percent (2.4%, 41)

2. If 1.3, 1.4, 1.5 or 1.6: In how many years do you anticipate potentially stopping your policy?

2.1. Unsure (20.3%, 26)
2.2. Likely in between 1-5 years (28.9%, 37)
2.3. Likely in between 6-10 years (27.3%, 35)
2.4. Likely in between 11-15 years (9.4%, 12)
2.5. More than 15 years (14.1%, 18)

3. What do estimate is the chance that other people with your type of life insurance policy might stop their policy (or lapse) before it expires?

3.1. I have not given it much thought (67.1%, 1131)
3.2. Between 0 – 5 percent (7.7%, 129)
3.3. Between 5 – 10 percent (6.6%, 112)
3.4. Between 10 – 25 percent (10.7%, 181)
3.5. Between 25 – 50 percent (5.8%, 97)
3.6. Over 50 percent (2.1%, 35)

4. What do estimate is the chance that you might someday stop your policy due to divorce or death of a spouse?

4.1. I have not given it much thought (34.7%, 584)

4.2. Between 0 – 5 percent (55.0%, 927)
4.3. Between 5 – 10 percent (4.9%, 83)
4.4. Between 10 – 25 percent (2.6%, 43)
4.5. Between 25 – 50 percent (1.0%, 17)
4.6. Over 50 percent (1.8%, 31)

5. What do estimate is the chance that you might someday stop your policy because you will need the money, maybe due to lower income or increased expenses?

5.1. I have not given it much thought (18.3%, 308)
5.2. Between 0 – 5 percent (63.6%, 1073)
5.3. Between 5 – 10 percent (9.0%, 151)
5.4. Between 10 – 25 percent (4.2%, 71)
5.5. Between 25 – 50 percent (3.0%, 51)
5.6. Over 50 percent (1.9%, 32)

6. What do estimate is the chance that you might someday stop your policy because you feel healthier than expected and would prefer to purchase a different policy?

6.1. I have not given it much thought (24.1%, 406)
6.2. Between 0 – 5 percent (70.2%, 1183)
6.3. Between 5 – 10 percent (0%, 0)
6.4. Between 10 – 25 percent (2.5%, 42)
6.5. Between 25 – 50 percent (1.4%, 23)
6.6. Over 50 percent (0.9%, 15)

7. At some point in the last 5 years, has your total household income decreased? This might be due to a salary cut, a job separation by you or your spouse, or because part of your total household income is partly tied to commissions or bonuses that tend to fluctuate.

7.1. Yes (32.5%, 545)
7.2. No (67.5%, 1132)

8. What are the chances that at some point in the next 5 years, your total household income would decrease substantially? This might due to a salary cut, a layoff of you or a spouse, retirement, or because part of your total household income is partly tied to commissions or bonuses that tend to fluctuate?

8.1. There is little chance that my income could fluctuate downward by a lot (43.6%, 730)
8.2. My income could fluctuate downward a lot with a chance less than 5% (13.7%, 229)
8.3. My income could fluctuate downward a lot with a chance between 5 – 10% (15.4%, 257)
8.4. My income could fluctuate downward a lot with a chance between 10 – 25% (13.7%, 229)
8.5. My income could fluctuate downward a lot with a chance between 25 – 50% (7.2%, 120)
8.6. My income could fluctuate downward a lot with a chance greater than 50% (6.5%, 108)

Responders who answered Question 1 indicating a 10% chance or more of lapsing (1.4, 1.5, or 1.6) completed the survey in 6.2 days on average in comparison to 6.7 days for responders who indicated a 10 percent or less chance (1.1, 1.2, or 1.3). The t-value for a test of equal means between the two independent samples is 0.78, thereby failing to reject a difference at conventional levels of significance.

Notice that subjects were asked to report both the total probability of lapsing (Question 1) and the probability of lapsing broken down by possible reasons (Questions 4, 5, and 6). In both cases, we find that they severely underestimate the chance of lapsing. Consistently with the Conjunction Fallacy (Tversky and Kahneman, 1983), we find that lapse probabilities broken down by each reason add up to more than the total probability of lapsing. For example, 6.2% of responders indicated a 10% or higher chance of lapsing in general (Question 1), whereas 9.1% of responders indicate a 10% or higher chance of lapsing because they might someday need money (Question 5). Overall, 13.0% of responders indicated a 10% or less chance of lapsing in Question 1 while simultaneously indicating a 10% or more chance of lapsing for any one or more specific reason when subsequently prompted in Questions 4, 5 and 6. The relevant probability for our purposes depends on whether, when purchasing life insurance, individuals think about the probability of lapsing as a whole, or if they think about each possible reason for lapsing in isolation. In either case, however, these numbers support the view that individuals severely underestimate the probability of lapsing.[52]

Figure 4 shows the average characteristics of responders ("Complete") and non-responders ("Incomplete"), with 95% confidence intervals shown with the black line. At the time of the survey, responders have had held their policy for about 26 days less than non-responders. The expected default rate of responders (55.8%) is slightly larger than non-responders (57.5%). Most other characteristics are very similar, including marital status, gender, length of policy, occupation, age at policy issuance and ultimate face value. A slightly higher proportion of college employees responded to our survey, perhaps because we identified ourselves in the opening page and stated that the survey would be used for academic research. This small difference of occupations might also explain why responders are slightly older and have slightly smaller policies, as many college employees get supplemental life insurance through their college as a voluntary group benefit, outside of the individual insurance market that we consider.

Lapsers Survey

A link to this survey was sent by email to the universe of 3,229 former LLI policyholders who lapsed their policies between 2012 and 2017, generating a response rate of 4.9 percent. Subjects were asked either 6 or 7 questions. Below are the questions for the Lapsers Survey along with percent responses for each question shown in parentheses. For Question 1, there are two percent responses shown: the first one corresponds to the raw responses and the second one corresponds to our subjective recoding for those who selected 1.9 in the raw data.

1. You have recently cancelled (or let "lapse") your life insurance policy. Many people cancel / lapse their policies for one or more of the reasons listed below. Which choice best reflects your

[52] Alternatively, subjects may infer from Question 1 that lapse probabilities are higher than expected, revising their answers to subsequent questions about lapsing upwards.

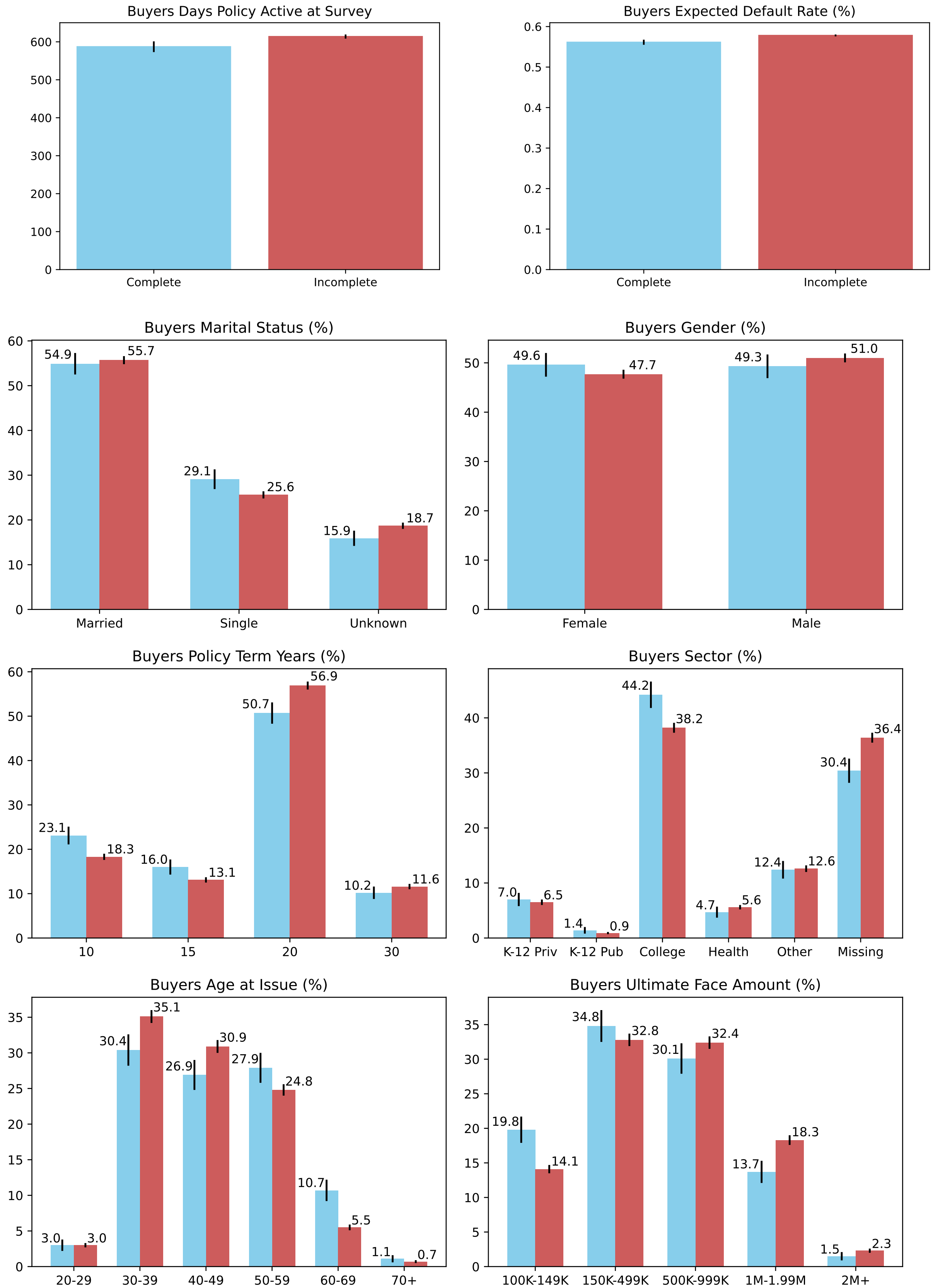


Figure 4: Descriptive statistics for respondents (blue) and non-respondents (red). Black lines represent 95% confidence intervals. Starting from the top, the figures represent days that policy has been active at point of survey; expected default rate of policies based on LLI's historical annual default rate of 5.2%; marital status; gender; term of policy; occupation; age and ultimate face amount.

reason?

1.1. My income decreased (7.6%, 12; 8.3%, 13)

1.2. I needed the money (1.3%, 2; 7.1%, 11)

1.3. My family situation changed due to divorce (4.5%, 7; 4.5%, 7)

1.4. My family situation changed due to death of spouse (0.6%, 1; 0.6%, 1)

1.5. I recently retired (12.7%, 20; 13.5%, 21)

1.6. I was healthier than expected and bought another policy (1.3%, 2; 6.4%, 10)

1.7. I forgot to make my insurance premium payments (12.1%, 19; 23.1%, 36)

1.8. I believe that I didn't cancel my policy (12.7%, 20; 14.7%, 23)

1.9. Other (please explain) (47.1%, 74; 12.8%, 20)

1.10. Family situation changed for reasons other than divorce or death of spouse (N/A%, N/A; 9.0%, 14)

2. If 1.8: Our records indicate that you cancelled your [type] insurance with a death benefit of [$$$$$$] on [DATE]. Do you recall that cancellation?

2.1. Yes (5.0%, 1)

2.2. No (95.0%, 19)

3. At some point in the last 5 years, has your total household income decreased? This might be due to a salary cut, a job separation by you or your spouse, or because part of your total household income is partly tied to commissions or bonuses that tend to fluctuate.

3.1. Yes (44.2%, 69)

3.2. No (55.8%, 87)

4. IF 3.1: By how much did your total household income decrease in the last five years?

4.1. Less than 5% (0.0%, 0)

4.2. Between 5 and 15% (23.5%, 16)

4.3. Between 15 and 25% (26.5%, 18)

4.4. Between 25 and 50% (23.5%, 16)

4.5. More than 50% (26.5%, 18)

5. Have you experienced one of the following options in the last 5 years? Check all options that apply:

5.1. A divorce (18.6%, 13)

5.2. Retirement by you or your spouse (61.4%, 43)

5.3. Hospitalization by you or your spouse (37.1%, 26)

6. Since you cancelled your policy, have you purchased a new one?

6.1. No. (82.5%, 127)

6.2. Yes. I purchased a smaller policy (9.1%, 14).

6.3. Yes. I purchased a larger policy (8.4%, 13).

7. What is your annual household income?

7.1. Less than $50,000 per year (12.2%, 19)

7.2. Between $50,000 - $125,000 per year (39.7%, 62)

7.3. Between $125,000 - $250,000 per year (25.6%, 40)
7.4. Over $250,000 per year (5.1%, 8)
7.5. I prefer to not answer (17.3%, 27)

Figure 5 shows the average characteristics of responders ("Complete") and non-responders ("Incomplete"), with 95% confidence intervals shown with the black line. Notice that responders tend to be slightly older (mean age of 49) relative to non-responders (mean age of 45), have smaller policies, and have lapsed a bit more recently. The mean number of days that policies were active before lapsing nearly are identical for responders and non-responders. Both groups are similar in terms of gender and marital status.

Appendix C: Competing Models

Model of Risk Reclassification

This section considers reclassification risk model based on Hendel and Lizzeri (2003), Daily, Hendel, and Lizzeri (2008), and Fang and Kung (2010). We demonstrate that a reasonably calibrated rational model with liquidity shocks produces back-loaded policies, the opposite loading of observable contracts.

The main distinction between the model considered here and the other ones in the literature is in the timing of shocks. Hendel and Lizzeri (2003) study a model in which consumers are subject to health shocks only. Lack of commitment on the side of the consumer motivates lapsing following positive health shocks. Preventing lapses is then welfare improving and front-loaded fees (i.e., payments before the realization of the health shock that cannot be recuperated if the consumer drops the policy) are an effective way to do so. Daily, Hendel, and Lizzeri (2008) and Fang and Kung (2010) introduce bequest shocks in this framework. In their model, there is one period in which both bequest and health shocks may happen. Lapsing is efficient if it is due to a loss of the bequest motive and is inefficient if motivated by a positive health shock. The solution then entails some amount of front loading as a way to discourage lapses.

Markov Transition Matrix (25 year old Male; 5 years)

	1	2	3	4	5	6	7	8
1	*.989*	*.001*	*.000*	*.000*	*.000*	*.000*	*.000*	*.011*
2	*.932*	*.028*	*.000*	*.000*	*.000*	*.000*	*.000*	*.039*
3	*.927*	*.030*	*.000*	*.000*	*.000*	*.000*	*.000*	*.042*
4	*.918*	*.034*	*.000*	*.000*	*.000*	*.000*	*.000*	*.046*
5	*.860*	*.056*	*.000*	*.000*	*.000*	*.000*	*.000*	*.082*
6	*.914*	*.038*	*.000*	*.000*	*.000*	*.000*	*.000*	*.048*
7	*.850*	*.060*	*.000*	*.000*	*.001*	*.000*	*.000*	*.088*

Table 1: Probability of five-year ahead changes in health states at age 25.

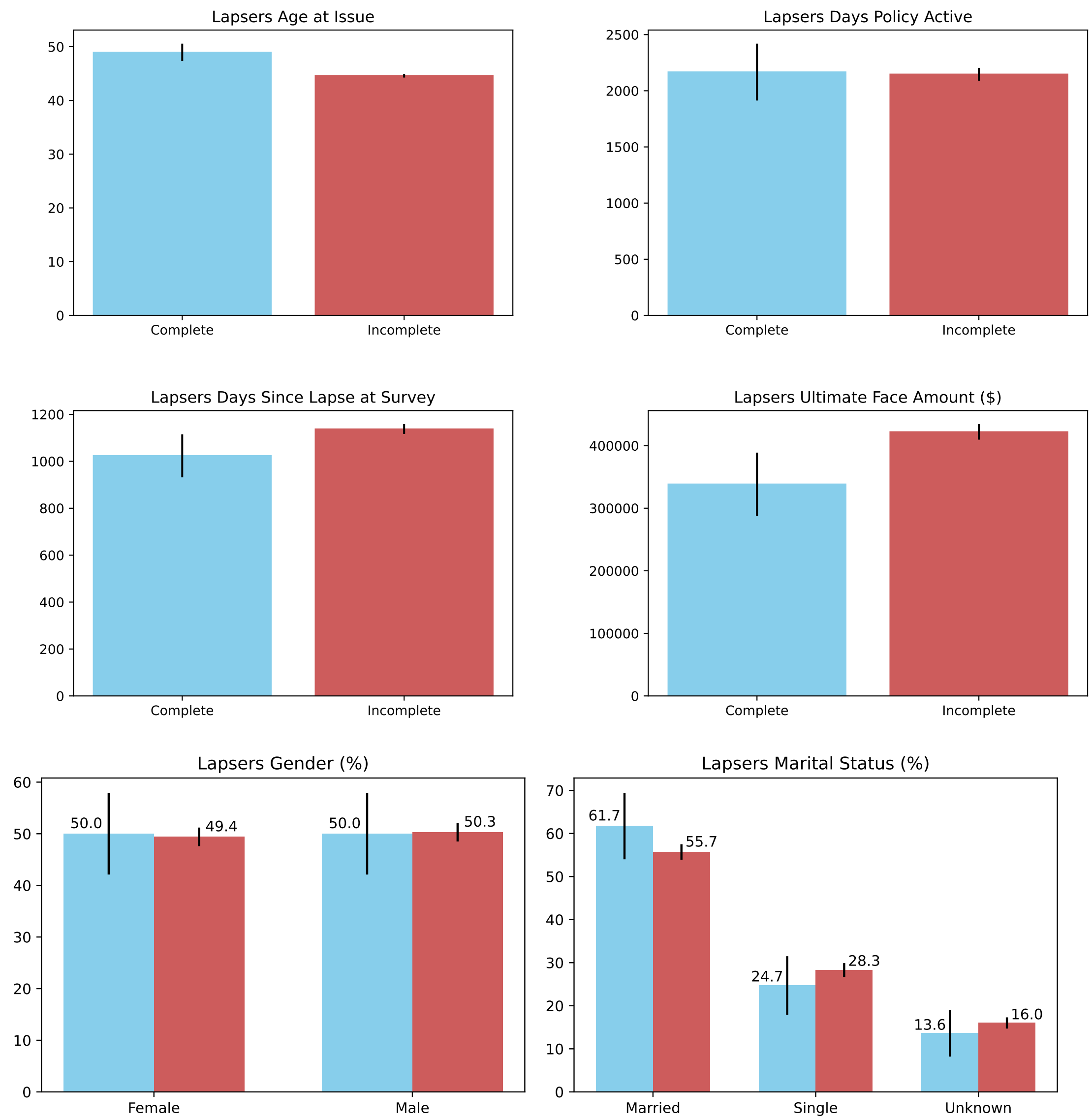


Figure 5: Descriptive statistics for respondents (blue) and non-respondents (red). Black lines represent 95% confidence intervals. Starting from the top, the figures represent age at issuance; number days before policy lapsed; number of days between lapse and survey; ultimate face amount; gender; and marital status. Differences between other variables not shown were typically not statistically significant, likely due to low response rate.

	1	2	3	4	5	6	7	8
	Markov Transition Matrix (50 year old Male; 5 years)							
1	*.942*	*.014*	*.001*	*.000*	*.001*	*.001*	*.000*	*.041*
2	*.544*	*.252*	*.009*	*.004*	*.011*	*.006*	*.002*	*.172*
3	*.515*	*.259*	*.010*	*.005*	*.012*	*.006*	*.002*	*.190*
4	*.446*	*.285*	*.012*	*.007*	*.020*	*.007*	*.003*	*.219*
5	*.257*	*.273*	*.020*	*.020*	*.065*	*.007*	*.005*	*.353*
6	*.430*	*.296*	*.014*	*.009*	*.027*	*.008*	*.004*	*.212*
7	*.229*	*.267*	*.021*	*.022*	*.074*	*.007*	*.006*	*.374*

Table 2: Probability of five-year ahead changes in health states at age 50.

Markov Transition Matrix (75 year old Male; 5 years)

	1	2	3	4	5	6	7	8
1	*.645*	*.103*	*.014*	*.005*	*.014*	*.016*	*.008*	*.195*
2	*.129*	*.235*	*.038*	*.016*	*.040*	*.036*	*.024*	*.482*
3	*.094*	*.198*	*.035*	*.017*	*.048*	*.032*	*.025*	*.551*
4	*.046*	.136	*.031*	*.023*	*.078*	*.025*	*.033*	*.629*
5	*.011*	*.046*	*.016*	*.019*	*.095*	*.011*	*.032*	*.771*
6	*.052*	*.150*	*.035*	*.021*	*.079*	*.052*	*.048*	*.562*
7	*.009*	*.036*	*.013*	*.015*	*.087*	*.013*	*.039*	*.787*

Table 3: Probability of five-year ahead changes in health states at age 75.

The composition of shocks changes significantly along the life cycle. The tables above show “snap shots” across different ages of five-year ahead Markov health transition matrices based on hazard rates provided by Robinson (1996). State 1 represents the healthiest state while State 8 represents the worst (death). As the matrices show, younger individuals are unlikely to suffer negative health shocks and the ones who do experience such shocks typically recover within the next 5 years (with the obvious exception of death, which is). Older individuals are more likely to suffer negative heath shocks, and those shocks are substantially more persistent. Consistently, policyholders younger than about 65 rarely surrender due to health shocks whereas health shocks are considerably more important for older policyholders (c.f., Fang and Kung 2012).

Consistently with these observations, we consider a stylized model in which the period of shocks is broken down in two periods. In the first period, consumers are subject to non-health shocks only. In the second period, they are only subject to health shocks. As a result, optimal contracts are *back loaded*: they do not discourage lapses in the first period but discourage lapses in the second period. Because only health-related lapsing is inefficient, lapse fees should be high only in periods in which health shocks are relatively prevalent. Empirically, these periods occur much later in life.

Formally, there are 4 periods: $t = 0,1,2,3$. Period 0 is the contracting stage. Consumers are subject to a liquidity shock $L > 0$ (with probability $l > 0$) in period 1. They are subject to a health shock in period 2. The health shock is modeled as follows. With probability $\pi > 0$, the consumer finds out that he has a high risk of dying (*type H*). With complementary probability, he finds out that he has a low risk of death (*type L*). Then, in period 3, a high-risk consumer dies with probability α_H and a low-risk consumer dies with probability α_L, where $0 < \alpha_L < \alpha_H < 1$. We model lapses as motivated by liquidity/income shocks rather than bequest shocks because, as shown by First, Fang and Kung (2012), bequest shocks are responsible for a rather small proportion of lapses, whereas other (i.e. non-health and non-bequest shocks) are responsible for most of it, especially for individuals below a certain age. The assumption that mortality shocks only happen in the last period is for simplicity only. Our result remains if we assume that there is a positive probability of death in each period.

The timing of the model is as follows:

- Period 0: The consumer makes a take-it-or-leave-it offer of a contract to a non-empty set of firms. A contract is a vector of state-contingent payments to the firm

$$\left\{t_0, t_1^s, t_2^{s,h}, t_3^{d,s,h}\right\}_{s=S,NS\ h=H,L\ d=D,A},$$

  where: t_0 is paid in period 0 before any information is learned; t_1^s is paid conditional on the liquidity shocks in period 1, $s = S, NS$; $t_2^{s,h}$ is paid conditional on the health shock $h \in \{H, L\}$ in period 2 and liquidity shock s in period 1; $t_3^{d,s,h}$ is paid conditional on being either dead $d = D$ or alive $d = A$ in period 3 conditional on previous shocks s and h.

- Period 1: The consumer observes the realization of the liquidity shock s. He then decides whether to keep the original contract, thereby paying t_1^s, or obtaining a new contract in a competitive secondary market. The competitive secondary market is again modeled by having the consumer make a take-it-or-leave-it offer a (non-empty) set of firms.

- Period 2: The realization of the health shock is publicly observed. The consumer decides to keep the contract, thereby paying $t_2^{s,h}$, or substitute by a new one, obtained again in a competitive environment (in which the consumer makes a take-it-or-leave it offer to firms).

- Period 3: The mortality shock is realized. The consumer receives a payment of $-t_3^{d,s,h}$.

As before, we assume that consumers and firms discount the future at the same rate and normalize the discount rate to zero. Consumers get utility $u_A(c)$ of consuming c units (while alive). Consumers get utility $u_D(c)$ from bequeathing c units. The functions u_A and u_D satisfy the Inada condition: $\lim_{c \searrow 0} u_d(c) = -\infty$, $d = A, D$.

With no loss of generality, we can focus on period-0 contracts that the consumer never finds it optimal to drop. That is, we may focus on contracts that satisfy "non-reneging constraints." Of course, this is not to say that the equilibrium contracts will never be dropped in the same way that the revelation principle does not say that in the real world people should be "announcing their types." To wit, any allocation implemented by a non-reneging contract can also be implemented by a mechanism in which the consumer is given resources equal to the expected amount of future consumption and gets a new contract (from possibly a different firm) in each period. In particular, the model cannot distinguish between lapsing an old contract and substituting it by a new (state-contingent) contract and having an initial contract that is never lapsed and features state-dependent payments that satisfy the non-reneging constraint. However, the model determines payments in each state.

Consistently with actual (whole) life insurance policies, one can interpret the change of terms following a liquidity shock in period 1 as the lapsing of a policy at some predetermined cash value and the purchase of a new policy, presumably with a smaller coverage. We ask the following question: Is it possible for a firm to profit from lapses motivated by a liquidity shock? In other words, it is possible for the firm to get higher expected profits conditional on the consumer experiencing a liquidity shock in period 1 than conditional on the consumer not experiencing a liquidity shock? As we have seen in the evidence described in Section 2, firms do profit from such lapses, which are the most common source of lapses for policyholders below a certain age. However, as we show below, this is incompatible with the reclassification risk model described here.

The intuition for the result is straightforward. The reason why individuals prefer to purchase insurance at 0 rather than 1 is the risk of needing liquidity and therefore facing a lower wealth. If the insurance company were to profit from the consumers who suffer the liquidity shock, it would need to charge a higher premium if the consumer suffers the shock. However, this would exacerbate the liquidity shock. In that case, the consumer would be better off by waiting to buy insurance after the realization of the shock.

As in the text, there is no loss of generality in working with the space of state-contingent consumption rather than transfers. The consumer's expected utility is

$$u_A(c_0) + l \left\{ \begin{array}{c} u_A\left(c_1^S\right) + \pi\left[u_A\left(c_2^{S,H}\right) + (1-\alpha_H)\,u_A\left(c_3^{S,H,A}\right) + \alpha_H u_D\left(c_3^{S,H,D}\right)\right] \\ +(1-\pi)\left[u_A\left(c_2^{S,L}\right) + (1-\alpha_L)\,u_A\left(c_3^{S,L,A}\right) + \alpha_L u_D\left(c_3^{S,L,D}\right)\right] \end{array} \right\}$$

$$+(1-l) \left\{ \begin{array}{c} u_A\left(c_1^{NS}\right) + \pi\left[u_A\left(c_2^{NS,H}\right) + (1-\alpha_H)\,u_A\left(c_3^{NS,H,A}\right) + \alpha_H u_D\left(c_3^{NS,H,D}\right)\right] \\ +(1-\pi)\left[u_A\left(c_2^{NS,L}\right) + (1-\alpha_L)\,u_A\left(c_3^{NS,L,A}\right) + \alpha_L u_D\left(c_3^{NS,L,D}\right)\right] \end{array} \right\}.$$

The equilibrium contract maximizes this expression subject to the following constraints. First, the firm cannot be left with negative profits:

$$c_0 + l \left\{ \begin{array}{l} c_1^S + \pi \left[c_2^{S,H} + (1-\alpha_H) c_3^{S,H,A} + \alpha_H c_3^{S,H,D} \right] \\ + (1-\pi) \left[c_2^{S,L} + (1-\alpha_L) c_3^{S,L,A} + \alpha_L c_3^{S,L,D} \right] \end{array} \right\}$$

$$+ (1-l) \left\{ \begin{array}{l} c_1^{NS} + \pi \left[c_2^{NS,H} + (1-\alpha_H) c_3^{NS,H,A} + \alpha_H c_3^{NS,H,D} \right] \\ + (1-\pi) \left[c_2^{NS,L} + (1-\alpha_L) c_3^{NS,L,A} + \alpha_L c_3^{NS,L,D} \right] \end{array} \right\}$$

$$\leq W + I \left[2 - \pi \alpha_H - (1-\pi) \alpha_L \right] - lL$$

Second, allocation has to satisfy the incentive-compatibility constraints (which state that the consumer prefers the report of the liquidity shock honestly):

$$u_A\left(c_1^S\right) + \pi \left[u_A\left(c_2^{S,H}\right) + (1-\alpha_H) u_A\left(c_3^{S,H,A}\right) + \alpha_H u_D\left(c_3^{S,H,D}\right) \right]$$

$$+ (1-\pi) \left[u_A\left(c_2^{S,L}\right) + (1-\alpha_L) u_A\left(c_3^{S,L,A}\right) + \alpha_L u_D\left(c_3^{S,L,D}\right) \right] \geq$$

$$u_A\left(c_1^{NS} - L\right) + \pi \left[u_A\left(c_2^{NS,H}\right) + (1-\alpha_H) u_A\left(c_3^{NS,H,A}\right) + \alpha_H u_D\left(c_3^{NS,H,D}\right) \right]$$

$$+ (1-\pi) \left[u_A\left(c_2^{NS,L}\right) + (1-\alpha_L) u_A\left(c_3^{NS,L,A}\right) + \alpha_L u_D\left(c_3^{NS,L,D}\right) \right],$$

and

$$u_A\left(c_1^{NS}\right) + \pi \left[u_A\left(c_2^{NS,H}\right) + (1-\alpha_H) u_A\left(c_3^{NS,H,A}\right) + \alpha_H u_D\left(c_3^{NS,H,D}\right) \right]$$

$$+ (1-\pi) \left[u_A\left(c_2^{NS,L}\right) + (1-\alpha_L) u_A\left(c_3^{NS,L,A}\right) + \alpha_L u_D\left(c_3^{NS,L,D}\right) \right] \geq$$

$$u_A\left(c_1^S + L\right) + \pi \left[u_A\left(c_2^{S,H}\right) + (1-\alpha_H) u_A\left(c_3^{S,H,A}\right) + \alpha_H u_D\left(c_3^{S,H,D}\right) \right]$$

$$+ (1-\pi) \left[u_A\left(c_2^{S,L}\right) + (1-\alpha_L) u_A\left(c_3^{S,L,A}\right) + \alpha_L u_D\left(c_3^{S,L,D}\right) \right].$$

The third set of constraints requires contracts to be non-reneging after it has been agreed upon (that is, in periods 1 and 2). The period-2 non-reneging constraints are

$$u_A\left(c_2^{s,h}\right) + (1-\alpha_h) u_A\left(c_3^{A,NS,h}\right) + \alpha_h u_D\left(c_3^{D,s,h}\right) \geq \max_{\{\hat{c}\}} \left\{ \begin{array}{c} u_A(\hat{c}_2) + (1-\alpha_h) u_A(\hat{c}_3) + \alpha_h u_D(\hat{c}_3) \\ \text{s.t. } \hat{c}_2 + (1-\alpha_h)\hat{c}_3 + \alpha_h \hat{c}_3 = (2-\alpha_h) I \end{array} \right\}, \tag{52}$$

for $h = H, L$ and $s = S, NS$. The period-1 non-reneging constraints are

$$u_A\left(c_1^s\right) + \pi \left[u_A\left(c_2^{s,H}\right) + (1-\alpha_H) u_A\left(c_3^{s,H,A}\right) + \alpha_H u_D\left(c_3^{s,H,D}\right) \right]$$

$$+ (1-\pi) \left[u_A\left(c_2^{s,L}\right) + (1-\alpha_L) u_A\left(c_3^{s,L,A}\right) + \alpha_L u_D\left(c_3^{s,L,D}\right) \right] \geq$$

$$\max_{\{\hat{c}\}} \begin{array}{l} u_A\left(\hat{c}_1^s\right)+\pi\left[u_A\left(\hat{c}_2^{s,H}\right)+\left(1-\alpha_H\right)u_A\left(\hat{c}_3^{s,H,A}\right)+\alpha_H u_D\left(\hat{c}_3^{s,H,D}\right)\right] \\ \quad +\left(1-\pi\right)\left[u_A\left(\hat{c}_2^{s,L}\right)+\left(1-\alpha_L\right)u_A\left(\hat{c}_3^{s,L,A}\right)+\alpha_L u_D\left(\hat{c}_3^{s,L,D}\right)\right] \end{array}$$

subject to

$$\hat{c}_1^s+\pi\left[\hat{c}_2^{s,H}+\left(1-\alpha_H\right)\hat{c}_3^{s,H,A}+\alpha_H\hat{c}_3^{s,H,D}\right]+\left(1-\pi\right)\left[\hat{c}_2^{s,L}+\left(1-\alpha_L\right)\hat{c}_3^{s,L,A}+\alpha_L\hat{c}_3^{s,L,D}\right]$$

$$\leq I\left[2-\pi\alpha_H-\left(1-\pi\right)\alpha_L\right]-\chi_{s=S}L,$$

and

$$u_A\left(\hat{c}_2^{s,h}\right)+\left(1-\alpha_h\right)u_A\left(\hat{c}_3^{A,NS,h}\right)+\alpha_h u_D\left(\hat{c}_3^{D,s,h}\right)\geq \max_{c_2,c_3^A,c_3^D}\left\{\begin{array}{c} u_A\left(c_2\right)+\left(1-\alpha_h\right)u_A\left(c_3^A\right)+\alpha_h u_D\left(c_3^D\right) \\ \text{s.t. } c_2+\left(1-\alpha_h\right)c_3^A+\alpha_h c_3^D=\left(2-\alpha_h\right)I \end{array}\right\},$$

for $s=S,NS$, where χ_x denotes the indicator function.

We will define a couple of "indirect utility" functions that will be useful in the proof by simplifying the non-reneging constraints. First, for $h=H,L$ we introduce the function $U_h:\mathbb{R}_+\to\mathbb{R}$ defined as

$$U_h\left(W\right)\equiv\max_{c^A,c^D}\left\{\begin{array}{c} \left(2-\alpha_h\right)u_A\left(c^A\right)+\alpha_h u_D\left(c^D\right) \\ \text{s.t. } \left(2-\alpha_h\right)c^A+\alpha_h c^D\leq W \end{array}\right\}.$$

It is straightforward to show that U_h is strictly increasing and strictly concave. Next, we introduce the function $\mathscr{U}:\mathbb{R}_+\to\mathbb{R}$ defined as

$$\mathscr{U}\left(W\right)\equiv\max_{c,C^L,C^H}\left\{\begin{array}{c} u_A\left(c\right)+\pi U(C^H)+\left(1-\pi\right)U(C^L) \\ s.t.\ c+\pi C^H+\left(1-\pi\right)C^L\leq W \\ \left(2-\alpha_H\right)I\leq C^H \\ \left(2-\alpha_L\right)I\leq C^L \end{array}\right\}. \tag{53}$$

It is again immediate to see that $\mathscr{U}$ is strictly increasing. The following lemma establishes that it is also strictly concave:

Lemma 13. *$\mathscr{U}$ is a strictly concave function.*

Proof. Let

$$\mathscr{U}_0(W)\equiv\max_{C^L,C^H}\left\{u_A\left(W-\pi C^H-\left(1-\pi\right)C^L\right)+\pi U(C^{s,H})+\left(1-\pi\right)U(C^{s,L})\right\},$$

$$\mathscr{U}_1(W)\equiv\max_{C^L,C^H}\left\{\begin{array}{c} u_A\left(W-\pi C^H-\left(1-\pi\right)C^L\right)+\pi U(C^{s,H})+\left(1-\pi\right)U(C^{s,L}) \\ \text{subject to } \left(2-\alpha_L\right)I=C^H \end{array}\right\}, \text{ and}$$

$$\mathscr{U}_2(W) \equiv \max_{C^L, C^H} \left\{ \begin{array}{c} u_A\left(W - \pi C^H - (1-\pi) C^L\right) + \pi U(C^{s,H}) + (1-\pi) U(C^{s,L}) \\ \text{subject to } \ (2-\alpha_H) I = C^H \\ (2-\alpha_L) I = C^L \end{array} \right\}.$$

Notice that $\mathscr{U}_0(W) \geq \mathscr{U}_1(W) \geq \mathscr{U}_2(W)$, and $\mathscr{U}_0$, $\mathscr{U}_1$, and $\mathscr{U}_2$ are strictly concave. It is straightforward to show that there exist W_L and $W_H > W_L$ such that:

- $\mathscr{U}(W) = \mathscr{U}_0(W)$ for $W \geq W_H$,
- $\mathscr{U}(W) = \mathscr{U}_1(W)$ for $W \in [W_L, W_H]$, and
- $\mathscr{U}(W) = \mathscr{U}_2(W)$ for $W \leq W_L$.

Moreover, by the envelope theorem, $\mathscr{U}_0'(W_H) = \mathscr{U}_1'(W_H)$ and $\mathscr{U}_1'(W_L) = \mathscr{U}_2'(W_L)$. Therefore,

$$\mathscr{U}'(W) = \left\{ \begin{array}{c} \mathscr{U}_0'(W) \text{ for } W \geq W_H \\ \mathscr{U}_1'(W) \text{ for } W_L < W \leq W_H \\ \mathscr{U}_2'(W) \text{ for } W < W_L \end{array} \right. .$$

Because $\mathscr{U}'$ is strictly decreasing in each of these regions and is continuous, it then follows that $\mathscr{U}$ is strictly concave. □

Let X^s be the sum of the insurance company's expected expenditure at time t=1 conditional on s in the original contract:

$$X^s \equiv c_1^s + \pi \left[c_2^{s,H} + (1-\alpha_H) c_3^{s,H,A} + \alpha_H c_3^{s,H,D} \right] + (1-\pi) \left[c_2^{s,L} + (1-\alpha_L) c_3^{s,L,A} + \alpha_L c_3^{s,L,D} \right] + \chi_{s=S} L.$$

Our main result establishes that in any optimal mechanism the insurance company gets negative profits from consumers who suffer a liquidity shock and positive profits from those who do not suffer a liquidity shock. Expected profits conditional on the liquidity shock $s = S, NS$ equal

$$\Pi^s \equiv W + I\left[2 - \pi\alpha_H - (1-\pi)\alpha_L\right] - (c_0 + X^s).$$

By zero profits, we must have $l\Pi^S + (1-l)\Pi^{NS} = 0$. We can now prove our main result:

Proposition 6. *In any equilibrium contract, the insurance company gets negative profits from consumers who suffer a liquidity shock and positive profits from those who do not suffer a liquidity shock:*

$$\Pi^S \leq 0 \leq \Pi^{NS}. \tag{54}$$

Proof. Suppose we have an initial contract in which the firm profits from the liquidity shock in period 1 (that is, inequality 54 does not hold). Then, by the definition of Π^s, we must have that the total expenditure conditional on $s = NS$ exceeds the one conditional on $s = S$: $X^{NS} > X^S$.

Consider the alternative contract that allocates the same consumption at $t = 0$ as the original one but implements the best possible renegotiated contract at $t = 1$ conditional on the liquidity shock. More precisely, consumption in subsequent periods is defined by the solution to

$$\max_{\left(c_1^s,c_2^{s,h},c_3^{s,h,d}\right)_{h=H,L,\,d=A,D}} u_A\left(c_1^s\right)+\pi\left[u_A\left(c_2^{s,H}\right)+\left(1-\alpha_H\right)u_A\left(c_3^{s,H,A}\right)+\alpha_H u_D\left(c_3^{s,H,D}\right)\right] \tag{55}$$

$$+\left(1-\pi\right)\left[u_A\left(c_2^{s,L}\right)+\left(1-\alpha_L\right)u_A\left(c_3^{s,L,A}\right)+\alpha_L u_D\left(c_3^{s,L,D}\right)\right]$$

subject to

$$\left\{\begin{array}{c} c_1^s+\pi\left[c_2^{s,H}+\left(1-\alpha_H\right)c_3^{s,H,A}+\alpha_H c_3^{s,H,D}\right] \\ +\left(1-\pi\right)\left[c_2^{s,L}+\left(1-\alpha_L\right)c_3^{s,L,A}+\alpha_L c_3^{s,L,D}\right] \end{array}\right\}\leq I\left[2-\pi\alpha_H-\left(1-\pi\right)\alpha_L\right]-\chi_{s=S}L,$$

$$u_A\left(c_2^{s,h}\right)+\left(1-\alpha_h\right)u_A\left(c_3^{A,s,h}\right)+\alpha_h u_D\left(c_3^{D,s,h}\right)\geq \tag{56}$$

$$\max_{\{\hat{c}\}}\left\{\begin{array}{c} u_A\left(\hat{c}_2\right)+\left(1-\alpha_h\right)u_A\left(\hat{c}_3\right)+\alpha_h u_D\left(\hat{c}_3\right) \\ \text{s.t. } \hat{c}_2+\left(1-\alpha_h\right)\hat{c}_3+\alpha_h\hat{c}_3=\left(2-\alpha_h\right)I \end{array}\right\},\ h=L,H.$$

By construction, this new contract satisfies the non-reneging and incentive-compatibility constraints. We claim that the solution entails full insurance conditional on the shock: $u_A'\left(c_2^{s,h}\right)=u_A'\left(c_3^{A,NS,h}\right)=u_D'\left(c_3^{D,s,h}\right)$ for all s,h (starting from any point in which this is not satisfied, we can always increase the objective function while still satisfying both the zero-profit condition and the non-reneging constraints by moving towards full insurance). Let $C^{s,h}\equiv c_2^{s,h}+\left(1-\alpha_h\right)c_3^{A,s,h}+\alpha_h c_3^{D,s,h}$ denote the total expected consumption at periods 2 and 3. Then, $c_2^{s,h}$ and $c_3^{d,s,h}$ maximize expected utility in period 2 conditional on the shocks s,h given the total expected resources:

$$u_A\left(c_2^{s,h}\right)+\left(1-\alpha_h\right)u_A\left(c_3^{A,s,h}\right)+\alpha_h u_D\left(c_3^{D,s,h}\right)=\max_{c,c^A,c^D}\left\{\begin{array}{c} u(c)+\left(1-\alpha_h\right)u_A\left(c^A\right)+\alpha_h u_D\left(c^D\right) \\ \text{s.t. } c+\left(1-\alpha_h\right)c^A+\alpha_h c^D\leq C^{s,h} \end{array}\right\}$$

$$=\max_{c^A,c^D}\left\{\begin{array}{c} \left(2-\alpha_h\right)u_A\left(c^A\right)+\alpha_h u_D\left(c^D\right) \\ \text{s.t. } \left(2-\alpha_h\right)c^A+\alpha_h c^D\leq C^{s,h} \end{array}\right\}=U_h\left(C^{s,h}\right).$$

The non-reneging constraints (56) can be written as

$$U_h\left(C^{s,h}\right)\geq U_h\left(\left(2-\alpha_h\right)I\right),\ \ h=L,H.$$

Using the fact that U_h is strictly increasing, they can be further simplified to

$$\left(2-\alpha_h\right)c_3^{A,s,h}+\alpha_h c_3^{D,s,h}\geq\left(2-\alpha_h\right)I,\ \ h=L,H.$$

With these simplifications, we can rewrite Program (55) as

$$\max_{c_1^s, C^{s,H}, C^{s,L}} u_A\left(c_1^s\right) + \pi U(C^{s,H}) + (1-\pi)\, U(C^{s,L})$$

subject to

$$c_1^s + \pi C^{s,H} + (1-\pi)\, C^{s,L} \leq I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right] - \chi_{s=S} L,$$

$$(2-\alpha_H)\, I \leq C^{s,H},$$

$$(2-\alpha_L)\, I \leq C^{s,L}.$$

By equation (53), this expression corresponds to $\mathscr{U}\left(I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right] - \chi_{s=S} L\right)$.

The consumer's expected utility from this new contract (at time 0) equals

$$u(c_0) + l\mathscr{U}\left(I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right] - L\right) + (1-l)\,\mathscr{U}\left(I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right]\right). \tag{57}$$

The utility that the consumer attains with the original contract is bounded above by the contract that provides full insurance conditional on the amount of resources that the firm gets at each state in period 1: X^S and X^{NS} (note that this is an upper bound since we do not check for incentive-compatibility or non-reneging constraints). That is, the utility under the original contract is bounded above by

$$u(c_0) + l\mathscr{U}\left(X^S - L\right) + (1-l)\,\mathscr{U}\left(X^{NS}\right). \tag{58}$$

By zero profits, the expected expenditure in the original and the new contracts are the same. Moreover, because $X^S < I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right]$, it follows that the lottery $\{X^S - L, l; X^{NS}, 1-l\}$ is a mean-preserving spread of the lottery

$$\{I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right] - L, l; I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right], 1-l\}.$$

Thus, strict concavity of $\mathscr{U}$ yields:

$$l\mathscr{U}\left(X^S - L\right) + (1-l)\,\mathscr{U}\left(X^{NS}\right) <$$

$$l\mathscr{U}\left(I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right] - L\right) + (1-l)\,\mathscr{U}\left(I\left[2 - \pi\alpha_H - (1-\pi)\,\alpha_L\right]\right).$$

Adding $u(c_0)$ to both sides and comparing with expressions (57) and (58), it follows that the consumer's expected utility under the new contract exceed his expected utility under the original contract, thereby contradicting the optimality of the original contract. □

Therefore, in any equilibrium, firms cannot profit from consumers who suffer a liquidity shock and cannot lose money from those that do not.

Appendix D: Additional Empirical Evidence

HRS Data

This appendix reports the relationship between lapsing and health shocks using the data from the Health and Retirement Study (HRS) (various years). The construction of Table 4 follows the steps outlined in Fang and Kung (2012, Table 6) for the 1994 and 1996 longitudinal waves. We also included respondents in the more recent 2012 wave. As in Fang and Kung, Table 4 shows that there is a positive and statistically insignificant relationship between lapses and either the number of health conditions (Conditions) and changes in health conditions (ΔConditions). Therefore, individuals who lapse are not healthier than those who maintain coverage and they do not appear to lapse after positive health shocks, which is not in line with a reclassification risk channel.

Evidence of Lapsing in Health and Retirement Survey

	Logit Regression		Logit Marginal Effect	
Variable	*Coefficient*	*SE*	*Coefficient*	*SE*
Constant	1.73	0.987	na	na
Age	0.012	0.007	0.001	0.000
Logincome	-0.406***	0.046	-0.032***	0.003
Number of health conditions	0.008	0.026	0.007	0.002
Married	0.162	0.250	0.012	0.019
Has children	-0.775	0.657	-0.061	0.052
Age of youngest child	-0.004	0.005	0.000	0.000
ΔAge	0.202	0.327	0.016	0.026
(ΔAge)^2	-0.011	0.078	-0.001	0.006
ΔLogincome	0.076	0.051	0.006	0.004
(ΔLogincome)^2	-0.003	0.014	0.000	0.001
ΔConditions	0.106	0.139	0.008	0.011
(ΔConditions)^2	0.013	0.075	0.001	0.006
ΔMarried	-0.299	0.335	-0.023	0.026

Table 4: Logistic Regression using 1992 - 2012 RAND HRS Longitudinal (13,137 Observations) with dependent variable equal to maintaining (0) / lapsing (1) coverage. ***, **, and * represent significance at 0.001, 0.01, and 0.05, respectively.

MetLife and SBLI Data

As we examine in Subsection 5.2, in the presence of fixed costs that are proportional to coverage, such as sales commissions or illiquidity premiums, even consumers with rational expectations may demand policies with surrender fees. With commissions, one should not observe different surrender fees after all commissions have been paid, which typically happens in the first two or three years of the policy.

As described in Subsection 5.2, with an explanation based on the cost of liquidating investments, the optimal insurance policy sets a surrender fee that balances the higher return that can be obtained by facing a more predictable pattern of lapses against the cost that policyholders face when they are unable to smooth consumption after a liquidity shock. Let the liquidity premium be the proportion of the value of an investment that has to be given up in case of early liquidation. Since the (ex-ante) expected cost to policyholders of being unable to smooth consumption is increasing in the probability of a liquidity shock, this explanation predicts that, holding the liquidity premium constant, surrender fees (as a proportion of the amount invested) should *decrease* in the probability of facing a liquidity shock.

In order to evaluate the relationship between surrender fees and the probability of liquidity shocks, we hand-collected detailed whole life insurance policy offers from two national insurance companies, MetLife and SBLI, at the Lifequotes (2013) website.[53] MetLife and SBLI are two national life insurers with operations in most of the 50 states. MetLife is the largest U.S. life insurer with over $2 trillion in total life insurance coverage in force while SBLI is middle sized with $125 billion of coverage in force, thereby allowing us to ensure that premium data was not driven by idiosyncratic features associated with firm size. We collected policy information across both genders across with the following coverage amounts: $100,000; $250,000; $500,000; $750,000 and $1,000,000. We chose ages between 20 and 70 in 5-year increments and both genders. We focused on traditional whole life policies since future cash surrender values do not depend on the return of the insurer's portfolio.[54] MetLife policies mature at age 120, whereas SBLI policies mature at age 121. All policies assume no tobacco or nicotine use and excellent health ("preferred plus"). Premiums are annual, which is the most common frequency. An automation tool was used to effectively eliminate human coding error. For each policy, we obtained the cash surrender values for each of the 25 years after purchase.

Our data set covers all American States except for New York[55], where the companies did not offer these policies, for a total of 10,738 policies. MetLife offered policies for all coverage amounts noted above, for a total of 5,390 policies (2,695 per gender). SBLI did not offer policies with $100,000 coverage for individuals aged 60 and older in the states of Alabama, Alaska, Idaho, Minnesota, Montana, Nebraska, North Dakota, and Washington. In total, these missing data add up to 42 policies (21 per gender). Our results remain if we exclude these states from the sample. In sum, our data set consists of 5,348 policies (2,674 per gender) SBLI policies.

The surrender fee for each policy corresponds to the proportion of the discounted sum of insurance loads (i.e., present value of premiums paid in excess of the actuarially-fair price) that cannot be recovered as cash surrender value. Thus, the surrender fee is the fraction of pre-paid premiums that cannot be recovered if the policy is surrendered. To ensure the comparability of the policies, we kept the terms of each policy constant except for our controls (coverage, ages, and genders). We, therefore, focused on

[53]The choice of these two firms was dictated by data availability. Whole-life policies are typically used differently than Universal Life (UL) policies, as UL policies are often used a tax-preferred investment vehicle in addition to insurance.

[54]Unlike traditional whole policies, most universal life insurance policies only provide an estimate of future cash surrender values.

[55]Unfortunately, these two companies did not sell this type of policies in the state New York. SBLI does not operate in New York. MetLife whole policies in New York are issued separately from the ones in other states. In order to verify the robustness of our findings to other companies, we also collected data from other insurance firms for the state of California and obtained the same results.

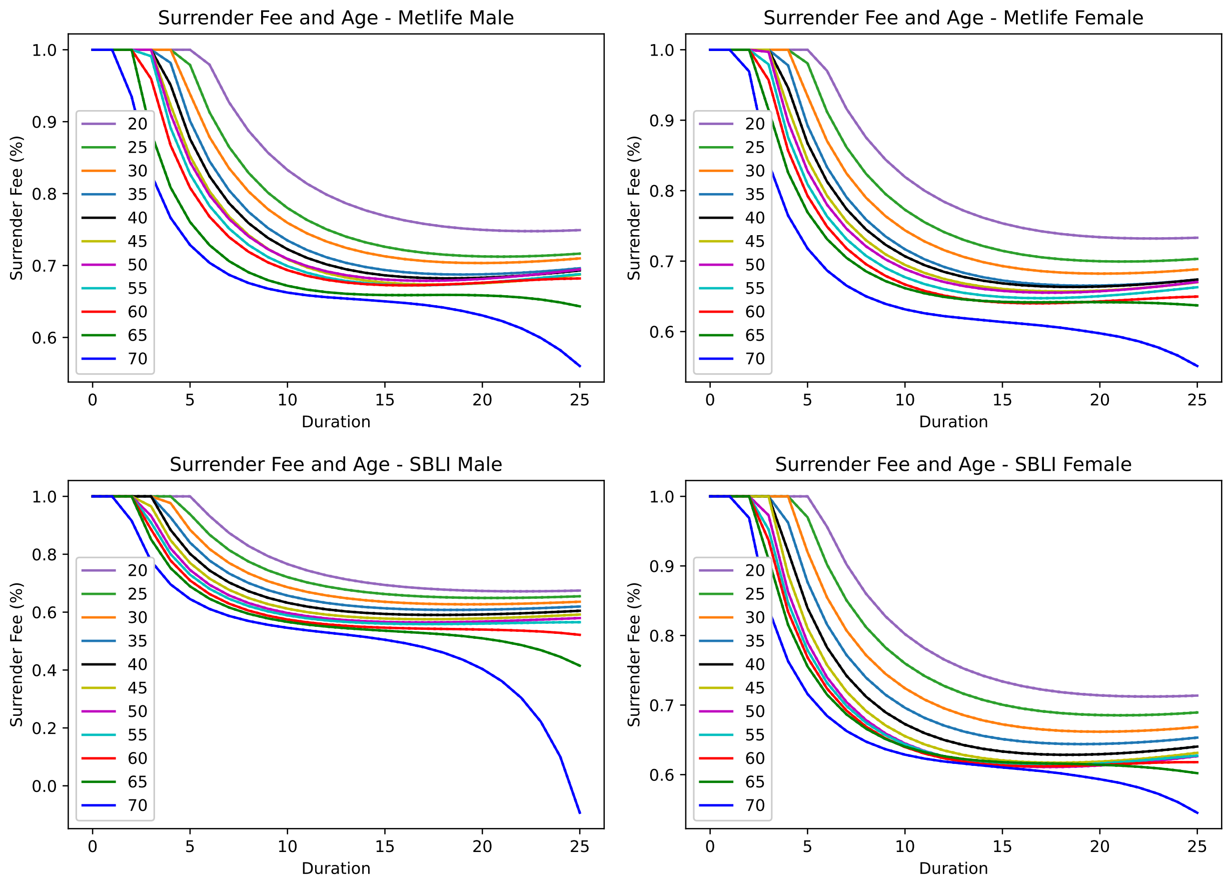


Figure 6: Mean surrender fees by policy duration for each age (different color line) and their 95% confidence intervals.

policies for the "preferred plus" health category that require a health exam.

To test the prediction that surrender fees should increase in the probability of liquidity shocks, we need observable measures of the probability of liquidity shocks. Since we have detailed policy data but no administrative information about the individuals who buy each policy, we need to proxy for the probability of liquidity shocks using the terms of the policies. We use two different proxies: age and coverage. It is widely documented that younger individuals are more likely to be liquidity constrained, and age is a frequently-used proxy for the presence of liquidity constraints.[56] In fact, consistently with these proxies, lapse rates are decreasing in both age and coverage (Section 2).

Figure 6 shows the mean surrender fees as a function of policy duration at each age along with their associated 95% confidence intervals.Because whole life policies do not have a cash surrender value during the first few years after purchasing, surrender fees start at 100% for each age. As policies mature, they accumulate cash value, reducing the surrender fee. Our main interest, however, is in the difference in surrender fees for policies sold to individuals of different ages. For both MetLife and SBLI policies, notice that the surrender fees decrease in age, at each duration. Thus, consistently with the differential

[56]See, for example, Jappelli (1990), Jappelli, Pischke, and Souleles (1998), Besley, Meads, and Surico (2010), and Zhang (2014).

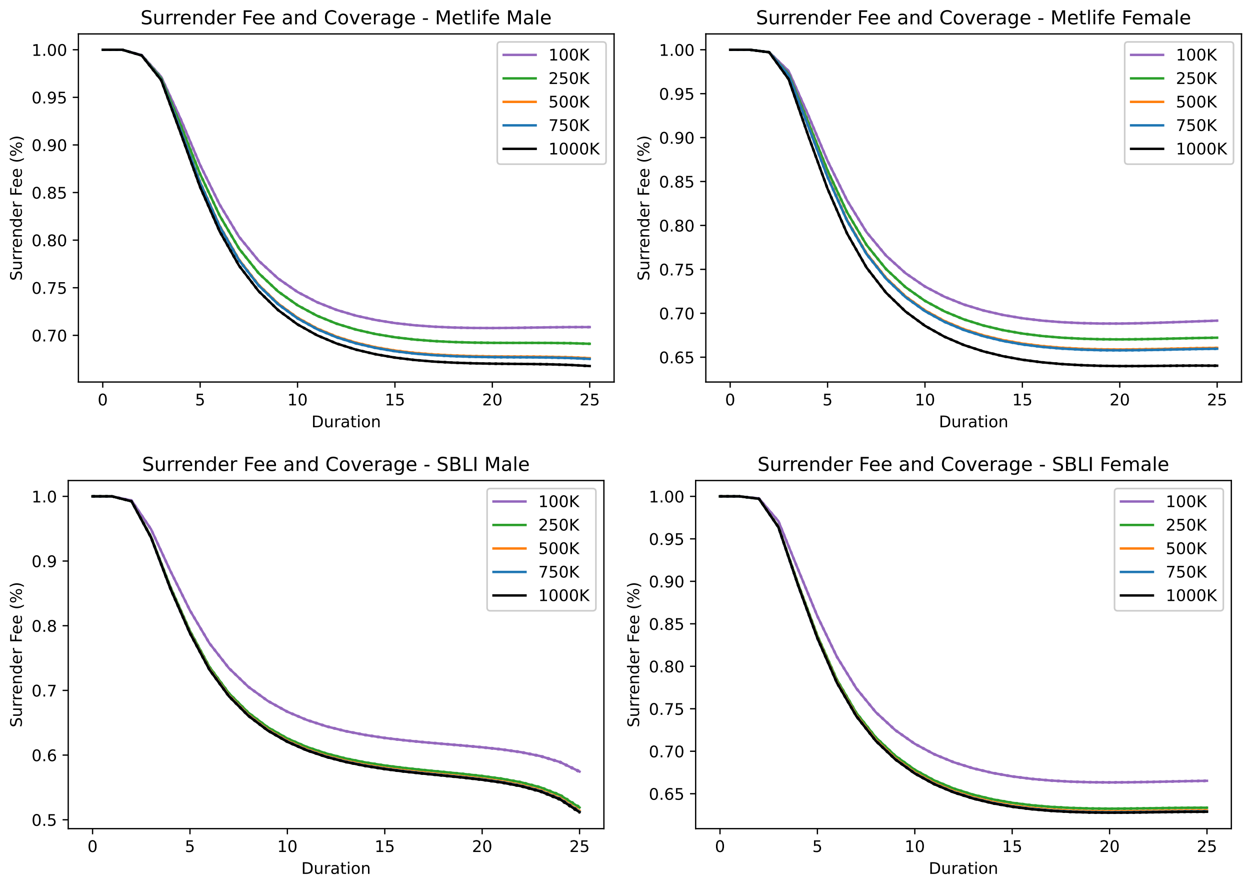


Figure 7: Mean surrender fees by policy duration for each face amount and their 95% confidence intervals.

attention and forgetfulness models, younger individuals face higher surrender fees. The differences by age are not only statistically significant; they are also economically large. For example, while a 20-year-old policyholder who surrenders after 5 years would not collect any cash value, a 70 year old collects about 30% of the amount paid in excess of the actuarially-fair prices. Figure 7 shows the mean surrender fees for different *coverage* amounts and their associated 95% confidence intervals. In contrast to the prediction of the rational model, surrender fees for both MetLife and SBLI policies are decreasing in coverage. However, while for MetLife the difference is always statistically significant, for SBLI policies with coverage above $100,000 are not statistically significant at 5% level.[57] Relative to age considered above, differences by coverage levels are slightly smaller. Nevertheless, the surrender fee on a $100,000 policy is, on average, between 5 and 10 percentage points larger than the surrender fee on a $1,000,000 policy.

[57]The lack of statistical significance for SBLI policies with more than $100,000 coverage could be due to the fact that, while lapse probabilities are much higher for smaller policies, the difference is not very large for policies with coverage above $200,000 (see Figure 2).

Compulife

Data on insurance policy quotes, current as of February 2013, were obtained from Compulife (2013). We gathered quotes for a $500,000 policy with a 20 year term for a male age 35, non-smoker, and a preferred-plus rating class. For the mortality table, we use the 2008 Valuation Basic Table (VBT) computed by the Society of Actuaries that captures the "insured lives mortality" based on the insured population. For Figure 3 in the main text, we assume a nominal interest rate of 6.5%. However, the results are very robust to the interest rate. The figures below repeat the exercise under extreme assumptions about the nominal interest rate and the inflation rate. Note that the only cost in actuarial profits is the death benefit. To obtain economic profits, one should subtract all other costs.

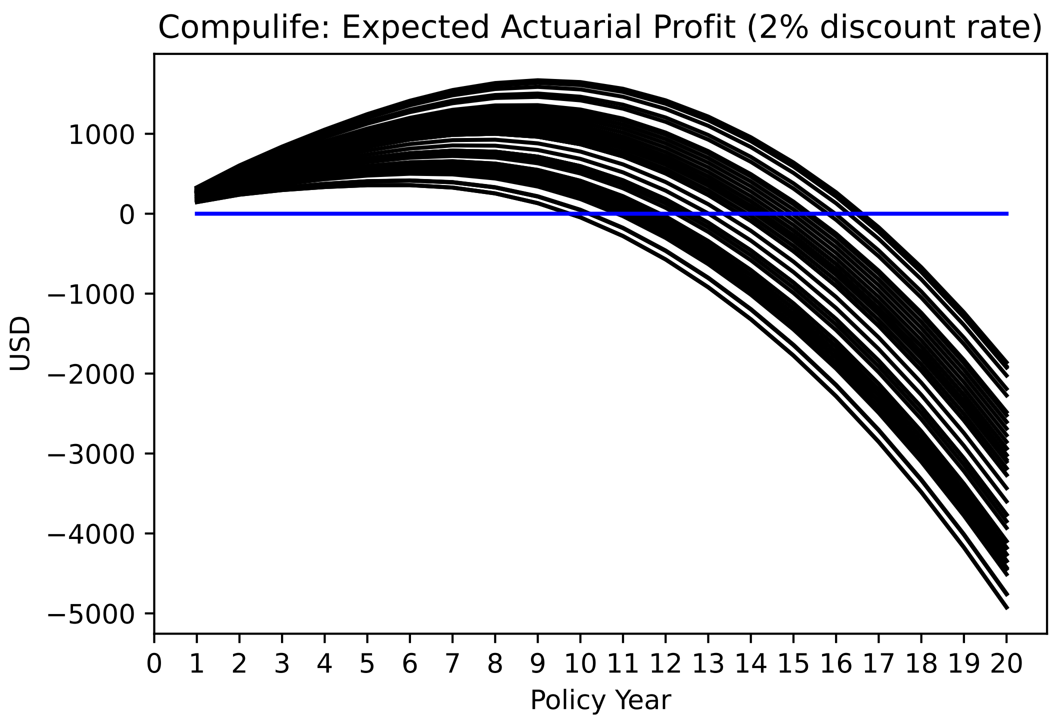


Figure 8: Insurer's profits if the consumer plans to hold policy for after N years under 2% nominal interest rate.

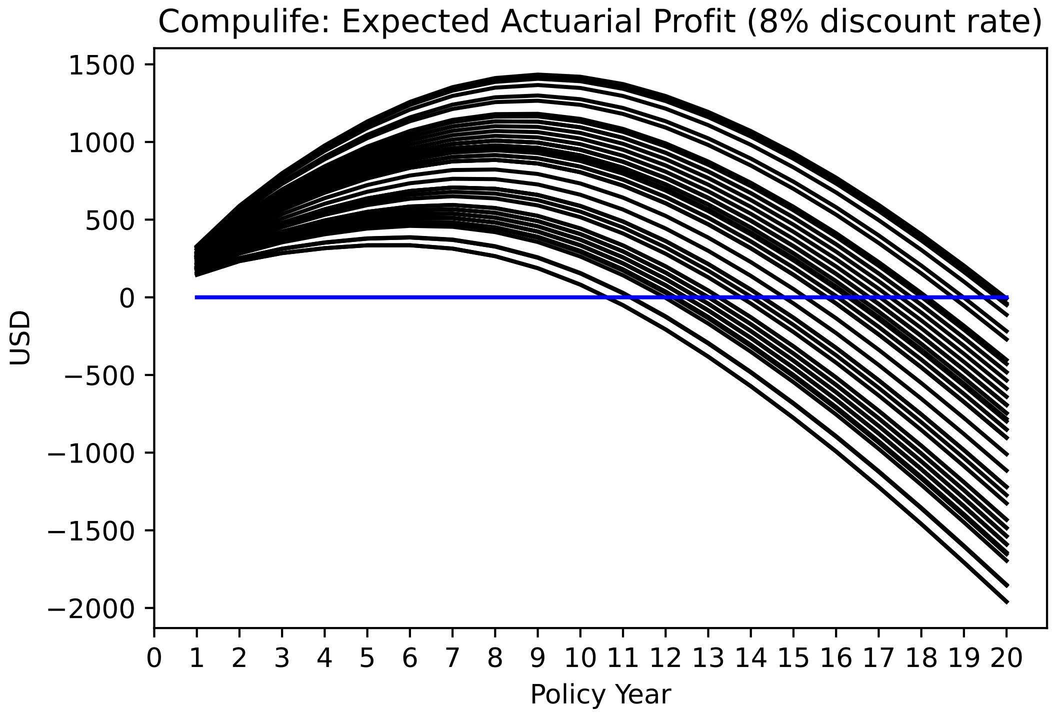


Figure 9: Insurer's profits if the consumer plans to hold policy for after N years under 8% nominal interest rate.

Large Life Insurer (LLI) Color-Coded Data

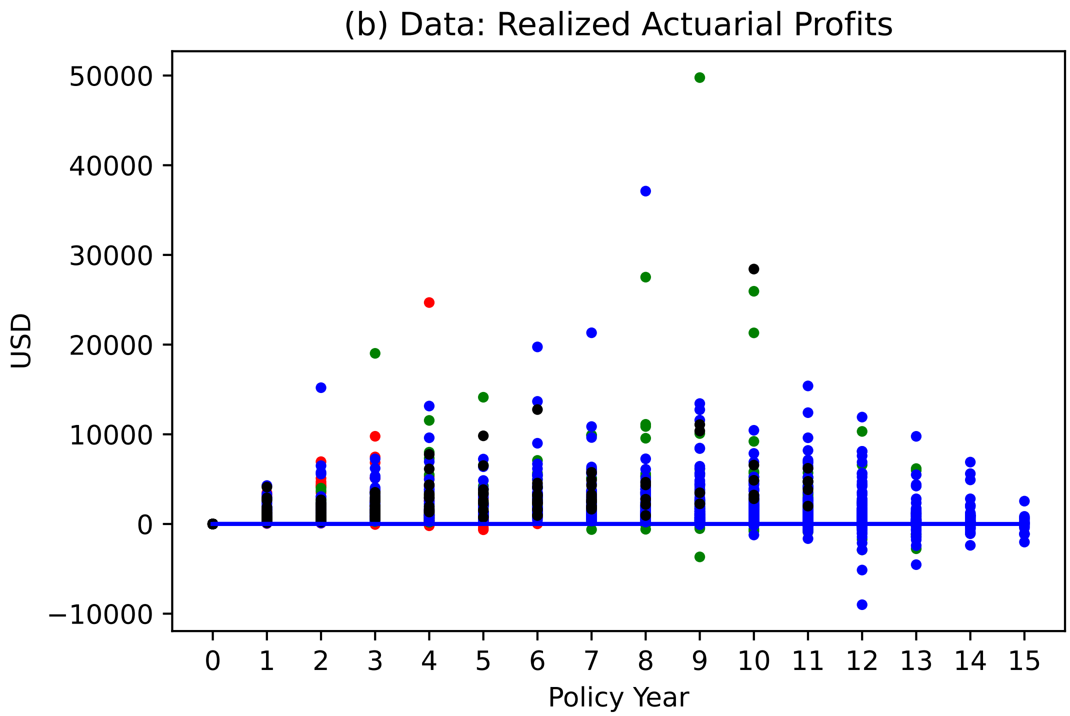


Figure 10: Figure 3(b) with color coding of policies: 10-year (red), 15-year (green), 20-year (blue) and 30-year (black).