Policyholder Exercise Behavior in Life Insurance: The State of Affairs*

Daniel Bauer, Jin Gao, Thorsten Moenig, Eric R. Ulm, Nan Zhu†

March 2015

Abstract

The paper presents a review of structural models of policyholder behavior in life insurance. We first discuss underlying drivers of policyholder behavior in theory and survey the implications of different models. We then turn to empirical behavior and appraise how well different drivers explain observations. The key contributions lie in the synthesis and the systematic categorization of different approaches. The paper should provide a foundation for future studies, and we describe some important directions for future research in the conclusion.

Keywords: Optimal Exercise Behavior. Frictions. Guaranteed Minimum Benefits. Life Settlements.

---

*This paper concludes our research under the Society of Actuaries’ (SOA) Center of Actuarial Excellence (CAE) Research Grant “Structural Models of Policyholder Behavior.” As such, this paper disproportionally surveys our own work, although we also include many references to related research. We are grateful to the SOA for the financial support.

†Bauer (corresponding author, email: dbauer@gsu.edu, phone: +1-(404)-413-7490) and Ulm are with the Department of Risk Management and Insurance at Georgia State University (CAE); Gao is with the Department of Finance and Insurance at Lingnan University; Moenig is with the Department of Mathematics at the University of St. Thomas (CAE); and Zhu is with the Department of Mathematics at Illinois State University (CAE).
1 Introduction

Pricing and risk management for most life insurance products, in one way or another, depend on policyholders’ future behavior. Examples include lapsing or surrendering term or whole life insurance; stopping payment of premiums or annuitizing benefits in participating or universal life policies; and withdrawing or transferring funds in Variable Annuities (VAs) with Guaranteed Minimum Benefits. Therefore, it is important that corresponding actuarial models accurately describe policyholders’ future actions—and policyholders deviating from the prescribed behavior presents a significant risk factor in selling these products that risk managers should consider.

In this paper, we discuss structural models of policyholder behavior that explicitly model the decision process, where we emphasize implications for practicing actuaries for pricing and risk management. A structural approach to policyholder behavior—rather than relying on historical data in order to build empirical models for predicting policyholder exercise as it is frequently done in practice—is important for at least three reasons.

First, an empirical approach will prove difficult for a newly introduced product line. Consider, for instance, a new generation of Guaranteed Living Benefits (GLBs) in VAs such as Guaranteed Minimum Withdrawal Benefits (GMWBs) introduced in the early 2000s or Guaranteed Lifetime Withdrawal Benefits (GLWBs) introduced early in this decade. When offering such new products while having available no or only a few years of observed withdrawal behavior, it is up to the actuary to make a reasonable and prudent assumption. But what is reasonable or prudent in this context?

Second, when regulatory or economic circumstances change, relying on historical data may be deceptive. For instance, a rise in market interest rates in the 1970s resulted in the so-called disintermediation in the whole life market with drastically more surrenders and policy loans (Black and Skipper, 2000, p. 111); also, misjudgment of exercise of Guaranteed Annuity Options (GAOs) in the face of falling interest rates contributed to the demise of the UK-based life insurer Equitable Life in 2000 (Boyle and Hardy, 2003). Hence, it is necessary to have an understanding of what drives these empirical exercise rules—and, particularly, under which circumstances they may fail.
And, third, to accurately appraise the risk of a systematic change in policyholder behavior—for instance, due to improved education by financial advisors—one needs to understand what policyholders should optimally do from their perspective as well as what is the worst case scenario from the insurer’s point of view. As we will describe in more detail, these vantage points may differ and answering both may require different structural models. This has important implications for the insurer’s risk exposure, and structural models can help insurers justify to the regulator less prudent (but more accurate) assumptions for reserving, thereby reducing its cost of capital. Conversely, regulators will benefit from understanding policyholder behavior, for instance in view of establishing uniform modeling requirements for insurers across the board.

We commence in Section 2 by describing the drivers for optimal policyholder behavior identified in the literature. Of course, the impact of policyholder exercise on the value of the insurance contract is a major factor in explaining policyholder exercise behavior, although we argue that in a world with frictions there are other aspects that affect how they behave: Taxes, preferences, etc. Equipped with the insight of what may drive policyholder exercise and how these aspects affect behavior, we go on in Section 3 with attempts to explain empirical patterns for different product categories. In particular, we connect to the empirical literature as well as so-called dynamic functions describing policyholder behavior, which are based on empirical exercise patterns and used by most companies. Here we emphasize that value-maximization alone does not rationalize various features, but these can be explained by considering said frictions as well as behavioral aspects. Finally, Section 4 concludes.
2 What Potentially Drives Policyholder Behavior? A Review of Theory

Value-Maximization

Consider the most basic situation, namely that of a complete and frictionless market for life insurance. In this case, assessing policyholder exercise behavior is straightforward in principle. Each agent would be able to replicate every possible cash flow using underlying (Arrow) securities, so that (optimal) policyholder behavior would be fully determined by value maximization, where a unique valuation is implied by the assumption of no-arbitrage (Duffie, 2010). Deviating from a value-maximizing strategy is not opportune since any consumption menu can be purchased.

This is not to say that actually determining the optimal strategies within a risk-neutral valuation framework is trivial. It may require the solution of optimal control problems akin to the valuation of American or Bermudan options, and a great number of contributions in actuarial science have taken this approach to evaluate various types of contracts (Milevsky and Salisbury, 2006; Ulm, 2006; Bauer et al., 2008; Chen and Forsyth, 2008; Dai et al., 2008; Bauer et al., 2010; Bacinello et al., 2011, among many others).

The value of the contract, thus, presents a primary driver for policyholder exercise. However, it does not appear to be sufficient: When comparing the derived optimal behavior to empirical patterns or resulting values to market prices, one frequently finds a significant dissonance. For instance, it would typically not be optimal to lapse a front-loaded term life insurance unless there is a substantial change in the economic environment, yet lapsation is common and considered in pricing all basic life insurance contracts. Similarly, discrepancies between calculations in a value-maximizing model and market prices have been pointed out for GLBs in VAs including GMWBs (Milevsky and Salisbury, 2006; Bauer et al., 2008; Chen et al., 2008), GLWBs (Piscopo, 2010), and Guaranteed Minimum Income Benefits (Marshall et al., 2010); as well as for surrender guarantees in participating products (Grosen and Jørgensen, 2000; Bauer et al., 2006).
The reason for this dissonance is that the life insurance market is neither frictionless nor complete. Consumers may face borrowing constraints or different borrowing and saving rates. Insurance contracts entail transaction costs as well as differential tax treatment. Policyholders face trading constraints, as typically there is no liquid secondary market for “used” life insurance policies. The insurance market is incomplete in the sense that the payoff depends on the policyholder’s survival and not all payoff profiles may be attained via existing securities. The information set of the insurance company and its customers may differ, giving rise to potential informational frictions. And, finally, policyholders may not make perfectly rational decisions and may be subject to behavioral biases—although the latter point has to be considered with care as it is all too enticing to point to “irrationality” for explaining exercise patterns (and some authors have).

Much recent research on policyholder behavior is concerned with the question of how these various frictions affect policyholder behavior and, particularly, of how to adjust the conventional value maximization framework to account for them.\(^1\) However, before heading down this path, it is worth pointing out that although a basic value-maximizing approach may fail at aligning theory with observations, this approach can be important for risk management. More precisely, the approach identifies the worst-case scenario from the insurer’s point of view that is robust to any exercise strategy, even lucky or prescient ones (Bauer et al., 2010, 2013). Thus, one way to account for the risk associated with policyholder behavior is to (i) determine the value associated with value-maximizing behavior and put up the difference to the market price as a policyholder behavior risk reserve; and (ii) manage embedded risk as if policyholders behaved as value maximizers. If policyholders, as expected, deviate from the value-maximizing behavior, this strategy will result in a surplus, the risk-adjusted expected present value of which—adjusted for potential capital charges—should equal exactly the risk reserve. Nonetheless, in order to understand how policyholder actually will behave, we consider situations with frictions in the remainder of the paper.

\(^1\)In what follows, for simplicity, we will also refer to market incompleteness and behavioral biases as frictions, although frequently researchers separate the concepts.
Taxes: Subjective Value Maximization

The inability to sell or repurchase the policy at its fair value (trading restriction) is relevant only if market incompleteness is material. Otherwise, the policyholder may set up a portfolio of underlying securities that replicate or offset the policy cash flow, so that exercise should be driven by value maximization after all.

However, even in this situation, preferred tax treatment of life insurance benefits requires modifying the basic risk-neutral approach since the tax considerations will affect the policyholder’s subjective valuation. This is extremely relevant since tax advantages are a primary reason for the popularity of many savings products offered by insurers—such as VAs (Milevsky and Panyagometh, 2001; Brown and Poterba, 2004). Given that tax treatment is a major driver for the purchase of the products, it is not surprising that it also may be of relevance to how policyholders behave after purchase.

This idea is taken up in Moenig and Bauer (2014a), where the authors show that such a subjective risk-neutral value maximization for GMWBs within VAs yields exercise patterns and prices that are in line with market observations. The key insight is that in a complete pre-tax market, it is possible to replicate any post-tax cash flow with a pre-tax investment in some benchmark securities, leading to a non-linear implicit equation for the subjective value (rather than a linear risk-neutral expected value). More precisely, Moenig and Bauer (2014a) show that the time-$t$ value of a post-tax cash-flow $X$ at time $t + 1$ in the absence of offsetting obligations is given by the equation:

$$V_t = \mathbb{E}_t^Q \left[ e^{-\int_{t+1}^{t+1} r_s ds} X \right] + \frac{\kappa}{1 - \kappa} \mathbb{E}_t^Q \left[ e^{-\int_{t+1}^{t+1} r_s ds} (X - V_t)^+ \right],$$

where $\mathbb{Q}$ is the risk-neutral measure in the complete pre-tax market, $r$ is the risk-free rate, and $\kappa$ is the effective capital gains tax rate. In particular, we recover the usual expected discounted value in case $\kappa = 0$, whereas for $\kappa > 0$ the value increases as it is also necessary to replicate incurred capital gains taxes. Applying this equation to a dynamic model for a VA plus GMWB that includes
an adequate treatment of withdrawals and terminal benefits, the authors show that the resulting optimal behavior can differ dramatically from the consideration without taxes.

As an important consequence, the consideration of taxes yields a difference in the policyholder’s and the insurer’s valuation of the policy cash flows. In particular, Moenig and Bauer (2014a) find that given this dissonance, empirical VA plus GMWB contracts present a worthwhile investment opportunity for the policyholder while at the same time being profitable to the insurer (see Table 1 in the next section).

This implicit wedge between the policyholder’s and the insurer’s valuation can lead to curious results, such as a negative marginal value for a basic return-of-premium Guaranteed Minimum Death Benefit (GMDB) in the presence of a GMWB (Moenig and Bauer, 2014b). The key point is that with a GMDB, policyholders will adjust their behavior in order to maximize their subjective value, net of taxes. This change, however, can yield a smaller total value of all policy cash flows when ignoring corresponding tax benefits, as it is relevant for the insurer. As pointed out in Moenig and Bauer (2014b), this may explain why a return-of-premium GMDB is included as a standard feature in most VA products.

**Incompleteness: The Impact of Preferences and Idiosyncratic Risks**

When trading restrictions and market incompleteness are important—although, of course, assessing this qualification is not trivial—the conventional approach is to build life-cycle optimization models that consider the policyholder’s insurance decision problem in a portfolio context, following early work by Fischer (1973) and Richard (1975). In addition to the contract’s value, in this setting we obtain a number of additional dimensions that influence behavior, particularly the level of risk aversion, subjective discount rates, the strength of the motive for leaving bequests in the case of death, and interactions with other relevant risk factors.

For instance, Gao and Ulm (2012) show that allocation decisions in a VA with a GMDB will be driven by the appreciation of additional consumption by the policyholders and their heirs—given that the latter group is protected by the GMDB. They emphasize this “argument” between the ben-
eficiaries and the insured as a driver for optimal investment: Due to the additional protection, the beneficiaries—assuming they exhibit the same level of risk aversion—will prefer a more aggressive strategy, and the strength of the disagreement depends on the constellation of the policy parameters. We also refer to Steinorth and Mitchell (2012) for a similar analysis of withdrawal behavior in VAs with GLWBs and to Eling and Kochanski (2013) for papers considering lapse/surrender behavior with regards to other product categories in life-cycle frameworks. However, the relevance and the effect of these additional dimensions associated with the policyholder’s preferences (risk aversion, discounting, wealth, and bequest motive) crucially depend on the model framework, and—as pointed out by Campbell (2006)—capturing all relevant aspects and risk factors within a life-cycle model is an ambitious task.

One important aspect is the universe of available financial instruments. For instance, Gao and Ulm (2015) show that the presence of labor income and the availability of term life insurance dramatically affects policyholder behavior and take-up in VAs with a GMDB rider. More precisely, they show that labor income dramatically increases the wedge between policyholder and beneficiaries since it is only earned when the policyholder survives, and therefore augments the “argument” between them—yielding a considerable change in the optimal allocation rule. Furthermore, they show that a simple term-life insurance is a satisfactory substitute for the GMDB contract, so that policyholders are willing to pay very little for the GMDB. Indeed, their analysis suggests that if fairly priced insurance is available, consumers optimally would not choose to purchase the death benefit rider in their model.

Similarly, Bauer and Moenig (2015) show that the presence of an outside savings opportunity considerably affects policyholder withdrawal behavior for a VA with GMWB, and that the optimal withdrawal strategy closely resembles that under subjective risk-neutral valuation (Horneff et al. (2013) also consider a life-cycle model with outside savings for a GMWB). The intuition is that preferences only matter to the extent that the market is incomplete (Bauer and Moenig, 2015). This insight echoes the basic logic of the so-called martingale approach to optimal control by Cox and Huang (1989, 1991): In a complete market, it is optimal to maximize value since it is
possible to purchase Arrow securities to attain any state-contingent allocation a consumer wishes to realize. Only if the market is sufficiently incomplete will there be a reason to deviate from value-maximizing behavior so that preferences could have an effect. To characterize the level of incompleteness, Bauer and Moenig (2015) contrast state allocations in a (hypothetical) complete market with the corresponding situation based on existing securities, following ideas by Koijen et al. (2014). It appears that while policyholder behavior for GMWBs is mostly driven by value maximization since survival probabilities in the relevant age range are low, policyholder behavior for GLWBs is affected by preferences since this product class changes the universe of investment options in a significant manner—essentially offering insurance coverage against states that combine longevity with adverse market developments.

As a straightforward consequence, access to a secondary market for insurance—such as in the form of life settlements—also has an effect on policyholder behavior as it increases the set of financial possibilities, and therefore potentially completes the market. However, the very existence of this market is linked to the possibility of mortality probabilities changing over the policyholder’s life-cycle, which brings us to the second important aspect: The relevant sources of uncertainty. When solely considering mortality risk, there would be no benefit to committing to long-term life insurance contracts (e.g., the life-cycle models by Fischer (1973); Richard (1975) include optimal one-period insurance contracts). The rationale for optimal long-term, front-loaded contracts arises with the relevance of re-classification risk (Hendel and Lizzeri, 2003), i.e. the possibility of moving to a worse rating class with higher mortality probabilities and having to pay more for insurance. Similarly, the possibility of a changing bequest motive may be a relevant risk factor.

Trivially, in the presence of a secondary insurance market, the immediate impact will be that lapse and surrender rates for conventional whole, term, or universal policies decrease as some policyholders—typically those with sub-par mortality prospects—have the possibility of settling. As a consequence, equilibrium insurance prices should go up, which transfers resources from early in life to late in life (Daily et al., 2008) and may erode possibilities for insuring reclassification risk (Fang and Kung, 2010a), both potentially decreasing consumer utility and thus welfare. In
particular, this is the case when lapsation is driven by bequest shocks. On the other hand, Zhu and Bauer (2011) show that the resources are also transferred from healthy to sick states of the world, which may have positive implications for the insured as the latter may be situations where resources are scarce (see Fang and Kung (2010b) for similar results). In particular, this is the case when lapsation is driven by liquidity needs related to health expenditures. Thus, understanding the drivers for policyholder behavior is necessary to appraise the merit of the life settlements market.

**Asymmetric Information: Adverse Selection and Moral Hazard**

When including policyholders’ willingness to purchase insurance under the umbrella of “policyholder behavior,” there exists an extensive literature on the effects of asymmetric information on policyholder behavior as the latter problem is very well studied in the literature (see, e.g., Dionne et al. (2013), Chiappori and Salanié (2013), Winter (2013), and references therein). One of the most robust predictions under asymmetric information that is frequently used for testing whether there exist information asymmetries in a certain insurance market is that of a “positive correlation of risk and average conditional on all public available information” (Chiappori and Salanié, 2013)—or, in other words, whether consumers that know they face higher risk will purchase more insurance.

Zhu and Bauer (2011, 2013) show that private information on mortality prospects, in a similar manner, also affects how policyholders behave vis-à-vis their decision to lapse, surrender, and/or—in the presence of a secondary market—settle an existing life insurance policy. In particular, they show that if there exists asymmetric information, a life settlement company will have to accommodate this in pricing the “used” policy. More precisely, if the (private) true life expectancy is underestimated by the life settlement company and the offer price is based on this (too short) estimate, a policyholder will be glad to accept this (high) offer. On the other hand, if the (private) true life expectancy is overestimated by the life settlement company and the offer price is based on this (too long) estimate, a policyholder will potentially walk away from this (low) offer. This asymmetry in how under- and overestimating mortality probabilities affect profitability will shift the pricing schedule in equilibrium, leading the life settlement company to offer less than the actuarial
present value for the policies.

Zhu and Bauer (2013) present corresponding pricing formulas in the context of a life-cycle model and show that the effect can be considerable. In particular, they show that asymmetric information will lead to a positive bias in expected return calculations based on the equivalence principle and best-estimate mortality rates—which is typically the benchmark used in practice. Hence, asymmetric information may serve as an explanation for the alleged underperformance in the life settlements market.

Private information can also affect behavior in advanced insurance contracts. As illustrated by Benedetti and Biffis (2013), when the evolution of mortality differs among policyholders but decisions are based on their own mortalities, the design of the contract affects the remaining pool of policyholders due to differences in policyholder behavior—and thus also the aggregate survival probabilities in the pool. In other words, the aggregate mortalities endogenously depend on how policyholders behave even if policyholders are ex-ante homogeneous, and their behavior in turn depends on the contract features. For instance, the authors show that for a VA with GMDB, over the course of the contract mortality probabilities will exceed aggregate population rates as policyholders with low (private) realizations will surrender their contracts. The resulting (endogenous) adjustments on population mortality depend on the contract parameters in a non-trivial manner.

**Beyond Rationality: Behavioral Aspects**

Beyond factors that could be rationalized, policyholder behavior may be affected by psychological, cognitive, or emotional factors—which is the central theme of the emerging field of *behavioral economics*. These behavioral mechanisms include *heuristics*, i.e. that individuals follow simple “rules of thumb” and/or focus on a single aspect of a complex problem; and *framing*, i.e. that individuals perceive a situation based on its presentation; among others (Shefrin, 2002). However, as a word of caution, while it is tempting to attribute certain behavior that is difficult to explain at first sight to “irrationality,” as detailed above, frictions can lead to complex exercise patterns even when policyholders are rational. Furthermore, it is important to distinguish between *exogenous* factors
outside of a given model such as idiosyncratic liquidity shocks, e.g. due to personal tragedy, on the one side, and *behavioral* aspects, which make individuals *systematically* deviate from optimal choices due to some psychological or neurological process, on the other side.

There have been several recent contributions that emphasize the importance of behavioral concepts in insurance and risk management focusing on the impact of different assumptions about preferences (Harrison and Martinez-Correa, 2012), the role of theory versus experiments (Richter et al., 2014), and implications for insurance regulation (Kunreuther and Pauly, 2014). We refer to these papers for more detailed reviews.

In view of the specific problem of policyholder behavior, Mulholland and Finke (2014) hypothesize that cognitive aspects influence policy lapsation. In particular, they argue that lapsing a policy is an important financial decision, and it has been demonstrated in other research that cognitive ability is positively related to sound financial decision-making—although a possible driver could be “information constraints” rather than preferential or psychological effects (Christelis et al., 2010).

Gottlieb and Smetters (2014) present a utility model, in which consumers exhibit *differential attention* when making life insurance decisions. More precisely, they overstate the risk of dying relative to other risk potentially leading to liquidity shocks. This differential attention could be due to *narrow framing*, i.e. consumers may think about risks in isolation and do not merge the consideration with the broader set of risks they are facing; another reason could be the so-called *disjunction fallacy* stating that consumers tend to attach inconsistent probability weightings to combined hypotheses (Costello, 2009). In any case, underweighting other risk factors will lead policyholders to lapse excessively relative to a model of rational insurance purchasing and lapsation.
3 What Actually Drives Policyholder Behavior? Empirical Evidence

Equipped with the insights of what may drive policyholder behavior in theory from the previous section, we now turn to the question of which aspects seem to explain actually observed—or empirical—behavior. We separate the discussion by different product categories. In particular, we start by analyzing lapsation and surrender in conventional (term, whole, or universal) policies before we consider behavior for more advanced, investment linked policies—especially VA contracts.

Lapsation, Surrender, and Settlement in Conventional Policies

Lapsation and surrender are extremely prevalent in conventional term and whole life policies. For instance, Figure 1 shows the total individual life insurance policy lapse rates in percent by policy year taken from SOA and LIMRA (2012) based on observation years 2007-2009. According to these lapse rates, conditional on the policyholder surviving, only slightly over 35% of all policies are active after 20 policy years and a mere 28% make it beyond policy year 30. We refer to SOA and LIMRA (2012) and more recent SOA/LIMRA life insurance persistency studies for details on how lapse rates differ by policyholder characteristics.

Eling and Kochanski (2013) survey research on life insurance lapsation and review more than 50 theoretical and empirical contributions. On the empirical side, literature has formed a number of hypotheses as to what factors drive lapsation. More precisely, according to the so-called interest rate hypothesis (IRH), policyholders lapse in response to changes in interest rates; the related and more recent policy replacement hypothesis (PRH) presumes that policies are lapsed with the intention to purchase another insurance contract as a replacement; and the so-called emergency fund hypothesis (EFH) contemplates that policyholders predominantly lapse to meet unexpected funding requirements.

Lapse occurs when the policyholder stops paying premiums and/or actively cancels the policy. Whether the policyholder is eligible for a cash benefit upon surrender depends on the policy characteristics. Minimum cash surrender values are regulated in the U.S. by the standard nonforfeit law for life insurance.
Figure 1: Total individual life insurance policy lapse rates in percent by policy year. Source: SOA and LIMRA (2012).

Note that here the IRH and the PRH are linked to value maximization as the driver for policyholder behavior. If interest rates increase, insurance will become cheaper (according to e.g. an early result by Lidstone (1905)) so that policyholders may be incentivized to lapse existing contracts and potentially purchase new coverage—although surrender values may entail considerable markdowns relative to the market reserves. The EFH, on the other hand, is primarily related to market incompleteness: Funds are required due to shocks that are not or only partially insured.

Eling and Kochanski (2013) point out that early empirical studies on lapsation based on aggregate industry data find more support for the IRH over the EFH, although other factors also appear relevant in the lapse decision including company characteristics. However, one important aspect seems to be that aggregate data have some limitations in view of testing the EFH. In contrast, a recent set of studies make use of household-level panel data to analyze lapse behavior (He, 2011; Fang and Kung, 2012; Fier and Liebenberg, 2013; Inderst and Sirak, 2014), and the “use of microlevel variations in income represents a major step forward compared to previous studies”

Section 1035 of the U.S. tax code offers tax protection for “policy exchanges”, which indirectly supports the PRH. Furthermore, the frontloaded, short-term compensation structure for life insurance agents and brokers further encourages frequent policy replacements.
These studies generally show support of the EFH over the IRH. In particular, relying on Cox proportional hazards regressions based on German data, Inderst and Sirak (2014) find that those with higher wealth and income are less likely to lapse their policy. Furthermore, different occupation groups show different lapse profiles and (recent) unemployment appears to be a key driver for lapsing. Unlike previous studies, they find that age does appear to be significant once one controls for wealth and policy years. They conclude that there is ample support for the EFH whereas they “rule out” the value-based hypotheses—although this may be due to the specific time period (2005-2011) in which value-driven lapsation may not have been opportune (due to the low/decreasing interest rate environment).

The support for the EFH does not seem surprising in view of the basic theoretical deliberations in the preceding section. So far, the secondary market for life insurance is relatively small and only few policyholders seem to have access to it—an observation that we will come back to later in this section. Hence, there definitely appear to be restrictions in trading the insurance asset. Furthermore, term and whole life policies are the basic instruments protecting against the risk of an early death, so they play an important role in completing feasible consumption profiles across states that can be attained by households. However, two sets of key questions emerge in this context: (i) What types of shocks will lead policyholders to lapse? And are these shocks anticipated correctly by policyholders? (ii) If life insurance holdings are governed by preferences for insurance and these preferences vary over the life-cycle, why then do consumers frequently elect to purchase long-term contracts in the first place? Would it not be more opportune to purchase one-year life contracts sequentially to exclude the loss from premature lapsation?

With regards to the former questions (i), Fang and Kung (2012) attempt to disentangle the drivers for lapsing a policy. Using a semi-structural discrete choice model—calibrated to life insurance holdings from the Health and Retirement Study (HRS) data—they conclude that a large portion of policy lapses are driven by idiosyncratic shocks that are largely unrelated to health, income, and bequest motives—especially when individuals are relatively young. However, as
the policyholders age, the shocks are more systematic and, initially, are predominantly related to income and health. Over the life-cycle, the bequest motive factor becomes increasingly significant.

In view of the latter questions (ii), as already pointed out in the previous section, one answer lies in the existence of additional risk factors such as morbidity risk: If there is uncertainty about the policyholder’s health status, long-term front-loaded contracts as observed in practice arise in a model with one-sided commitment (Hendel and Lizzeri, 2003). The authors test their implications using life insurance lapse data and conclude that all patterns in the data are congruent with their model. In other words, long-term, front-loaded contracts serve as protection against reclassification risk in addition to mortality risk.

However, Gottlieb and Smetters (2014) argue that liquidity shocks and reclassification risk alone do not account for the overall high fraction of lapsed policies. In contrast, according to the authors, their model with differential attention (cf. the previous section) is strongly supported by U.S. policy data whereas they conclude that the patterns are “generally inconsistent with the competing models.” More precisely, they posit that policyholders lapsing after a (negative) health shock and decreasing lapse/surrender fees are inconsistent with reclassification risk. This implies that in view of the second part of questions (i), there appear to be behavioral aspects that lead policyholders to lapse prematurely—although Gottlieb and Smetters (2014) point out that it is impossible to rule out asymmetric information in general as a source for long-term, front-loaded contracts that are lapsed frequently. Aside from differential attention due to narrow framing or cognitive fallacies, using the HRS, Mulholland and Finke (2014) show evidence for numeracy, i.e. basic numerical skills as measured by responses in the survey, as a key driver for lapses: Policyholders with higher levels of numeracy are significantly less likely to lapse their policies. Similarly, based on German household panel survey data, Nolte and Schneider (2015) conclude that policyholders display bounded rationality when it comes to policy lapsation, alluding to financial literacy and heuristics.

With regards to the role of informational frictions, He (2011) analyzes the presence of “dynamic adverse selection”—i.e. whether policyholders consider their own health state in the lap-
sation decision—in the context of the HRS. She finds that despite the substantial front-loading, policyholders take their mortality prospects into account when lapsing policies: Policyholders with higher mortality risk are less likely to lapse (see also Finkelstein et al. (2005) for similar results in the context of long-term care insurance).

Further evidence in this direction is provided by Bauer et al. (2014). Relying on data from a large U.S. life expectancy provider, the authors show the presence of asymmetric information in view of the policyholders’ decisions of whether to settle their policy. More precisely, comparing the mortality profiles for policyholders that settled their policy relative to policyholders that did not settle and controlling for observables (to the life settlement company), Bauer et al. (2014) find a positive correlation between settling and survival—consistent with the prediction under asymmetric/private information on the part of the policyholder. In other words, those who are being offered a “good deal” relative to their private information (that is, the company’s life expectancy estimate was low) tend to take it, whereas those offered a “bad deal” conditional on their private information (the company’s life expectancy estimate was high) tend to walk away from the transaction. Furthermore, the authors argue that the pattern of the differences in mortality between the two groups is in line with adverse selection on the initial health state.4

To summarize, a potpourri of aspects appear to factor in the lapse decision, but recent literature yields a decent understanding of the key drivers: (i) Value is an important aspect in that the predominant pattern is that lapses are decreasing in policy years (see Figure 1); also, policyholders make use of private information that affects the policy value. (ii) However, the triggers for a policy lapse are idiosyncratic (income, health, bequest, etc.) shocks that cannot be perfectly insured using other instruments (market incompleteness)—and the existence of these additional risk factors also serves as an explanation for the predominance of long-term, front-loaded contracts. Yet, there is evidence that policyholders—or at least some policyholders—do not correctly anticipate these shocks, and lapses are higher than predicted by a rational expectations model.

---

4These results are in contrast to contributions from the behavioral literature indicating that individuals fare poorly at forecasting their own mortality prospects (Elder, 2013; Payne et al., 2013). Furthermore, findings with regards to informational advantages upon purchasing life insurance are mixed (Cawley and Philipson, 1999; He, 2009).
Given that the primary drivers are idiosyncratic, yet the prevalence of these idiosyncrasies may vary by policy parameters such as age, underwriting method, risk class, etc., it is not surprising that insurers primarily consider these deterministic aspects when modeling lapsation for the purpose of pricing or policy valuation. Indeed, the theoretical and empirical literature provides positive support for such an approach. However, there are three caveats to this conclusion: (i) As indicated in the introduction, substantial shocks to the economic environment may lead to significant changes in lapse behavior. Given the very long recent period of low interest rates, recent lapse data may not properly reflect the impact a hike in interest rates would have on lapsation (Inderst and Sirak, 2014). (ii) Given the relevance of behavioral factors in explaining the predominance of lapsation, efforts to educate policyholders and to increase consumer financial literacy in general may affect lapse rates; this possibility should be considered in medium- to long-term forecasts of lapse behavior. (iii) Another relevant aspect is the development of a secondary market for life insurance or life-settlements market (Eling and Kochanski, 2013). The number of settled policies thus far is very low relative to the primary life insurance market, and settlements are typically limited to policies with a high face value. As described in the previous section, of course a change may considerably affect the primary insurance market. We believe that developing a thorough understanding of the likelihood of the secondary market blossoming, as well as an appraisal of whether such a development is desirable from the perspective of the insurance industry and/or society as a whole, are key open problems for research (see also Section 4).

Variable Annuities and other Equity-Linked Products

Modeling policyholder behavior is particularly relevant for equity/unit-linked products for several reasons. First, beyond the possibility to surrender the policy for a cash value, these products frequently entail additional options such as the possibility to transfer funds between sub-accounts, to (partially) annuitize the account value, and/or to withdraw a certain (guaranteed) amount free of charge every year. Furthermore, here the contract value immediately depends on the performance of an underlying investment portfolio, so that relatively multifaceted strategies for such behavior
are possible—and how one models policyholder behavior can have a substantial impact on pricing, hedging, and hedge efficiency of variable product lines (Kling et al., 2014).

The significance of this question is reinforced by the increasing importance of these variable products in the insurance landscape, particularly of VAs. Between 2011 and 2013, U.S. VA sales amounted to roughly $150 billion per year, and 76% of these contained GLBs (see the corresponding fact sheets in the LIMRA data bank). The total assets under management exceed $1.5 trillion. Figure 2 provides details on the percentage of policies sold that contained a GLB (blue solid line) and the prevalence of different types of GLBs among the policies (various dotted lines).\(^5\) It is evident that the great majority of all the products contain GLB features that directly depend on policyholder exercise: Withdrawal behavior for GLWBs/GMWBs and annuitization for Guaranteed Minimum Income Benefits (GMIBs). Furthermore, of course surrender behavior may be affected by all the embedded options, including those with Guaranteed Minimum Accumulation Benefits (GMABs) and also GMDBs.

Indeed, Knoller et al. (2014) analyze policyholder surrender behavior for VAs with simple

---

\(^5\)Note that these percentages have changed drastically over the last two decades. For instance, GMWBs used to be the most popular election in the mid-2000s (Sell, 2006).
GMABs based on Japanese data. Unlike the results for conventional products, the authors find that the value of the embedded guarantee has by far the largest explanatory power—whereas they find mixed evidence for the EFH but mild support that financial literacy impacts surrender.

The finding that value is the most important driver of policyholder behavior in variable products is broadly in line with the general approach in the actuarial literature, where these problems are commonly solved using value-maximizing approaches akin to the valuation of American or Bermudan option (see Bauer et al. (2010) and references therein). However, as already indicated, several studies have found discrepancies with corresponding results and market observations when following the value-maximizing approach. For instance, with regards to GMWBs Milevsky and Salisbury (2006) report an “underpricing of this feature [GMWBs] in a typically overpriced VA market” and Chen et al. (2008) posit that “only if several unrealistic modeling assumptions are made it is possible to obtain GMWB fees in the same range as is normally charged” (for similar assertions, see Dai et al. (2008); Blamont and Sagoo (2009)). Piscopo (2010) states that under a no-arbitrage valuation, “GLWB issued on the USA market are underpriced” and that “market fees are not sufficient to cover the market hedging cost of the guarantee.” And Marshall et al. (2010) conclude that according to their no-arbitrage valuation model “fee rates charged by insurance companies for the GMIB option may be too low.”

A potential resolution to this puzzle in the context of GMWBs is provided by the approach in Moenig and Bauer (2014a) that takes into account taxation. As detailed in the previous section, their approach considers the valuation of post-tax cash flows by replicating them with post tax cash flows of some benchmark securities. The authors apply their method to VA plus GMWB products and solve for the optimal withdrawal/surrender strategy. They find that taxes considerably affect the withdrawal behavior, and thus the pricing of the guarantees.

The latter point is illustrated by Table 1, which shows valuation results based on a calibrated version of their model for empirical VA/GMWB products offered in 2007, both with (top rows) and without (bottom rows) the consideration of taxes. As is evident from the table, valuation when not considering taxes results in a negative surplus from offering the GMWB for the insurer, im-
Table VII: Valuation results for empirical GMWB contracts from Table VI with $r = 4\%$.

<table>
<thead>
<tr>
<th>2007 Contracts:</th>
<th>ASL2</th>
<th>HF</th>
<th>HFP</th>
<th>MN</th>
<th>SB5</th>
<th>SB6</th>
<th>SB7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>With Taxes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_0$</td>
<td>99,870</td>
<td>102,300</td>
<td>103,490</td>
<td>102,300</td>
<td>102,820</td>
<td>101,690</td>
<td>100,740</td>
</tr>
<tr>
<td>VA fees</td>
<td>14,640</td>
<td>10,720</td>
<td>11,800</td>
<td>11,840</td>
<td>10,290</td>
<td>10,710</td>
<td>11,010</td>
</tr>
<tr>
<td>GMWB fees</td>
<td>2,770</td>
<td>4,620</td>
<td>2,040</td>
<td>4,690</td>
<td>4,670</td>
<td>4,860</td>
<td>5,000</td>
</tr>
<tr>
<td>Surrender fees</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VA cost</td>
<td>610</td>
<td>550</td>
<td>570</td>
<td>660</td>
<td>490</td>
<td>530</td>
<td>710</td>
</tr>
<tr>
<td>GMWB cost</td>
<td>4,100</td>
<td>4,250</td>
<td>1,160</td>
<td>4,200</td>
<td>7,120</td>
<td>4,620</td>
<td>2,340</td>
</tr>
<tr>
<td>Surplus from GMWB</td>
<td>-1,330</td>
<td>370</td>
<td>880</td>
<td>490</td>
<td>-2,450</td>
<td>240</td>
<td>2,660</td>
</tr>
<tr>
<td><strong>Without Taxes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_0$</td>
<td>98,330</td>
<td>95,400</td>
<td>94,630</td>
<td>95,080</td>
<td>97,920</td>
<td>95,330</td>
<td>94,830</td>
</tr>
<tr>
<td>VA fees</td>
<td>2,550</td>
<td>6,240</td>
<td>6,150</td>
<td>6,930</td>
<td>6,620</td>
<td>5,970</td>
<td>5,640</td>
</tr>
<tr>
<td>GMWB fees</td>
<td>540</td>
<td>2,690</td>
<td>1,060</td>
<td>3,210</td>
<td>3,010</td>
<td>2,710</td>
<td>2,560</td>
</tr>
<tr>
<td>Surrender fees</td>
<td>0</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>880</td>
<td>1,240</td>
<td>1,370</td>
</tr>
<tr>
<td>VA cost</td>
<td>190</td>
<td>320</td>
<td>410</td>
<td>340</td>
<td>370</td>
<td>360</td>
<td>330</td>
</tr>
<tr>
<td>GMWB cost</td>
<td>1,230</td>
<td>4,040</td>
<td>1,460</td>
<td>4,890</td>
<td>8,060</td>
<td>4,890</td>
<td>4,070</td>
</tr>
<tr>
<td>Surplus from GMWB</td>
<td>-690</td>
<td>-1,350</td>
<td>-400</td>
<td>-1,690</td>
<td>-5,050</td>
<td>-2,180</td>
<td>-1,510</td>
</tr>
</tbody>
</table>

Table 1: Valuation results for empirical VA/GMWB products with and without considering taxes, based on an initial investments of 100,000. Source: Moenig and Bauer (2014a).

plying that the fee rates by the insurance companies are too low—in line with the aforementioned literature. However, when accounting for taxes, the surplus results are slightly positive or negative with the average being close to zero (approximately $123). Thus, while the model does not perfectly replicate all the prices, there is no systematic deviation in one direction and the differences between model and market prices are notably smaller.

To provide an intuition for this result, Figure 3, taken from Moenig (2012), plots optimal withdrawal strategies for different valuation approaches in the context of a simple VA/GMWB at a certain point in time and for fixed policy parameters. The panels on the left-hand side, panels (a) and (c), provide optimal withdrawal patterns without and with the consideration of taxes, respectively. The key difference is that without taxes, we observe complete policy surrenders when the GWMB is out-of-the-money (OTM), whereas with taxation there are no withdrawals at all in the OTM region. The intuition is straightforward: When the option is OTM, typically taxes are due on potential withdrawals. In addition, outside investments are subject to capital gains taxa-
Figure 3: Optimal withdrawal patterns for a simple VA/GMWB product under different valuation approaches as functions of the VA account value, $X^-$. The study is based on a 15-year GMWB rider with initial investment 100 and annual guaranteed withdrawal amount 7. All graphs are snapshots from time $t = 10$, under the assumption that no prior withdrawal has been made. Source: Moenig (2012).
tion, whereas funds grow tax-deferred inside the VA (Moenig and Bauer, 2014a). As is detailed in Moenig and Bauer (2014a), this basic pattern is in line with information on dynamic functions that describe GMWB policyholder behavior as a function of whether and how much the guarantee is in-the-money (ITM) (see e.g. Attachment 5: Modeling Specifications in American Academy of Actuaries (2005)). Such functions are derived from empirical behavior and are used by most insurers (Society of Actuaries, 2008). More precisely, these stipulate that withdrawals are increasingly prevalent depending on the ITM ratio whereas the basic surrender schedule is not modified when the option is OTM. Thus, while again the model is not perfectly congruent with the dynamic function, the model reproduces the corresponding patterns at least to first order.

Hence, this subjective risk-neutral valuation (SRNV) approach accounting for tax advantages generates viable results for GMWBs, although it may be suitable to combine it with a deterministic surrender schedule that accounts for surrenders/lapses exogenous to the model. In particular, it appears that incompleteness does not play a major role in this context. This is also illustrated by the bottom panels of Figure 3, where the SRNV approach is presented in contrast to the optimal withdrawal behavior in a life-cycle utility model with taxes and outside investment opportunities. As is evident from the figure, the two models generate very similar patterns. However, as discussed in Bauer and Moenig (2015), it is conceivable that market incompleteness will be relevant for other product lines such as GLWBs.

Another relevant aspect is the strength of the value maximization motive. For instance, Ulm (2010) uses Morningstar and NAIC data to analyze transfer behavior between fixed and variable accounts within VAs with GMDBs. He finds that actual transfer behavior is not in line with value-maximizing transfer strategies as derived in Ulm (2006). In contrast, he shows that policyholders actually transfer in order to “chase returns,” transferring money into stocks if they have done well recently and out of stocks if they have performed poorly. This is a familiar feature found in the mutual funds literature, and a variety of explanations have been provided. For instance, we refer to Da et al. (2015) for an empirical analysis of how investor sentiment predicts mutual fund flows between equity and bond funds (see also references therein).
We summarize that while recent advances in the literature have led to a better grasp of what drives policyholder behavior in theory, research is necessary to better understand the interaction of these drivers and their relevance in the context of different product lines—particularly from an *ex ante* perspective.

4 Conclusion and Future Research

This paper reviews the state of the research on policyholder behavior in life insurance. We discuss theoretical drivers and align them with empirical evidence. Some general principles arise: The value of the insurance contract, at least when considered in isolation, is not sufficient to explain how policyholders lapse their policies and/or make use of exercise-dependent embedded options in advanced life insurance contracts. Depending on the contracts, to align the predictions of theoretical models with observations, it is necessary to incorporate frictions and/or to account for the incompleteness of the market with regards to shocks relevant to a household’s finances. Moreover, documented cognitive and behavioral biases also appear to influence policyholders’ decisions.

Thus far, most papers focus on explaining observed behavior for a certain product line or a certain behavior *ex post*, i.e. they provide insights on relevant frictions and drivers in the context of available data. For a practicing actuary, however, the situation may be more difficult as it may be necessary to form a view on policyholder behavior *ex ante* (e.g. when introducing a new product to the market). A structural understanding of the relevant drivers and their interaction is thus one direction for future research we believe to be highly relevant.

Another important direction for future research is an improved understanding of interactions between policyholder behavior and insurer. Clearly, changes in policyholder behavior may spawn changes in the insurer’s operations, in view of pricing and/or contract design, which in turn may generate different behavior and so on (see e.g. Ulm (2014) for a corresponding model in the context of guaranteed funds). Solving for resulting equilibria can provide insights on important aspects of insurance markets such as the drivers for financial innovation in the VA market. Another aspect that
needs to be considered in the context of an equilibrium is the size, form, and development of the secondary market for life insurance policies—and also its implications on social welfare (Daily et al., 2008; Zhu and Bauer, 2011; Fang and Kung, 2010a,b). As described by Fang and Kung (2012), here the understanding of drivers for policyholder behavior and the result of the welfare analysis go hand in hand.

Therefore, we conclude that although substantial progress has been made in view of understanding policyholder behavior, there are profound open problems and challenges that remain to be answered.

References


