The Fragility of Market Risk Insurance*

Ralph S.J. Koijen† Motohiro Yogo‡

December 31, 2019

ABSTRACT

Insurers sell retail financial products called variable annuities that package mutual funds with minimum return guarantees over long horizons. Variable annuities accounted for $1.5 trillion or 35% of U.S. life insurer liabilities in 2015. Sales fell and fees increased after the 2008 financial crisis as the higher valuation of existing liabilities stressed risk-based capital. Insurers also made guarantees less generous or stopped offering guarantees to reduce risk exposure. We develop an equilibrium model of insurance markets in which financial frictions and market power are important determinants of pricing, contract characteristics, and the degree of market incompleteness.

*A.M. Best Company, Morningstar, and the NAIC own the copyright to their respective data, which we use with permission under their license agreements with Princeton University. This paper is based upon work supported by the National Science Foundation grant 1727049 and the Julis-Rabinowitz Center for Public Policy and Finance. We thank Adam Xu and Zhen Ye for assistance on constructing data from Morningstar Annuity Intelligence. For comments and discussions, we thank Naoki Aizawa, Mark Flannery, Victoria Ivashina, Arvind Krishnamurthy, Emanuel Mönch, Borghan Narajabad, Theo Nijman, Don Noh, Anna Paulson, Richard Rosen, Donghwa Shin, and Haiyue Yu. We also thank seminar participants at Boston University; Federal Reserve Bank of Minneapolis; Federal Reserve Board; Michigan State University; NYU; Ohio State University; Princeton University; Temple University; UC Berkeley; UCLA; University of Chicago; University of Cincinnati; University of Delaware; University of Maryland; University of Michigan; University of Nebraska; UNC; University of South Carolina; UT Austin; UVA; Derivatives and Volatility 2017: The State of the Art; 2017 DNB/Riksbank Macroprudential Conference; 2017 SITE Workshop on Financial Regulation; 2017 NBER Conference on Financial Market Regulation; 2017 ICPM-Netspar Discussion Forum; 2017 IMF Conference on Monetary, Financial, and Prudential Policy Interactions in the Post-Crisis World; 2017 NBER Insurance Meeting; and 2018 Macroeconomics of Pensions and Retirement Financing Conference.

†University of Chicago, Booth School of Business, NBER, and CEPR (e-mail: Ralph.Koijen@chicagobooth.edu).

‡Princeton University, Department of Economics and NBER (e-mail: myogo@princeton.edu).

The traditional role of life insurers is to insure idiosyncratic risk through products like life annuities, life insurance, and health insurance. With the secular decline of defined benefit pension plans and Social Security around the world, life insurers are increasingly taking on the role of insuring market risk through minimum return guarantees. In the U.S., life insurers sell retail financial products called variable annuities that package mutual funds with minimum return guarantees over long horizons. Variable annuities have grown to be the largest category of life insurer liabilities, larger than traditional annuities and life insurance, and accounted for $1.5 trillion or 35% of U.S. life insurer liabilities in 2015. Variable annuities also represent an important share of the mutual fund sector because the underlying investments are mutual funds.

The large size of the variable annuity market reflects its importance for household welfare. In theory, minimum return guarantees could facilitate efficient risk sharing across heterogeneous agents (Dumas, 1989, Chan and Kogan, 2002) or overlapping generations (Allen and Gale, 1997, Ball and Mankiw, 2007). Investors cannot easily replicate minimum return guarantees because traded options have shorter maturity and model uncertainty exposes investors to basis risk in a dynamic hedging program. Therefore, insurers complete a missing market for long-maturity options by offering minimum return guarantees over long horizons.

From the insurers' perspective, minimum return guarantees are difficult to price and hedge because traded options have shorter maturity. Imperfect hedging leads to risk mismatch that stresses risk-based capital when the valuation of existing liabilities increases with a falling stock market, falling interest rates, or rising volatility. During the 2008 financial crisis, many insurers including Aegon, Allianz, AXA, Delaware Life, John Hancock, and Voya suffered large increases in variable annuity liabilities ranging from 27% to 125% of total equity. Hartford was bailed out by the Troubled Asset Relief Program in June 2009 because of significant losses on their variable annuity business.[1] Risk mismatch between general account assets and minimum return guarantees leads to negative duration and negative convexity for the overall balance sheet and poses a challenge for life insurers in the low interest rate environment after the financial crisis. As a consequence, the stock returns of U.S. life insurers have significant negative exposure to long-term bond returns after the financial crisis (Hartley et al., 2017). Given their size and potential risk, variable annuities are an essential piece of the puzzle for understanding the insurance sector more broadly.

To this end, we construct a new and comprehensive panel data set on the variable annuity market at the contract level. The data contain quarterly sales, fees, and contract characteristics from 1999:1 (first quarter) to 2015:4 (fourth quarter). We combine these data with the

[1] Other examples of risk mismatch due to minimum return guarantees include the financial distress of Japanese life insurers in the 1990s (Kashyap, 2002) and the failure of Equitable Life in 2000 (Roberts, 2012).

insurers' annual financial statements from 2005 to 2015, which contain information about the value of variable annuity liabilities and the share of these liabilities that are reinsured. The data provide a detailed account of how the variable annuity market has evolved over time as the changing valuation and risk exposure of existing liabilities affected the insurers' financial strength.

Quarterly sales of variable annuities grew robustly from $22 billion in 2005:1 to $41 billion in 2007:4 and subsequently fell to $27 billion in 2009:2. At the same time, the average annual fee on contracts with minimum return guarantees increased from 2.04% of account value in 2007:4 to 2.38% in 2009:2, suggesting an important role for a supply shock. After the financial crisis, insurers made the minimum return guarantees less generous or stopped offering guarantees to reduce risk exposure. In the cross section of insurers, sales fell more for insurers that suffered larger increases in the valuation of existing liabilities. These insurers moved their variable annuity liabilities off balance sheet through reinsurance, consistent with the importance of a risk-based capital constraint (Koijen and Yogo, 2016).

To interpret this evidence, we develop an equilibrium model of insurance markets in which financial frictions and market power are important determinants of pricing, contract characteristics, and the degree of market incompleteness. Insurers compete in an oligopolistic market by setting the fee and the rollup rate, which is a key contract characteristic that is equivalent to the strike price of a put option. Required capital increases in the rollup rate because of a risk-based capital or an economic risk constraint. An adverse shock to the valuation of existing liabilities increases the shadow cost of capital and drives up the marginal cost of issuing contracts. The insurer not only raises the fee but lowers the rollup rate to reduce risk exposure. When the shadow cost of capital is sufficiently high, the insurer stops offering minimum return guarantees to avoid additional risk exposure.

The demand for variable annuities could depend on factors other than the fee and the rollup rate. They include the attractiveness (such as a tax advantage) of variable annuities relative to other savings products, the menu of options within contracts, and insurer characteristics that capture reputation in the retail market. We model these factors through a differentiated product demand system for the variable annuity market, which implies estimates of demand elasticities to the fee and the rollup rate. We also estimate the insurer's optimality conditions for the fee and the rollup rate, which depend on the demand elasticities, the frictionless option value, and the shadow cost of capital. We find that marginal cost explains a 33 basis point increase in the average annual fee from 2007:4 to 2009:2, of which 8 basis points is due to the option value and 25 basis points is due to the shadow cost of capital. Thus, financial frictions are more important than the option pricing channel for explaining the increase in fees during the financial crisis.

Previous research shows that exposure to market and interest rate risk from variable annuities is one of the most important sources of risk for U.S. life insurers. Insurers with variable annuity liabilities became financially constrained and significantly reduced the prices of traditional annuities and life insurance during the financial crisis (Koijen and Yogo, 2015). Insurers with variable annuity liabilities hold less liquid bonds, and the common exposure to market risk through variable annuities makes these insurers more vulnerable to fire-sale dynamics in bond markets (Ellul et al., 2018). Insurers with variable annuity liabilities have negative duration and negative convexity, making them vulnerable to a prolonged period of low interest rates (Hartley et al., 2017). As a consequence, they continue to register high in systemic risk indicators long after the financial crisis (Acharya et al., 2017). These papers study the consequences of variable annuity liabilities on other parts of the balance sheet, but they do not study the source of risk directly. Our contribution is to use contract-level data to provide a deeper understanding of the variable annuity market itself and its impact on life insurers.

Another contribution is to develop a theory of market risk insurance, building on the work of Froot (2007) for catastrophe insurance. Variable annuities, which guarantee investment returns over long horizons, are essentially a private solution to a gap left by the secular decline of defined benefit pension plans and Social Security. Insurers complete a missing market for long-maturity options by offering market risk insurance over long horizons, but they do so only imperfectly because of financial frictions and market power. Analogous to Rothschild and Stiglitz (1976), our theory of insurance markets explains pricing, contract characteristics, and the degree of market incompleteness. However, financial frictions and market power instead of informational frictions are the important determinants of market equilibrium. Our theory could apply to other insurance markets in which insurers bear significant aggregate risk over long horizons, such as the long-term care insurance market (Cutler, 1996).

Our work also relates to the mutual fund literature. Previous research has shown that mutual fund flows depend on past performance (Chevalier and Ellison, 1997, Sirri and Tufano, 1998, Wermers, 2003) and tax efficiency (Bergstresser and Poterba, 2002, Sialm and Starks, 2012). At the same time, demand is significantly less elastic to fees than the law of one price implies, which suggests an important role for product differentiation and market power (Hortaçsu and Syverson, 2004). We study the determinants of supply and demand for variable annuities, which has received relatively little attention despite being the largest life insurer liability and an important share of the mutual fund sector.

The remainder of this paper proceeds as follows. Section I describes variable annuities and details about their regulation that are relevant for this paper. Section II describes the data construction and summarizes key facts about the variable annuity market. Section III presents a model of variable annuity supply that explains the evidence on pricing and contract characteristics. Section IV estimates a model of variable annuity supply and demand to quantify the importance of financial frictions. Section V concludes.

I. Institutional Background

We start with an example of an actual product to explain how variable annuities work. We then summarize risk-based capital regulation, which is important for understanding how an adverse shock to the valuation of existing liabilities could affect variable annuity supply. We also explain how an economic risk constraint could work in conjunction with a risk-based capital constraint. We then summarize economic and institutional reasons why insurers do not fully hedge variable annuity risk. Finally, we present evidence on interest risk mismatch after the financial crisis.

A. An Example of a Variable Annuity

Insurers sell long-term savings products called variable annuities, which are investments in mutual funds. For an additional fee, insurers offer an optional minimum return guarantee on the mutual fund. Thus, a variable annuity is a retail financial product that packages a mutual fund with a long-maturity put option on the mutual fund. To explain how variable annuities work, we start with an example of an actual product.

MetLife Investors USA Insurance Company (2008) offers a variable annuity contract called MetLife Series VA, which comes with various investment options and guaranteed living benefits. In 2008:3, one of the investment options was the American Funds Growth Allocation Portfolio, which is a mutual fund with a target equity allocation of 70% to 85% and an annual portfolio expense of 1.01%. One of the guaranteed living benefits was Guaranteed Lifetime Withdrawal Benefit (GLWB). MetLife Series VA has an annual base contract expense of 1.3% of account value, and GLWB has an annual fee of 0.5% of account value. Thus, the total annual fee for the variable annuity with a GLWB is 1.8%, which is on top of the annual portfolio expense on the mutual fund.

Suppose that an investor were to invest in the American Funds Growth Allocation Portfolio in 2008:3. After 2013:3, the investor withdraws a constant dollar amount each year that is 5% of the highest account value ever reached. For example, this behavior describes an investor who invests in a mutual fund five years prior to retirement and subsequently spends down her wealth by consuming a constant dollar amount each year. Figure 1 shows the path of account value per $1 of initial investment, with the shaded region covering the withdrawal period after 2013:3. The account value fluctuates over time because of uncertainty in investment returns.

The same investor could purchase a GLWB from MetLife and guarantee her investment returns. GLWB has an annual rollup rate of 5% prior to first withdrawal, which means that at each contract anniversary, the guaranteed amount steps up to the greater of the account value and the previous guaranteed amount accumulated at 5%. Thus, GLWB is a put option on the mutual fund that locks in every year to a strike price that accumulates at an annual rate of 5%. Figure 1 shows that the guaranteed amount can only increase during the five-year accumulation period, protecting the investor from downside risk in investment returns.

Once the investor enters the withdrawal period, she can annually withdraw up to 5% of the highest guaranteed amount ever reached. In our example, the guaranteed amount in 2013:3 is \$1.44, which means that the investor can withdraw up to $\$1.44 \times 0.05 = \$0.072$ per year. Each withdrawal gets deducted from both the account value and the guaranteed amount. GLWB is a lifetime guarantee in that the investor receives income (i.e., \$0.072 per year) as long as she lives, even after the account is depleted to zero. During the withdrawal period, the guaranteed amount steps up to the account value at each contract anniversary. In Figure 1, these step-ups occur in 2014:3 and 2016:3 because of high investment returns.

Because the annual rollup rate is 5% and the annual fee is 0.5%, one may be tempted to conclude that the guaranteed return on the variable annuity is 4.5% during the accumulation period. This logic turns out to be incorrect because the guaranteed amount of \$1.44 in 2013:3 is only payable as annual income of \$0.072 over 20 years (or until the investor's death). Because of the time value of money, the present value of \$0.072 per year over 20 years is worth substantially less than \$1.44. Appendix A shows the empirical relevance of this issue based on the historical term structure of interest rates.

GLWB is the most common type of guaranteed living benefit. The other three types of guaranteed living benefits are Guaranteed Minimum Withdrawal Benefit (GMWB), Guaranteed Minimum Income Benefit (GMIB), and Guaranteed Minimum Accumulation Benefit (GMAB). A GMWB is similar to a GLWB, except that the investor does not receive income after the account is depleted to zero. A GMIB is similar to a GLWB, except that guaranteed amount at the beginning of the withdrawal period converts to a life annuity (i.e., fixed income for life). A GMAB provides a minimum return guarantee much like the accumulation period of a GLWB, but it does not have a withdrawal period with guaranteed income.

If an investor were to die while the contract is in effect, her estate receives a standard death benefit that is equal to the remaining account value. For an additional fee, insurers offer four types of guaranteed death benefits (highest anniversary value, rising floor, earnings enhancement benefit, and return of premium) that enhance the death benefit during the accumulation period. Our main focus is on guaranteed living benefits, so we will not go into the details of guaranteed death benefits in this paper.

Even without guaranteed living benefits, variable annuities may be attractive to investors because of a potential tax advantage in non-qualified accounts. Earnings on variable annuities can be deferred and accumulate tax free if the first withdrawal occurs after age 59.5. However, all earnings including capital gains are taxed at the ordinary income tax rate, which is higher than the capital gains tax rate. Therefore, the tax advantage can justify the higher fees on variable annuities only if the accumulation period is very long. In an illustrative example, Brown and Poterba (2006, Table 5.2) show that the accumulation period must be longer than 40 years to justify an annual fee of 0.25% under the 2003 tax rates and an 8% pre-tax return (with 2% from dividends and 6% from capital gains).

B. Risk-Based Capital Regulation

Insurance regulators and rating agencies use risk-based capital as an important metric of an insurer's financial strength. Risk-based capital is the ratio of accounting equity to required capital:

$$\text{RBC} = \frac{\text{Assets} - \text{Reserves}}{\text{Required capital}}. \tag{1}$$

Reserves in the numerator is an accounting measure of liabilities that may not coincide with market value. Required capital in the denominator is a measure of how much equity could be lost in an adverse scenario. For a sufficiently high risk-based capital ratio, insurance regulators view that equity capital is adequate to meet the insurer's existing liabilities even in an adverse scenario.

Variable annuity liabilities enter both reserves and required capital in risk-based capital. As summarized in Junus and Motiwalla (2009), Actuarial Guideline 43 since December 2009 determines the reserve value of variable annuities, and the C-3 Phase II regulatory standard since December 2005 determines the contribution of variable annuities to required capital. Actuarial Guideline 43 is a higher reserve requirement than its precursor Actuarial Guideline 39, so insurers were given a phase-in period until December 2012 to fully comply with the new requirement.

To compute reserves and required capital, insurance regulators provide various scenarios for the joint path of Treasury, corporate bond, and equity prices. Insurers simulate the path of equity deficiency for their variable annuity business (net of the hedging programs and reinsurance) under each scenario and keep the highest present value of equity deficiency along each path. Insurers then compute reserves as a conditional mean over the upper 30% of equity deficiencies. This conditional tail expectation builds in a degree of conservatism that is conceptually similar to a correction for risk premia, but reserves do not coincide with the market value of liabilities. Insurers use the same methodology for required capital, except that they compute a conditional mean over the upper 10% of equity deficiencies.

More generous guarantees with higher rollup rates or better coverage of downside market risk relative to fees require higher reserves and more capital. Moreover, minimum return guarantees are long-maturity put options on mutual funds whose value increases when the stock market falls, interest rates fall, or volatility rises. Therefore, an adverse scenario like the financial crisis increases both reserves and required capital and puts downward pressure on risk-based capital. Insofar as insurers want to avoid a rating downgrade or regulatory action, an adverse shock to the valuation of existing liabilities could affect their ability to issue new liabilities. In Section III, we present a model that formalizes this mechanism through which financial frictions affect variable annuity supply.

In addition to the risk-based capital constraint, the insurer could have an economic risk constraint as part of risk management. An economic risk constraint works similarly to a risk-based capital constraint, except that the relevant measure of assets and liabilities is market value. For example, let ϵ be a multiplicative shock to the leverage ratio due to risk mismatch from variable annuities, whose cumulative distribution function is F. Consider a value-at-risk constraint under which the probability that assets exceed liabilities must exceed a threshold:

$$\Pr\left(\frac{\text{Liabilities}}{\text{Assets}}\epsilon \leq 1\right) = F\left(\frac{\text{Assets}}{\text{Liabilities}}\right) \geq \kappa. \tag{2}$$

We can rewrite this constraint as

$$\frac{\text{Assets} - \text{Liabilities}}{(F^{-1}(\kappa) - 1)\text{Liabilities}} \geq 1, \tag{3}$$

which is similar to risk-based capital (1). An insurer with more conservative risk management has higher $F^{-1}(\kappa)$, either through higher κ or lower risk reflected in the distribution of ϵ.

As a consequence of the financial crisis, the insurer could learn that model uncertainty is higher than previously recognized. In response, the insurer could make risk management more conservative, tightening the economic risk constraint. Thus, an economic risk constraint could work in conjunction with a risk-based capital constraint and affect variable annuity supply.

C. Reasons for Risk Mismatch

In theory, insurers could hedge uncertainty in the valuation of variable annuity liabilities through offsetting derivative positions. In practice, insurers do not fully hedge variable annuity risk for various economic and institutional reasons (Drexler et al., 2017, Koijen and Yogo, 2017, Ellul et al., 2018, Sen, 2019).

Insurers may not be able to fully hedge because minimum return guarantees have longer maturity than traded options. Insurers are exposed to unexpected changes in implied volatility if they attempt to hedge minimum return guarantees by rolling over shorter maturity options. A dynamic hedging program would be subject to basis risk because of model uncertainty, especially regarding long-run volatility (Sun, 2009, Sun et al., 2009). A deeper economic question is why the market for long-maturity options is incomplete if insurers would want to hedge such risks. A potential reason is that someone must bear aggregate risk by market clearing, and insurers may have comparative advantage over other types of institutions because their liabilities have a longer maturity and are less vulnerable to runs (Paulson et al., 2012).

Insurers, especially stock rather than mutual companies, may not want to hedge because of risk shifting motives that arise from limited liability and the presence of state guaranty funds (Lee et al., 1997). Another reason that insurers may not want to hedge is that existing regulation does not properly reward hedging of market equity (Sen, 2019). Insurers report accounting equity under statutory accounting principles at the operating company level and under generally accepted accounting principles (GAAP) at the holding company level. Therefore, hedge positions differ depending on whether the insurer targets economic, statutory, or GAAP capital. A hedging program that smoothes market equity could actually increase the volatility of accounting equity under statutory accounting principles or GAAP (Credit Suisse, 2012).

Whether insurers target market or accounting equity depends on whether the more important friction is economic (i.e., value-at-risk constraint) or regulatory (i.e., risk-based capital constraint). For reducing regulatory frictions, reinsurance with an off-balance-sheet entity within the same insurance group is more efficient than hedging (Koijen and Yogo, 2016). Consistent with this view, Section IV shows that insurers used reinsurance to move variable annuity liabilities off balance sheet during the financial crisis.

D. Evidence on Interest Risk Mismatch

If the minimum return guarantees have higher duration and higher convexity than the general account assets, the overall balance sheet is potentially exposed to interest rate risk.

Equity capital decreases with unexpected decreases in interest rates, especially when the level of interest rates is low. Consistent with this hypothesis, Hartley et al. (2017) find that the stock returns of U.S. life insurers have significant negative exposure to long-term bond returns in the prolonged period of low interest rates after the financial crisis. We replicate their results here and later refer to this evidence as a potential explanation for why the variable annuity market never fully recovered after the financial crisis.

We construct monthly returns on a value-weighted portfolio of publicly traded U.S. life insurers that have variable annuity liabilities, which are listed in Appendix B. We regress excess portfolio returns, relative to the 1-month T-bill rate, onto excess stock market returns and excess 10-year Treasury bond returns. Table I reports the betas and the monthly alpha from the factor regression.

Over the sample period from January 1999 to December 2017, the stock market beta is 1.33, and the 10-year Treasury beta is 0.03 and statistically insignificant. On average, insurers do not have significant exposure to interest rate risk, controlling for exposure to the overall stock market. However, the 10-year Treasury beta varies over time when we break the sample into three periods: pre-crisis (1999–2007), financial crisis (2008–2009), and post-crisis (2010–2017). In the post-crisis subsample, the 10-year Treasury beta is -1.12 with a t-statistic greater than 7. That is, unexpected decreases in interest rates are bad news for U.S. life insurers during this prolonged period of low interest rates. Hartley et al. (2017) show that property-casualty insurers and U.K. life insurers (that do not have variable annuities) do not have such exposure to interest rates. They conclude that U.S. life insurers are uniquely exposed to interest rate risk because of negative duration and negative convexity from variable annuities.

II. Data on the Variable Annuity Market

A. Data Construction

We use three sources to construct a comprehensive panel data set on the variable annuity market at the contract level. The first data source is Morningstar (2016a), which has quarterly sales of variable annuities at the contract level since 1999. Morningstar provides a textual summary of the prospectus for each contract, from which we extract the history of fees and contract characteristics. The key contract characteristics are the base contract expense, the number of investment options, and the types of guaranteed living and death benefits that are offered.[2] For each guaranteed living benefit, the key characteristics are the

[2] We use assets under management by subaccount from Morningstar (2016b) to compute a measure of investment options that adjusts for the non-uniform distribution of assets across subaccounts within a con-

type (i.e., GLWB, GMWB, GMIB, or GMAB), the fee, the rollup rate, and the withdrawal rate. Morningstar provides the open and close dates for each contract and guaranteed living benefit, from which we construct the history of when different benefits were offered.

Sales are available at the contract level but not at the benefit level. Therefore, we must aggregate fees and rollup rates over all guaranteed living benefits that a contract offers to construct a panel data set on sales, fees, and characteristics at the contract level. For each date and contract, we first average the fees and the rollup rates by the type of guaranteed living benefit. We then use the average fee and rollup rate in the order of GLWB, GMWB, GMIB, and GMAB, based on availability. For example, if a contract does not offer a GLWB but offers a GMWB, we use the average fee and rollup rate on the GMWB. Because GLWB is the most common type of guaranteed living benefit and a GMWB is the closest substitute to a GLWB, our procedure yields a representative set of fees and rollup rates that are comparable across contracts.

The second data source is the annual financial statements of insurers, which are filed with the NAIC (National Association of Insurance Commissioners, 2005–2015). General Interrogatories Part 2 Table 9.2 of the financial statements reports the total related account value, the gross amount of reserves, and the reinsurance reserve credit on variable annuities. As we described in Section I, the total related account value is the market value of the mutual funds. The gross amount of reserves is the accounting value of the minimum return guarantees net of the hedging programs. We define variable annuity liabilities as the total related account value plus the gross amount of reserves minus the reinsurance reserve credit. For each insurer, we construct the *reserve valuation* as the ratio of gross amount of reserves to total related account value. The reserve valuation is an important measure of the option value of variable annuity liabilities (net of the hedging programs). In the cross section, reserve valuation is higher for insurers that have sold more generous guarantees. In the time series, reserve valuation increases when the stock market falls, interest rates fall, or volatility rises.

The third data source is A.M. Best Company (2006–2016), which provides a cleaned and organized version of the main parts of the annual financial statements. Following A.M. Best's definition of financial groups, we aggregate insurance companies' balance sheets up to the group level. Total liabilities are aggregate reserves for life contracts plus liabilities from the separate account statement. Total equity is capital and surplus. We convert the A.M. Best financial strength rating (coded from A++ to D) to a cardinal measure (coded from 175% to 0%) based on risk-based capital guidelines (A.M. Best Company, 2011, p. 24).

tract. Our measure is the inverse of the Herfindahl index over the subaccount shares within each contract, which is the number of investment options when the subaccounts are uniformly distributed.

We merge the A.M. Best data and the NAIC data by the NAIC company code. We then merge the Morningstar data and the NAIC data by company name. The final data set is a quarterly panel on the variable annuity market from 2005:1 to 2015:4, where the start date is dictated by the availability of the NAIC data. For some of the summary statistics that only involve the Morningstar data, we use a longer sample from 1999:1.

B. Summary of the Variable Annuity Market

Table II summarizes the variable annuity market. In 2005, variable annuity liabilities across all insurers were $1.071 trillion or 35% of total liabilities. Variable annuity liabilities have ranged from 34% to 41% of total liabilities as its value fluctuates with the market value of the mutual funds. Most recently in 2015, variable annuity liabilities were $1.499 trillion or 35% of total liabilities. The variable annuity market is fairly concentrated as measured by the number of insurers. The total number of insurers fell from 44 in 2008 to 38 in 2015.

As we explained previously, the reserve valuation (i.e., the ratio of gross amount of reserves to total related account value) measures the option value of variable annuity liabilities. Table II shows that the reserve valuation aggregated across all insurers increased sharply from 0.8% in 2007 to 4.1% in 2008. Since 2008, the reserve valuation is volatile and remains high relative to the level prior to the financial crisis.

Table III reports the top insurers ranked by their variable annuity liabilities in 2007. Eight of these insurers (Aegon, Allianz, AXA, Delaware Life, Hartford, Jackson National, Metropolitan Life, and Voya) suffered large increases in the reserve valuation ranging from 2.9 to 7.6 percentage points. These increases in the reserve valuation are significant shocks because these insurers have high leverage (i.e., the ratio of total liabilities to total assets) that range from 92% to 97%. For five of the eight insurers, the increases in gross amount of reserves are a significant share of total equity, ranging from 29% to 125%.

Figure 2 reports quarterly sales of variable annuities across all contracts from 1999:1 to 2015:4. Sales grew robustly from $22 billion in 2005:1 to its peak at $41 billion in 2007:4. Sales subsequently fell during the financial crisis to $27 billion in 2009:2, picked up again to $34 billion in 2011:2, and are $20 billion most recently in 2015:4. For comparison, the same figure shows the aggregate sales of U.S. open-end stock and bond mutual funds (excluding money market funds and funds of funds), which is a larger market and shown on a different scale. Interestingly, sales of variable annuities and mutual funds moved closely together through 2008, but the two time series diverge thereafter as mutual fund sales grew.

The decline in variable annuity sales after 2008 is partly explained by insurers that have stopped offering guaranteed living benefits. Figure 3 reports the number of insurers and contracts offering guaranteed living benefits from 1999:1 to 2015:4. Eleven insurers stopped offering guaranteed living benefits from 2008 to 2015, during which six insurers stopped selling variable annuities altogether as reported in Table II. This means that some insurers have opted to remain in the variable annuity market but to stop offering minimum return guarantees. Without minimum return guarantees, variable annuities are essentially mutual funds with longevity insurance and a potential tax advantage.

The upper panel of Figure 4 reports the average annual fee on open (for sale) guaranteed living benefits from 1999:1 to 2015:4. The increase in fees during the financial crisis coincides with the decline in sales, suggesting an important role for a supply shock. The average annual fee on guaranteed living benefits increased from 0.59% of account value in 2007:4 to 0.97% in 2009:2. Including the base contract expense, the total annual fee increased from 2.04% in 2007:4 to 2.38% in 2009:2. Since then, fees have remained stable. The average annual fee on guaranteed living benefits was 1.08% (2.33% including the base contract expense) in 2015:4.

The lower panel of Figure 4 reports two lines that summarize rollup rates on open contracts from 1999:1 to 2015:4. Conditional on offering a guaranteed living benefit, the dashed line shows that the average rollup rate increased from 2.4% in 2005:1 to 4.0% in 2007:4, coinciding with the period of robust sales growth. The average rollup rate remained high through the financial crisis and decreased only after 2011. However, the solid line shows that the share of contracts with guaranteed living benefits dropped immediately during the financial crisis, consistent with Figure 3. That is, many insurers responded to the financial crisis through the extensive margin by not offering contracts with guaranteed living benefits, rather through the intensive margin of lowering rollup rates.

To summarize Figures 2–4, variable annuity sales fell, fees increased, and many insurers stopped offering guaranteed living benefits during the financial crisis. As we discussed in Section I, this evidence is consistent with a supply shock as a consequence of tightening risk-based capital and economic risk constraints. As we also discussed in Section I, two factors could explain why variable annuity supply did not fully recover long after the financial crisis. The primary factor is interest risk mismatch that reduced equity capital during the prolonged period of low interest rates. A secondary factor is a higher capital requirement under Actuarial Guideline 43. Despite the enormous attention it received in the industry, the actual impact of Actuarial Guideline 43 on variable annuity supply is difficult to identify because of its gradual implementation from 2009 to 2012.

III. A Model of Variable Annuity Supply

As we discussed in Section I, risk-based capital and economic risk constraints are important determinants of variable annuity supply and explain the aggregate facts in Section II.

Insurers suffered an adverse shock to risk-based capital from the increased valuation of existing liabilities during the financial crisis. Moreover, insurers could have made risk management more conservative in response to higher model uncertainty. As the shadow cost of capital increased, insurers raised fees to pass through a higher marginal cost. Insurers also lowered rollup rates or stopped offering minimum return guarantees to reduce risk exposure. Higher fees and lower rollup rates make variable annuities less attractive to investors, explaining the decline in sales.

We formalize this narrative through a model of how an insurer chooses the fee and the rollup rate in the presence of financial frictions and market power. Insurers compete in an oligopolistic market and have market power because of product differentiation along contract characteristics other than the fee and the rollup rate, which we parameterize through a differentiated product demand system in Section IV. To simplify the exposition in this section, we present the optimization problem of a single insurer with the understanding that all insurers solve the same problem. To simplify the notation, we suppress subscripts that denote the identity of the insurer.

Also to simplify the exposition, we model the optimization problem as a one-time choice. We refer to our previous work for a dynamic version in which the insurer chooses the optimal price in every period (Koijen and Yogo, 2015, 2016). Relative to our previous work, the novel modeling ingredient is the optimal choice of contract characteristics, and the novel insight is that the insurer changes contract characteristics or stops offering minimum return guarantees to reduce risk exposure. Thus, we develop a more complete theory of the supply side of insurance markets that explains pricing, contract characteristics, and the degree of market incompleteness.

A. *Variable Annuity Market*

We start with high-level assumptions about financial markets that are standard in an option pricing model. There is a mutual fund whose price evolves exogenously over time. To simplify the notation, we assume no portfolio expense on the mutual fund. Let S_t be the mutual fund price per share in period t. By the absence of arbitrage, there exists a strictly positive stochastic discount factor $M_{t,t+s}$ that discounts a payoff in period $t + s$ to its price in period t. For example, the mutual fund price satisfies $S_t = \mathbb{E}_t[M_{t,t+s}S_{t+s}]$.

In period t, an insurer sells a variable annuity, which is a combination of the mutual fund and a minimum return guarantee. The variable annuity fee is P_t per dollar of account value, which we assume is paid upfront in a lumpsum for simplicity. The minimum return guarantee is over two periods, and the rollup rate r_t is the guaranteed return per period.

Thus, the payoff of the minimum return guarantee in period $t+2$ is

$$X_{t,t+2} = \max\left\{(1+r_t)^2 - \frac{S_{t+2}}{S_t}, 0\right\}. \tag{4}$$

The minimum return guarantee is a long-maturity put option whose strike price is the cumulative rollup rate. When $r_t = -1$, the variable annuity is a mutual fund because the put option is always worthless. We assume that the investor cannot insure downside market risk over long horizons outside of variable annuities, so the insurance market is incomplete when $r_t = -1$.

The option value of the minimum return guarantee in period t is

$$V_{t,t} = \mathbb{E}_t[M_{t,t+2}X_{t,t+2}] \tag{5}$$

per dollar of account value. More generally, $V_{t-s,t}$ denotes the option value in period t of a minimum return guarantee that was issued in period $t-s$. Although this notation is slightly cumbersome, it will be important to distinguish the option value of existing liabilities $V_{t-1,t}$ from the option value of new contracts $V_{t,t}$.

For the purposes of our theory, we do not need parametric assumptions about the option pricing model (e.g., Black and Scholes, 1973). We simply assume that the partial derivatives of option value have the usual signs. Namely, the put option value decreases in the mutual fund price, decreases in the riskless interest rate, increases in volatility, and increases in the rollup rate. In the language of Greeks in the option pricing literature, delta is negative, rho is negative, vega is positive, and dual delta is positive. Furthermore, we assume that the second derivative of option value with respect to the rollup rate is positive.

We also do not need parametric assumptions about variable annuity demand. We simply assume that demand is continuous, continuously differentiable, strictly decreasing in the fee, and strictly increasing in the rollup rate. In an oligopolistic market, demand for a contract depends on the fees and the rollup rates of all other competing contracts. To simplify the notation, we denote demand for a contract as Q_t with the understanding that it depends on the fees and the rollup rates of all other contracts. An institutional feature of the variable annuity market is that the rollup rate is always positive (i.e., $r_t \geq 0$) or $r_t = -1$ in the case of mutual funds with no minimum return guarantees. That is, insurers never offer a variable annuity with a negative rollup rate in the range $r_t \in (-1, 0)$, presumably because investors have a psychological aversion to “negative interest rates”. To model this institutional feature, we simply assume that the choice of the rollup rate is constrained to be in the set $\mathcal{R} = \{-1\} \bigcup [0, \infty)$.

B. *Balance Sheet Dynamics*

Let B_t be the total account value of mutual funds, or separate accounts in actuarial terms, at the end of period t. Let A_t be the general account assets at the end of period t. Let L_t be the general account liabilities, which represents the option value of existing minimum return guarantees, at the end of period t. The following T account summarizes the balance sheet.

Assets	Liabilities	
B_t	B_t	(separate account)
A_t	L_t	(general account)
	$A_t - L_t$	(equity)

There is no risk mismatch for mutual funds in the separate account. Equity fluctuates because of risk mismatch between assets and minimum return guarantees in the general account.

We now describe how variable annuity sales affect the balance sheet. Let Q_t be the account value of new contracts, excluding the option value of minimum return guarantees, that the insurer sells in period t. The account value evolves according to

$$B_t = \frac{S_t}{S_{t-1}} B_{t-1} + Q_t. \tag{6}$$

Current account value is the previous account value revalued at the current mutual fund price plus the account value of new contracts.

The general account assets evolve according to

$$A_t = R_{A,t} A_{t-1} + P_t Q_t, \tag{7}$$

where $R_{A,t}$ is an exogenous gross return on assets in period t. Current assets are the gross return on previous assets plus the fees on new contracts. Section I discussed economic and institutional reasons why insurers do not fully hedge variable annuity risk. Following that discussion, we assume that $R_{A,t}$ could be imperfectly correlated with the option value of existing liabilities, leading to risk mismatch.

The general account liabilities evolve according to

$$L_t = \frac{V_{t-1,t}}{V_{t-1,t-1}} L_{t-1} + V_{t,t} Q_t. \tag{8}$$

Current liabilities are previous liabilities revalued at current option value plus the cost of new contracts. The principle of reserving requires that the cost $V_{t,t}$ be recorded on the liability side to back the fees P_t on the asset side.

C. *Financial Frictions*

We define statutory capital at the end of period t as

$$K_t = \underbrace{A_t - L_t}_{\text{equity}} - \underbrace{\phi_t L_t}_{\text{required capital}}. \tag{9}$$

Statutory capital is equity minus required capital that is proportional to liabilities.[3] For simplicity, we assume that $\phi_t > 0$ is an exogenous parameter that does not depend on the fee or the rollup rate. Following the discussion in Section I, $1 + \phi_t$ represents the ratio of reserve to market value under Actuarial Guideline 43. Alternatively, ϕ_t represents the risk weight on minimum return guarantees under the C-3 Phase II regulatory standard. As equation (8) shows, required capital increases in the option value of existing liabilities $V_{t-1,t}$. Therefore, required capital increases when the stock market falls, interest rates fall, or volatility rises. Required capital also increases in the option value of new contracts $V_{t,t}$. Therefore, required capital for new contracts increases in the rollup rate, decreases in interest rates, and increases in volatility.

Following the discussion in Section I, low statutory capital could lead to a rating downgrade or regulatory action, which have adverse consequences in both retail and capital markets. We model the cost of financial frictions through a cost function

$$C_t = C(K_t), \tag{10}$$

which is continuous, twice continuously differentiable, strictly decreasing, and strictly convex. The cost function is decreasing because higher statutory capital reduces the likelihood of a rating downgrade or regulatory action. The cost function is convex because these benefits of higher statutory capital have diminishing returns.[4] Statutory capital would not matter if equity issuance were costless. Therefore, implicit in the specification of the cost function are financial frictions that make equity issuance costly.

An alternative interpretation of equation (9) is that the insurer has an economic risk constraint, such as the value-at-risk constraint described in Section I. As a consequence

[3]This formulation of statutory capital as a difference rather than as a ratio is for mathematical convenience in the derivations that follow. Koijen and Yogo (2015) show that the two formulations are similar because a constraint on statutory capital such as $K_t \geq 0$ can be rewritten as a risk-based capital constraint $\frac{A_t - L_t}{\phi_t L_t} \geq 1$.

[4]We refer to Ellul et al. (2015) and Koijen and Yogo (2015) for evidence that asset allocation and liability pricing decisions are especially sensitive to risk-based capital at low levels, which implies that the cost function is convex.

of the financial crisis, the insurer learned that model uncertainty is higher than previously recognized and made risk management more conservative. An increase in ϕ_t could capture a tighter economic risk constraint. A permanent increase in ϕ_t could lead to persistent effects on variable annuity supply that is consistent with the evidence in Section II.

D. *Optimal Pricing and Contract Characteristics*

Firm value is the profit from variable annuity sales minus the cost of financial frictions:

$$J_t = (P_t - V_{t,t})Q_t - C_t. \tag{11}$$

The insurer chooses the fee P_t and the rollup rate $r_t \in \mathcal{R}$ on the variable annuity to maximize firm value in an oligopolistic market. To simplify the exposition, we present the optimality conditions for a single insurer with the understanding that all insurers have the same optimality conditions in a Nash equilibrium.

To simplify the notation, we define the semi-elasticity of demand to the fee as $\epsilon_{P,t} = -\frac{\partial \log(Q_t)}{\partial P_t}$ and to the rollup rate as $\epsilon_{r,t} = \frac{\partial \log(Q_t)}{\partial r_t}$. We also define the negative of the partial derivative of the cost function as

$$c_t = -\frac{\partial C_t}{\partial K_t} > 0. \tag{12}$$

This partial derivative represents the importance of financial frictions, which decreases in statutory capital by the convexity of the cost function. The following proposition, which we prove in Appendix C, characterizes the optimal fee and rollup rate.

PROPOSITION 1: *The optimal fee is*

$$P_t = \frac{1}{\epsilon_{P,t}} + \underbrace{\lambda_t V_{t,t}}_{\textit{marginal cost}}, \tag{13}$$

where the shadow cost of capital is

$$\lambda_t = \frac{1 + c_t(1 + \phi_t)}{1 + c_t} > 1. \tag{14}$$

At an interior optimum, the rollup rate satisfies

$$\frac{\epsilon_{r,t}}{\epsilon_{P,t}} = \lambda_t \frac{\partial V_{t,t}}{\partial r_t}. \tag{15}$$

Otherwise, $r_t \in \{-1, 0\}$ *is optimal.*

The optimal fee (13) is the sum of two terms. The first term represents Bertrand pricing due to market power, under which the optimal fee decreases in the semi-elasticity of demand to the fee. The second term is the marginal cost of issuing contracts, which is the product of the shadow cost of capital and the option value. Marginal cost is greater than the option value because of financial frictions. The shadow cost of capital decreases in statutory capital through c_t and increases in the capital requirement ϕ_t.

For readers who are familiar with our earlier work (Koijen and Yogo, 2015), we clarify two potential points of confusion. First, equation (13) in this paper is identical to equation (21) in Koijen and Yogo (2015). The reason that they may appear different is that $\epsilon_{P,t}$ is the semi-elasticity of demand in this paper, while it is the full elasticity of demand in the earlier work. Semi-elasticity is the natural formulation in this paper because fees are already in percentages of account value. Second, equation (13) implies that marginal cost *decreases* in the shadow cost of capital if $\phi_t < 0$. In Koijen and Yogo (2015), the prices of traditional annuities decreased during the financial crisis because the effective ϕ_t was negative for those products.

The optimal rollup rate depends on three terms in equation (15). On the left side is the demand channel through which the insurer optimally chooses the rollup rate to exploit market power. The first term on the right side is the shadow cost of capital. The second term on the right side is the sensitivity of option value to the rollup rate, which increases in the rollup rate (i.e., $\frac{\partial^2 V_{t,t}}{\partial r_t^2} > 0$). The intuition behind equation (15) is especially clear when the left side is constant, which is the case for the constant semi-elasticity demand function in Appendix C and a special case of the differentiated demand system in Section IV. In this case, the optimal rollup rate decreases in the shadow cost of capital. An insurer that faces a higher shadow cost of capital must reduce risk exposure by lowering the rollup rate on new contracts. When the shadow cost of capital is sufficiently high, the insurer offers mutual funds with no minimum return guarantees (i.e., $r_t = -1$). That is, the insurer stops offering minimum return guarantees to avoid additional risk exposure. The general insight is that financial frictions affect contract characteristics and could even lead to market incompleteness in the extreme case.

The shadow cost of capital is not directly observed. However, the reserve valuation is a relevant empirical proxy that we described in Section II. The reserve valuation most closely corresponds to $(1 + \phi_t)V_{t-1,t}$ in the model, which is positively related to the shadow cost of capital. Therefore, we derive comparative statics for the optimal fee and rollup rate with respect to the reserve valuation. For a general demand function, equations (13) and (15) do not yield clean comparative statics because the semi-elasticities of demand could depend on the fee and the rollup rate. For the purposes of obtaining analytical insights, we assume constant semi-elasticities of demand in the following corollary to Proposition 1. We refer to Appendix C for a proof and an example of a constant semi-elasticity demand function.

COROLLARY 1: *If semi-elasticities of demand ϵ_P and ϵ_r are constant, the optimal fee increases in the reserve valuation (i.e., $\frac{\partial P_t}{\partial (1+\phi_t)V_{t-1,t}} > 0$), and the optimal rollup rate decreases in the reserve valuation (i.e., $\frac{\partial r_t}{\partial (1+\phi_t)V_{t-1,t}} < 0$). Therefore, sales decrease in the reserve valuation (i.e., $\frac{\partial Q_t}{\partial (1+\phi_t)V_{t-1,t}} < 0$).*

Corollary 1 provides a narrative for the aggregate facts in Figures 2–4. Insurers suffered an adverse shock to risk-based capital as the reserve valuation increased during the financial crisis. Moreover, insurers could have made risk management more conservative in response to higher model uncertainty. As the shadow cost of capital increased, insurers raised fees to pass through a higher marginal cost. Insurers also lowered rollup rates or stopped offering minimum return guarantees to reduce risk exposure. Higher fees and lower rollup rates make variable annuities less attractive to investors, explaining the decline in sales.

E. Evidence from the Cross Section of Insurers

We now provide evidence from the cross section of insurers that is consistent with Corollary 1. We look for broad patterns at the insurer level that could be summarized by a scatter plot and leave more formal analysis at the contract level for Section IV. Depending on the contract characteristics of existing liabilities, different insurers could experience different shocks to the reserve valuation during the financial crisis. Insurers that sold more generous guarantees prior to the financial crisis would have suffered larger increases in the reserve valuation than those that sold less generous guarantees. Moreover, insurers that sold more generous guarantees could have made risk management more conservative after the financial crisis as they learned that model uncertainty is higher than previously recognized. Thus, changes in the reserve valuation should be negatively related to sales growth in the cross section of insurers.

The upper panel of Figure 5 is a scatter plot of sales growth versus the change in the reserve valuation from 2007 to 2010. The linear regression line shows that sales growth is negatively related to the change in the reserve valuation. On the bottom right are insurers like AXA and Genworth that essentially closed their variable annuity business as they suffered large increases in the reserve valuation. On the left side of the figure are a cluster of six insurers (Fidelity Investments, MassMutual, New York Life, Northwestern, Ohio National, and Thrivent Financial) that did not offer a GLWB in 2007, which is a more generous guarantee than other types of guaranteed living benefits. The reserve valuation did not change much for these insurers because they sold less generous guarantees.

Insurers could relax a risk-based capital constraint through reinsurance with an off-balance-sheet entity within the same insurance group (Koijen and Yogo, 2016). If insurers that suffered large increases in the reserve valuation were in fact constrained, they should move variable annuity liabilities off balance sheet. The bottom panel of Figure 5 is a scatter plot of the change in the share of variable annuity reserves reinsured versus the change in the reserve valuation from 2007 to 2010. The linear regression line shows that the change in the share of variable annuity reserves reinsured is positively related to the change in the reserve valuation. On the one hand, AXA increased the share of variable annuity reserves reinsured by 61 percentage points as its reserve valuation increased by 11 percentage points from 2007 to 2010. On the other hand, the six insurers that did not offer a GLWB in 2007 did not experience any change in the reserve valuation or reinsurance activity. This evidence suggests an important role for a risk-based capital constraint because reinsurance within the same insurance group does not relax an economic risk constraint.

IV. Importance of Financial Frictions

Variation in fees across insurers and over time could come from supply- or demand-side effects. A model of variable annuity demand is necessary to disentangle these effects and to quantify the importance of financial frictions in explaining variable annuity supply. Therefore, we estimate a differentiated product demand system for the variable annuity market at the contract level, which provides an internally consistent framework to model market equilibrium and to decompose fees into markups versus marginal cost.

A. A Model of Variable Annuity Demand

A life-cycle model of consumption and portfolio choice is a fully structural approach to modeling variable annuity demand (Horneff et al., 2009, 2010, Koijen et al., 2011). These models could explain the demand for variable annuities relative to other savings products, but they are not designed to explain heterogeneous demand across insurers and contracts. Moreover, we do not have data on the demographics of investors that purchase variable annuities at the contract level. For these reasons, we take a different approach and model variable annuity demand through the random coefficients logit model (Berry et al., 1995), which is a tractable and micro-founded model of product differentiation and market power.

Let $P_{i,t}$ be the annual fee and $r_{i,t}$ be the rollup rate on contract i in period t. Let $\mathbf{x}_{i,t}$ be a vector of other observable characteristics of contract i in period t, which are determinants of demand. Let $\xi_{i,t}$ be an unobserved (to the econometrician) characteristic of contract i in period t. The probability that an investor with a realized coefficient α_P on the fee buys contract i in period t is

$$q_{i,t}(\alpha_P) = \frac{\exp\{\alpha_P P_{i,t} + \alpha_r r_{i,t} + \beta' \mathbf{x}_{i,t} + \xi_{i,t}\}}{1 + \sum_{j=1}^{I} \exp\{\alpha_P P_{j,t} + \alpha_r r_{j,t} + \beta' \mathbf{x}_{j,t} + \xi_{j,t}\}}, \tag{16}$$

where I is the total number of contracts across all insurers. The denominator of equation (16) captures how demand for a contract depends on the fees and characteristics of all other competing contracts. If the investor does not buy a variable annuity, she buys an "outside asset" instead, which occurs with probability $1 - \sum_{i=1}^{I} q_{i,t}(\alpha_P)$.

Let $F(\alpha_P)$ be the cumulative distribution function for the coefficient on the fee. The negative of the coefficient on the fee $-\alpha_P$ is lognormally distributed, ensuring a positive demand elasticity. Integrating equation (16) over the distribution of investors, the market share for contract i in period t is

$$Q_{i,t} = \int q_{i,t}(\alpha_P)\ dF(\alpha_P). \tag{17}$$

The semi-elasticity of demand to the fee for contract i in period t is

$$\epsilon_{P,t} = \frac{1}{Q_{i,t}} \int -\alpha_P q_{i,t}(\alpha_P)(1 - q_{i,t}(\alpha_P))\ dF(\alpha_P). \tag{18}$$

Through equation (13), the markup is inversely related to the semi-elasticity of demand.

We assume that the coefficient on the fee is independently and identically distributed over time. However, the semi-elasticity of demand to the fee can vary over time through the changing distribution of market shares. To see this, consider a special case of equation (18) for the logit model, for which the coefficient on the fee is constant (i.e., $\alpha_P = \overline{\alpha}_P < 0$). The semi-elasticity of demand to the fee simplifies to

$$\epsilon_{P,t} = -\overline{\alpha}_P(1 - Q_{i,t}) > 0. \tag{19}$$

The semi-elasticity of demand to the fee decreases as the market share increases, implying a higher markup through equation (13). Thus, one mechanism through which fees on guaranteed living benefits could increase (see Figure 4) is increasing market concentration as insurers exit the variable annuity market after the financial crisis (see Figure 3).

The estimation sample is all variable annuity contracts from 2005:1 to 2015:4. Because sales are at the contract level, we measure the total annual fee as the sum of the annual base contract expense and the annual fee on the guaranteed living benefit. We assign a type of guaranteed living benefit to each contract following the procedure described in Section II.

The rollup rate is 0% for contracts with guaranteed living benefits but no step ups and −100% for contracts without guaranteed living benefits. This treatment of the rollup rate is consistent with the model of variable annuity supply in Section III, in which we assumed that demand is continuously differentiable in the rollup rate. We specify the outside asset as sales of open-end stock and bond mutual funds (excluding money market funds and funds of funds).

The other contract characteristics in our specification are the number of investment options and a dummy for whether the contract offers a GLWB. These two characteristics capture the menu or the complexity of options within contracts (Célérier and Vallée, 2017). We also include the A.M. Best rating and insurer fixed effects to capture reputation in the retail market, which could vary across insurers and over time. Investors could substitute across insurers based on ratings, or they could substitute from variable annuities to mutual funds (i.e., the outside asset) if they are concerned about the stability of the insurance sector. The unobserved characteristic $\xi_{i,t}$ in equation (16) captures other demand factors that are difficult to measure such as a relative tax advantage. Finally, the intercept captures the attractiveness (such as a tax advantage) of variable annuities relative to mutual funds.

B. Identifying Assumptions

According to the model of variable annuity supply in Section III, the insurer optimally chooses the fee and the rollup rate, so they are jointly endogenous with demand. We start with the usual identifying assumption that observed characteristics other than the fee and the rollup rate are exogenous. Furthermore, we assume that the reserve valuation and the share of variable annuity reserves reinsured are valid instruments that affect marginal cost, but they do not enter demand directly. To ensure exogeneity, we construct both instruments in year t based only on contracts that the insurer sold in prior years but are still on the balance sheet in year t. Thus, the instruments do not depend directly on sales or contract characteristics in year t. Because our specification includes insurer fixed effects, the demand elasticities are identified from the time-series variation in the instruments within each insurer.

We motivate the reserve valuation as a relevant and valid instrument, based on the model of variable annuity supply in Section III. According to Corollary 1, the reserve valuation $(1+\phi_t)V_{t-1,t}$ is a relevant instrument that is correlated with the fee and the rollup rate. The reserve valuation depends on the option value of existing liabilities $V_{t-1,t}$, which is different from the option value of new contracts $V_{t,t}$. Recall that $1+\phi_t$ represents the ratio of reserve to market value under Actuarial Guideline 43. Thus, the reserve valuation is an accounting value that does not coincide with the market value or the investors' valuation that enters demand. As we discussed in Section I, insurers compute reserves and required capital as a conditional tail expectation using the insurance regulators' scenarios, which ultimately depend on contract characteristics. However, investors value these characteristics differently than insurers because their marginal utility depends on the usefulness of variable annuities for aggregate risk sharing, insuring longevity risk, and tax management. Therefore, contract characteristics enter demand differently than how they enter the insurer's conditional tail expectation. Thus, we have plausibly exogenous variation in the reserve valuation that affects demand only through marginal cost, conditional on contract characteristics in our specification.

We have a similar motivation for the share of variable annuity reserves reinsured as an instrument. Koijen and Yogo (2016) show that most of reinsurance is with less regulated and unrated off-balance-sheet entities within the same insurance group, which relaxes regulatory constraints and reduces tax liabilities. Thus, reinsurance lowers marginal cost, but it does not affect demand directly conditional on contract and insurer characteristics in our specification. This assumption is plausible insofar as investors have little motive or knowledge to condition demand on reinsurance activity beyond what is already reflected in ratings.

We estimate the random coefficients logit model by two-step generalized method of moments. In addition to the reserve valuation and the share of variable annuity reserves reinsured, we use the square of these instruments to help identify the variance of the random coefficient on the fee. We approximate the integral over the distribution of the coefficient on the fee through a simulation with 500 draws.

C. Estimating Variable Annuity Demand

Table IV reports the estimated mean and standard deviation of the random coefficients for the model of variable annuity demand. The mean coefficient on the fee is -3.10 with a standard error of 0.17. The standard deviation of the random coefficient on the fee is 0.22 and statistically significant. The coefficient on the rollup rate is 0.18 with a standard error of 0.02. The signs of the coefficients confirm that demand decreases in the fee and increases the rollup rate.

Demand also increases in the number of investment options and the A.M. Best rating. The coefficient on the number of investment options is 0.07 with a standard error of 0.01. The coefficient on the A.M. Best rating, which is standardized, is 0.56 with a standard error of 0.14. This means that a standard deviation increase in the rating increases demand by 56%. The dummy for GLWB enters with the opposite sign to what we expected, but it is not a statistically significant determinant of demand.

We compute the semi-elasticity of demand for each contract through equation (18). For contracts with guaranteed living benefits, the semi-elasticity of demand to the fee has a mean of 14.8 and a standard deviation of 0.5 across contracts in 2007:4. The average semi-elasticity of demand to the fee falls to 14.2 in 2009:2 and ultimately to 13.8 in 2015:4. The decrease in demand elasticity is consistent with increasing market concentration as insurers exit the variable annuity market after the financial crisis. However, the increase in market concentration is too small to generate a large decrease in the demand elasticity.

A semi-elasticity of 14.8 means that demand decreases by 14.8% per 1 basis point increase in the fee. In comparison, the average semi-elasticity of demand to the fee is 7 for S&P 500 index funds, implied by the average fee and estimated marginal cost during 1995–2000 (Hortaçsu and Syverson, 2004). The demand elasticity for variable annuities is higher than that for S&P 500 index funds, which is consistent with the fact that variable annuity investors are wealthier and less risk averse than the average household (Brown and Poterba, 2006). We also conjecture that variable annuity investors are more likely to shop around, especially given the irreversible nature of their investment.

Our baseline specification limits the random coefficients to the fee. For robustness, we have estimated a richer model in which the coefficients on the rollup rate or the A.M. Best rating are also random. However, the standard deviation of the random coefficient converged to zero or had large standard errors that indicated that the richer model is poorly identified. The identification problem arises from the fact that the variation in market shares can only identify a limited covariance structure for the random coefficients.

D. *Parameterizing Variable Annuity Supply*

For contract i sold by insurer n in period t, equation (13) for the optimal fee in logarithms is

$$\log\left(P_{i,t} - \frac{1}{\epsilon_{P,i,t}}\right) = \log(V_{i,t}) + \log(\lambda_{n,t}). \tag{20}$$

This equation provides a decomposition of marginal cost into the option value and the shadow cost of capital. The option value explains within-insurer variation in marginal cost along contract characteristics, while the shadow cost of capital explains between-insurer variation in marginal cost along insurer characteristics.

For contract i sold by insurer n in period t, equation (15) for the optimal rollup rate in logarithms is

$$\log\left(\frac{\epsilon_{r,i,t}}{\epsilon_{P,i,t}}\right) = \log\left(\frac{\partial V_{i,t}}{\partial r_{i,t}}\right) + \log(\lambda_{n,t}). \tag{21}$$

At an interior optimum, the marginal benefit of a higher rollup rate through demand is equal to the marginal cost of a higher rollup rate through financial frictions. Because the same shadow cost of capital enters both equations (20) and (21), both fees and rollup rates contribute toward the identification of the shadow cost of capital. Intuitively, a high shadow cost of capital must simultaneously lead to a high fee and a low rollup rate across all contracts that an insurer offers.

To transform equations (20) and (21) into estimation equations, we take four steps. First, we parameterize the option value to depend on the rollup rate and a vector $\mathbf{y}_{i,t}$ of other contract characteristics, which are the number of investment options and a dummy for GLWB. The option value of contract i in period t is

$$V_{i,t} = \exp\{\delta'\mathbf{y}_{i,t} + \exp\{\Delta'\mathbf{y}_{i,t}\}r_{i,t} + \eta_{i,t}\}, \tag{22}$$

where the residual $\eta_{i,t}$ represents unobserved (to the econometrician) contract characteristics. This specification implies that the partial derivative of option value with respect to the rollup rate is

$$\frac{\partial V_{i,t}}{\partial r_{i,t}} = \exp\{\Delta'\mathbf{y}_{i,t}\}V_{i,t}. \tag{23}$$

Thus, the coefficients δ determine the level of option value, and the coefficients Δ determine the slope of option value with respect to the rollup rate.

Second, we parameterize the shadow cost of capital to depend on a vector $\mathbf{z}_{n,t}$ of insurer characteristics, which are the A.M. Best rating, log reserve valuation, and the share of reserves reinsured. The shadow cost of capital for insurer n in period t is

$$\lambda_{n,t} = \exp\{\Gamma'\mathbf{z}_{n,t} + \gamma_n\}, \tag{24}$$

where γ_n are insurer fixed effects.

Third, we derived equations (13) and (15) under the assumption that the insurer offers only one contract. In reality, the insurer offers multiple contracts and presumably chooses the fees and the rollup rates accounting for demand elasticities across contracts. In Appendix C, we derive a more general version of equations (13) and (15) for a multi-product insurer and describe how to compute semi-elasticities of demand from the estimated model of variable annuity demand.

Fourth, equation (21) becomes an inequality at the corner solutions when the rollup rate is 0% or −100% (i.e., a contract without a guaranteed living benefit). We transform a moment inequality into a moment equality, following Luttmer (1996). We include a dummy $\mathbb{1}_{r_{i,t}}(0)$ for whether the rollup rate is 0%, and a dummy $\mathbb{1}_{r_{i,t}}(-1)$ for whether the contract does not offer a guaranteed living benefit. We also introduce a residual $\nu_{i,t}$ that represents unobserved (to the econometrician) deviations from the moment equality (21).

Thus, the estimation equations are

$$\log\left(P_{i,t} - \frac{1}{\epsilon_{P,i,t}}\right) = \delta' \mathbf{y}_{i,t} + \exp\{\Delta' \mathbf{y}_{i,t}\} r_{i,t} + \Gamma' \mathbf{z}_{n,t} + \gamma_n + \eta_{i,t}, \tag{25}$$

$$\log\left(\frac{\epsilon_{r,i,t}}{\epsilon_{P,i,t}}\right) - \log\left(P_{i,t} - \frac{1}{\epsilon_{P,i,t}}\right) = \Delta' \mathbf{y}_{i,t} + \rho' \begin{bmatrix} \mathbb{1}_{r_{i,t}}(0) \\ \mathbb{1}_{r_{i,t}}(-1) \end{bmatrix} + \nu_{i,t}. \tag{26}$$

Equation (25) for the optimal fee identifies the level of option value and the shadow cost of capital. After subtracting marginal cost, equation (26) for the optimal rollup rate identifies only the slope of option value with respect to the rollup rate. This clean separability comes from the fact that the optimal rollup rate depends on the shadow cost of capital only through marginal cost, given our parametric assumptions. We estimate equations (25) and (26) jointly by two-step generalized method of moments.

The intercept in equation (25) is the unconditional mean of marginal cost, from which we cannot separately identify the unconditional mean of the option value versus the shadow cost of capital. This issue is inconsequential for our main findings, which concern the time-series variation in the option value and the shadow cost of capital. For the purposes of presentation, we normalize the unconditional mean of the shadow cost of capital so that $\log(\lambda_{n,t}) = 0$ for the lowest realized value in our sample. This procedure leads to an upper bound of the option value and a lower bound of the shadow cost of capital for each contract.

E. Estimating Variable Annuity Supply

Table V reports estimates of equations (25) and (26). The signs of the coefficients on the insurer characteristics are consistent with the hypothesis that they capture the shadow cost of capital. That is, the shadow cost of capital decreases in the A.M. Best rating and increases in log reserve valuation and the share of reserves reinsured. These estimates also validate the "first stage" of the demand estimation in Table IV, which relies on log reserve valuation and the share of reserves reinsured as relevant instruments for fees and rollup rates.

Figure 6 summarizes the importance of the shadow cost of capital for the time series of fees. We decompose the total annual fee for contracts with guaranteed living benefits from 2005:1 to 2015:4, averaged across contracts with sales weights. The total annual fee increased by 34 basis points from 2.04% of account value in 2007:4 to 2.38% in 2009:2. This increase in annual fees reflects an increase in marginal cost by 33 basis points from 1.97% of account value in 2007:4 to 2.30% in 2009:2. We further decompose this increase in marginal cost into 8 basis points in the option value and 25 basis points in the shadow cost of capital. Thus, financial frictions are more important than the option pricing channel for explaining the increase in fees during the financial crisis.

In Section II, we discussed two factors that could explain why variable annuity supply did not fully recover long after the financial crisis. The primary factor is interest risk mismatch that reduced equity capital during the prolonged period of low interest rates. A secondary factor is a higher capital requirement under Actuarial Guideline 43. These two factors ultimately drive up the shadow cost of capital, which is consistent with the evdience in Figure 6.

V. Conclusion

The traditional insurance literature focuses on products such as life annuities, life insurance, and health insurance that insure idiosyncratic risk. This literature shows that informational frictions lead to variation in prices and contract characteristics across different types of individuals (Finkelstein and Poterba, 2004). However, the main business of life insurers is now savings products that insure market risk through minimum return guarantees. Although we focus on the U.S. because of data availability, guaranteed return products are important globally and represent a major share of life insurer liabilities in Austria, Denmark, France, Germany, Netherlands, and Sweden (European Systemic Risk Board, 2015, Hombert and Lyonnet, 2017). The key frictions in this market are financial frictions and market power, which lead to variation in prices and contract characteristics across insurers and over time.

This paper also has important implications for the literature on financial intermediation. Mutual funds are traditionally pure pass-through institutions with no risk mismatch. However, an important and growing part of the mutual fund sector that is sold through life insurers is subject to risk mismatch through minimum return guarantees. In that sense, life insurers are becoming more like pension funds because they have risky assets and guaranteed liabilities. The persistent under-funding of pension funds may foreshadow similar problems for life insurers in the future (Novy-Marx and Rauh, 2011). The fact that life insurers are publicly traded and subject to market discipline could lead to additional challenges that are not present for under-funded pension funds.

REFERENCES

Acharya, Viral V., Thomas Philippon, and Matthew Richardson, 2017, Measuring systemic risk for insurance companies, in Felix Hufeld, Ralph S. J. Koijen, and Christian Thimann, eds., *The Economics, Regulation, and Systemic Risk of Insurance Markets*, chapter 5, 100–123 (Oxford University Press, Oxford).

Allen, Franklin, and Douglas Gale, 1997, Financial markets, intermediaries, and intertemporal smoothing, *Journal of Political Economy* 105, 523–546.

A.M. Best Company, 2006–2016, *Best's Statement File: Life/Health, United States* (A.M. Best Company, Oldwick, NJ).

A.M. Best Company, 2011, Best's credit rating methodology: Global life and non-life insurance edition, A.M. Best Methodology.

Ball, Laurence, and N. Gregory Mankiw, 2007, Intergenerational risk sharing in the spirit of Arrow, Debreu, and Rawls, with applications to Social Security design, *Journal of Political Economy* 115, 523–547.

Bergstresser, Daniel, and James Poterba, 2002, Do after-tax returns affect mutual fund inflows?, *Journal of Financial Economics* 63, 381–414.

Berry, Steven, James Levinsohn, and Ariel Pakes, 1995, Automobile prices in market equilibrium, *Econometrica* 63, 841–890.

Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Brown, Jeffrey R., and James M. Poterba, 2006, Household ownership of variable annuities, in James M. Poterba, ed., *Tax Policy and the Economy*, volume 20, chapter 5, 163–191 (MIT Press, Cambridge, MA).

Célérier, Claire, and Boris Vallée, 2017, Catering to investors through security design: Headline rate and complexity, *Quarterly Journal of Economics* 132, 1469–1508.

Chan, Yeung Lewis, and Leonid Kogan, 2002, Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices, *Journal of Political Economy* 110, 1255–1285.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Credit Suisse, 2012, US variable annuities, Credit Suisse Connections Series.

Cutler, David M., 1996, Why don't markets insure long-term risk?, Working paper, Harvard University.

Drexler, Alejandro, Thanases Plestis, and Richard J. Rosen, 2017, How much risk do variable annuity guarantees pose to life insurers?, Chicago Fed Letter 384.

Dumas, Bernard, 1989, Two-person dynamic equilibrium in the capital market, *Review of Financial Studies* 2, 157–188.

Ellul, Andrew, Chotibhak Jotikasthira, Anastasia Kartasheva, Christian T. Lundblad, and Wolf Wagner, 2018, Insurers as asset managers and systemic risk, Working paper, Indiana University.

Ellul, Andrew, Chotibhak Jotikasthira, Christian T. Lundblad, and Yihui Wang, 2015, Is historical cost accounting a panacea? Market stress, incentive distortions, and gains trading, *Journal of Finance* 70, 2489–2538.

European Systemic Risk Board, 2015, Issues note on risks and vulnerabilities in the EU financial system, Working paper, European Systemic Risk Board.

Finkelstein, Amy, and James Poterba, 2004, Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market, *Journal of Political Economy* 112, 183–208.

Froot, Kenneth A., 2007, Risk management, capital budgeting, and capital structure policy for insurers and reinsurers, *Journal of Risk and Insurance* 74, 273–299.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2007, The U.S. Treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291–2304.

Hartley, Daniel, Anna Paulson, and Richard J. Rosen, 2017, Measuring interest rate risk in the life insurance sector: The U.S. and the U.K., in Felix Hufeld, Ralph S. J. Koijen, and Christian Thimann, eds., *The Economics, Regulation, and Systemic Risk of Insurance Markets*, chapter 6, 124–150 (Oxford University Press, Oxford).

Hombert, Johan, and Victor Lyonnet, 2017, Intergenerational risk sharing in life insurance: Evidence from France, Working paper, HEC Paris.

Horneff, Wolfram J., Raimond H. Maurer, Olivia S. Mitchell, and Michael Z. Stamos, 2009, Asset allocation and location over the life cycle with investment-linked survival-contingent payouts, *Journal of Banking and Finance* 33, 1688–1699.

Horneff, Wolfram J., Raimond H. Maurer, Olivia S. Mitchell, and Michael Z. Stamos, 2010, Variable payout annuities and dynamic portfolio choice in retirement, *Journal of Pension Economics and Finance* 9, 163–183.

Hortaçsu, Ali, and Chad Syverson, 2004, Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds, *Quarterly Journal of Economics* 119, 403–456.

Junus, Novian, and Zohair Motiwalla, 2009, A discussion of Actuarial Guideline 43 for variable annuities, Milliman Research Report.

Kashyap, Anil K., 2002, Sorting out Japan's financial crisis, *Federal Reserve Bank of Chicago Economic Perspectives* 26, 42–55.

Koijen, Ralph S. J., Theo E. Nijman, and Bas J. M. Werker, 2011, Optimal annuity risk management, *Review of Finance* 15, 799–833.

Koijen, Ralph S. J., and Motohiro Yogo, 2015, The cost of financial frictions for life insurers, *American Economic Review* 105, 445–475.

Koijen, Ralph S. J., and Motohiro Yogo, 2016, Shadow insurance, *Econometrica* 84, 1265–1287.

Koijen, Ralph S. J., and Motohiro Yogo, 2017, Risk of life insurers: Recent trends and transmission mechanisms, in Felix Hufeld, Ralph S. J. Koijen, and Christian Thimann, eds., *The Economics, Regulation, and Systemic Risk of Insurance Markets*, chapter 4, 79–99 (Oxford University Press, Oxford).

Lee, Soon-Jae, David Mayers, and Clifford W. Smith, Jr., 1997, Guaranty funds and risk-taking: Evidence from the insurance industry, *Journal of Financial Economics* 44, 3–24.

Luttmer, Erzo G. J., 1996, Asset pricing in economies with frictions, *Econometrica* 64, 1439–1467.

MetLife Investors USA Insurance Company, 2008, MetLife Series VA prospectus.

Morningstar, 2016a, *Morningstar Annuity Intelligence* (Morningstar, Inc., Chicago, IL).

Morningstar, 2016b, *Morningstar Direct* (Morningstar, Inc., Chicago, IL).

National Association of Insurance Commissioners, 2005–2015, *Annual Life InfoPro* (National Association of Insurance Commissioners, Kansas City, MO).

Novy-Marx, Robert, and Joshua Rauh, 2011, Public pension promises: How big are they and what are they worth?, *Journal of Finance* 66, 1211–1249.

Paulson, Anna, Richard Rosen, Zain Mohey-Deen, and Robert McMenamin, 2012, How liquid are U.S. life insurance liabilities?, Chicago Fed Letter 302.

Roberts, Richard, 2012, Did anyone learn anything from the Equitable Life? Lessons and learning from financial crises, Working paper, King's College London.

Rothschild, Michael, and Joseph E. Stiglitz, 1976, Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *Quarterly Journal of Economics* 90, 630–649.

Sen, Ishita, 2019, Regulatory limits to risk management, Working paper, Harvard Business School.

Sialm, Clemens, and Laura Starks, 2012, Mutual fund tax clienteles, *Journal of Finance* 67, 1397–1422.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Sun, Peter, 2009, The VA industry: An analysis of recent activities, Milliman Research Report.

Sun, Peter, Ken Mungan, Joshua Corrigan, and Gary Finkelstein, 2009, Performance of insurance company hedging programs during the recent capital market crisis, Milliman Research Report.

Wermers, Russ, 2003, Is money really "smart"? New evidence on the relation between mutual fund flows, manager behavior, and performance persistence, Working paper, University of Maryland.

Table I
Risk Exposure of U.S. Life Insurers

We construct monthly returns on a value-weighted portfolio of publicly traded U.S. life insurers that have variable annuity liabilities, which are listed in Appendix B. This table reports the betas and monthly alpha from a factor regression of excess portfolio returns, relative to the 1-month T-bill rate, onto excess stock market returns and excess 10-year Treasury bond returns. Heteroscedasticity-robust standard errors are reported in parentheses. The sample period is January 1999 through December 2017.

		By subsample		
Factor		1999–2007	2008–2009	2010–2017
Stock market	1.33	0.55	2.47	1.13
	(0.17)	(0.14)	(0.22)	(0.08)
10-year Treasury	0.03	-0.37	1.16	-1.12
	(0.30)	(0.27)	(0.52)	(0.15)
Alpha (%)	-0.20	0.32	-0.98	0.33
	(0.44)	(0.45)	(1.60)	(0.28)
Observations	228	108	24	96

Table II
A Summary of the Variable Annuity Market

The reserve valuation is the ratio of gross amount of reserves to total related account value.

	VA liabilities			
Year	Billion $	% of total liabilities	Number of insurers	Reserve valuation (%)
2005	1,071	35	45	0.9
2006	1,276	38	47	0.8
2007	1,435	41	46	0.8
2008	1,068	34	44	4.1
2009	1,195	35	43	3.4
2010	1,344	36	43	2.5
2011	1,358	35	42	4.9
2012	1,434	36	39	3.9
2013	1,606	37	40	1.8
2014	1,599	37	38	2.3
2015	1,499	35	38	2.9

Table III
Top Insurers by Variable Annuity Liabilities

The reserve valuation is the ratio of gross amount of reserves to total related account value. The change in gross amount of variable annuity reserves is reported as a share of total equity in 2007. The sample includes all insurers with at least $1 billion of variable annuity sales in 2007.

	VA liabilities	Change from 2007 to 2008	
Insurer	in 2007 (billion $)	Reserve valuation (%)	Reserves (% of equity)
AXA	140	7.6	125
Metropolitan Life	129	2.9	6
Prudential	122	1.4	13
Voya	121	4.2	42
Hartford	120	2.9	13
AIG	99	0.8	2
Lincoln	97	1.3	15
John Hancock	95	1.8	27
Ameriprise	81	1.0	13
Aegon	63	7.3	29
Pacific Life	56	1.5	13
Nationwide	46	1.7	18
Jackson National	33	3.6	13
Delaware Life	24	3.7	44
Allianz	23	5.3	35
New York Life	19	2.2	2
Genworth	17	0.5	1
Northwestern	12	0.2	0
Ohio National Life	11	2.2	22
Fidelity Investments	10	1.0	8
Security Benefit	10	1.3	12
MassMutual	6	1.7	0
Thrivent Financial	3	0.4	0

Table IV
Estimated Model of Variable Annuity Demand

The random coefficients logit model of demand is estimated by two-step generalized method of moments. The specification includes insurer fixed effects whose coefficients are not reported for brevity. The instruments are log reserve valuation, the share of variable annuity reserves reinsured, and the squares of these variables. Heteroscedasticity-robust standard errors are reported in parentheses. The sample includes all contracts from 2005:1 to 2015:4.

Variable	Mean	Standard deviation
Fee	-3.10	0.22
	(0.17)	(0.08)
Rollup rate	0.18	
	(0.02)	
Investment options	0.07	
	(0.01)	
GLWB	-1.52	
	(2.55)	
A.M. Best rating	0.56	
	(0.14)	
Observations	32,419	

Table V
Estimated Model of Variable Annuity Supply

A model of variable annuity supply is estimated by two-step generalized method of moments. Both the level and the slope of option value with respect to the rollup rate depend on the number of investment options and a dummy for GLWB. The shadow cost of capital depends on the A.M. Best rating, log reserve valuation, the share of variable annuity reserves reinsured, and insurer fixed effects whose coefficients are not reported for brevity. Heteroscedasticity-robust standard errors are reported in parentheses. The sample includes all contracts from 2005:1 to 2015:4.

Variable	Coefficient
Level of option value	
Investment options	0.03
	(0.02)
GLWB	-6.98
	(0.41)
Slope of option value	
Investment options	-0.06
	(0.02)
GLWB	-2.64
	(0.48)
Shadow cost	
A.M. Best rating	-1.23
	(0.31)
Reserve valuation	1.26
	(0.35)
Reserves reinsured	0.42
	(0.27)
Dummies on rollup rate	
0% rollup rate	11.36
	(0.49)
No guaranteed living benefit	46.96
	(0.51)
Observations	32,419

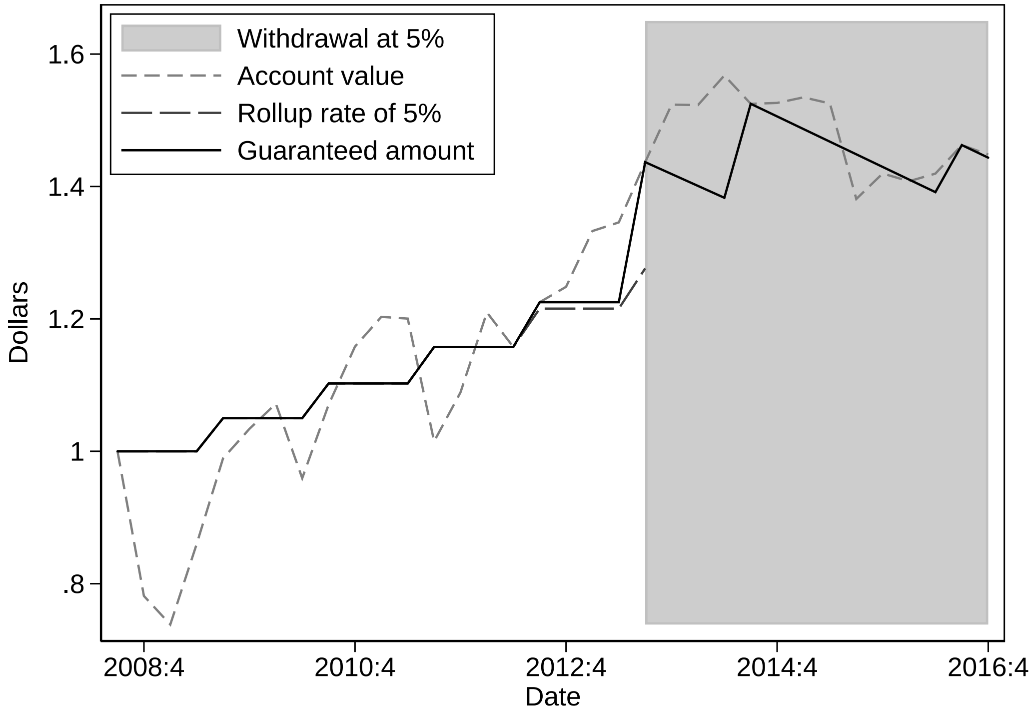

Figure 1. An example of a GLWB. The evolution of account value and the guaranteed amount are shown for MetLife Series VA with a GLWB from 2008:3 to 2016:4. The investment option is the American Funds Growth Allocation Portfolio. The investor is assumed to annually withdraw 5% of the highest guaranteed amount after 2013:3. For simplicity, this example abstracts from the impact of fees on account value and the guaranteed amount.

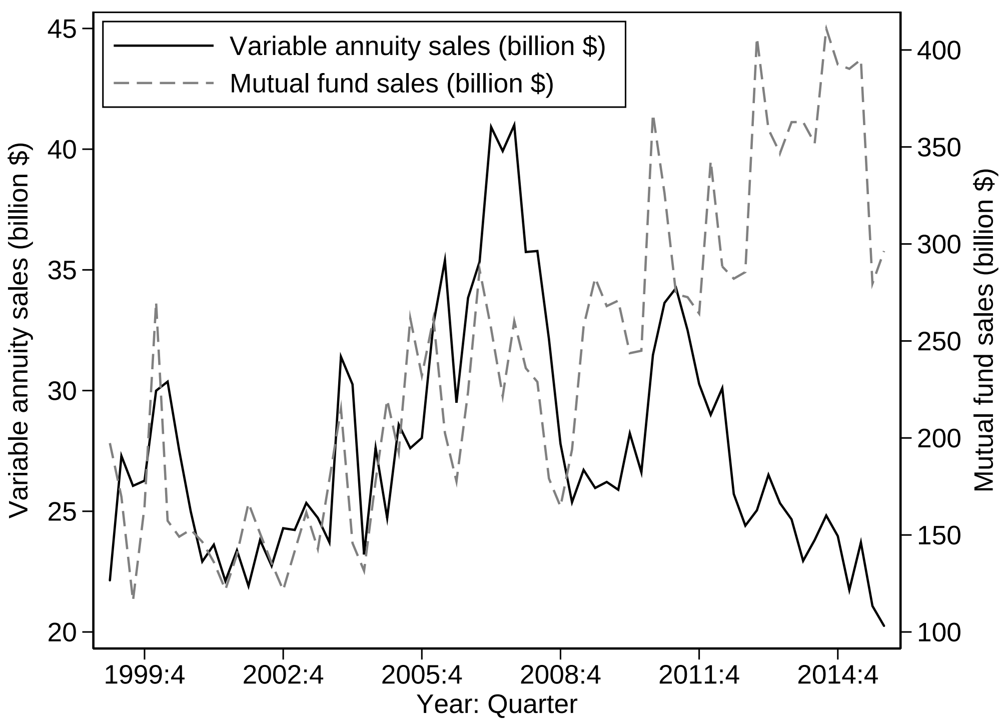

Figure 2. Variable annuity sales. The left axis reports quarterly sales of variable annuities across all contracts from 1999:1 to 2015:4. The right axis reports the aggregate sales of U.S. open-end stock and bond mutual funds (excluding money market funds and funds of funds).

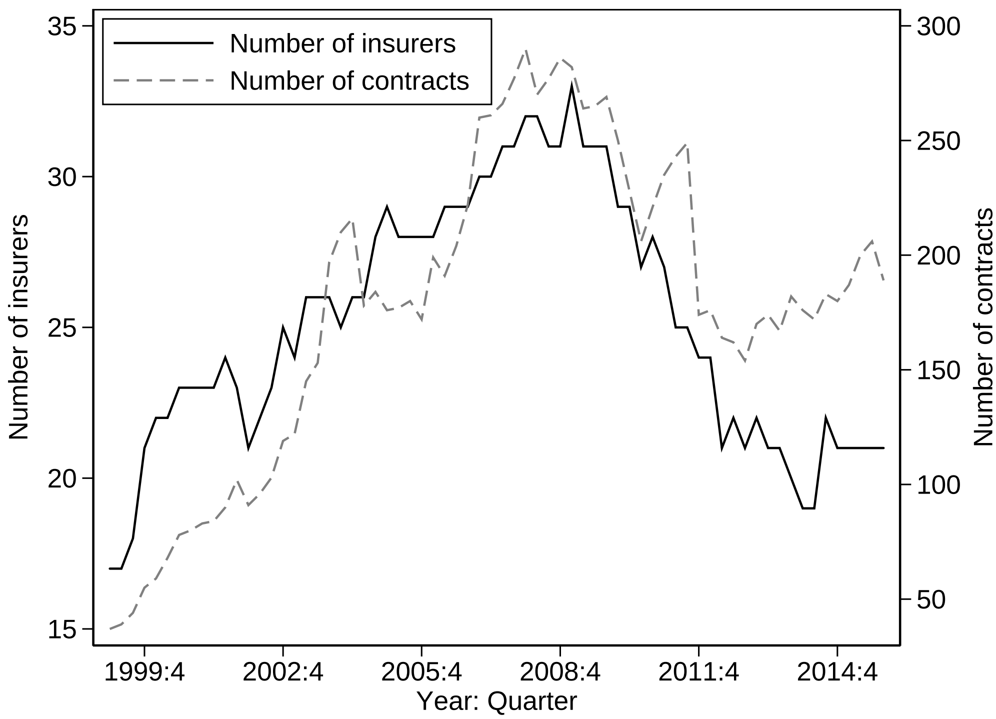

Figure 3. Number of insurers and contracts offering guaranteed living benefits. The sample includes all contracts with guaranteed living benefits from 1999:1 to 2015:4.

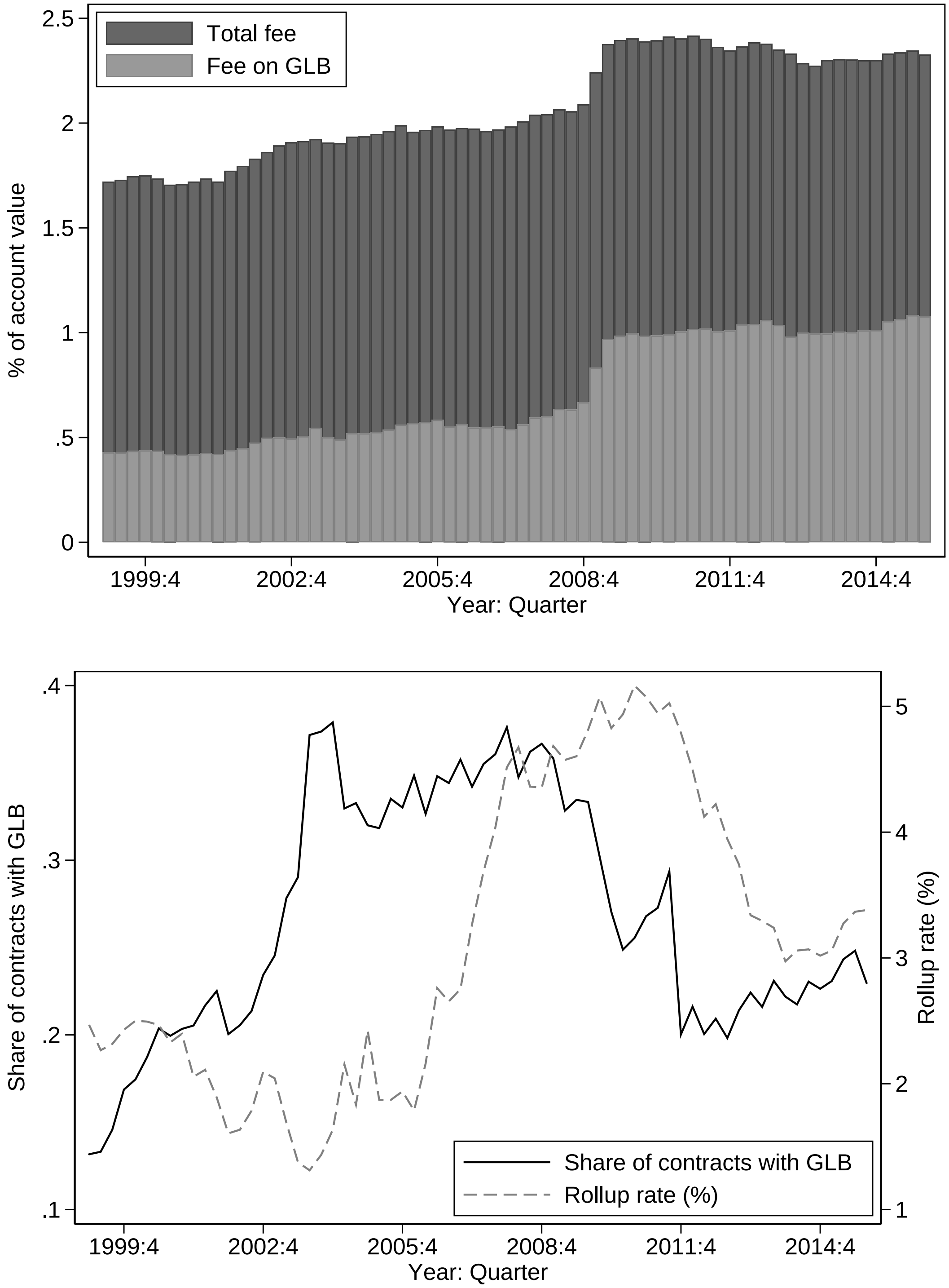

Figure 4. Fees and rollup rates on guaranteed living benefits. The upper panel reports the annual fee on open guaranteed living benefits, averaged across contracts with sales weights. The total annual fee includes the base contract expense. The lower panel reports the rollup rate on open guaranteed living benefits, averaged across contracts with sales weights, and the share of contracts with guaranteed living benefits. The sample includes all contracts with guaranteed living benefits from 1999:1 to 2015:4.

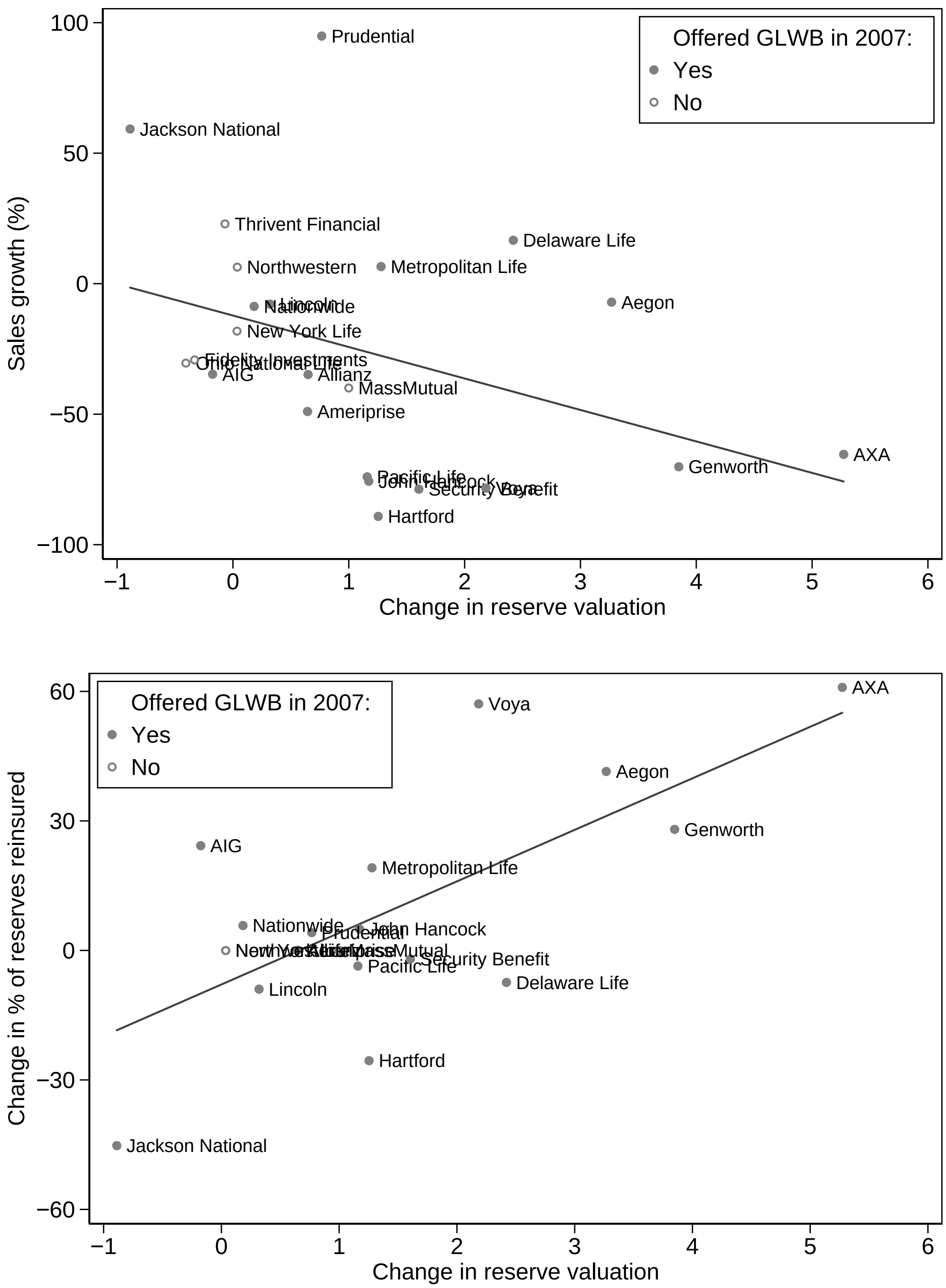

Figure 5. Cross section of insurers during the financial crisis. The upper panel is a scatter plot of sales growth versus the change in the reserve valuation from 2007 to 2010. The lower panel is a scatter plot of the change in the share of reserves reinsured versus the change in the reserve valuation from 2007 to 2010. Both panels report a linear regression line through the scatter points. The sample includes all insurers with at least $1 billion of variable annuity sales in 2007.

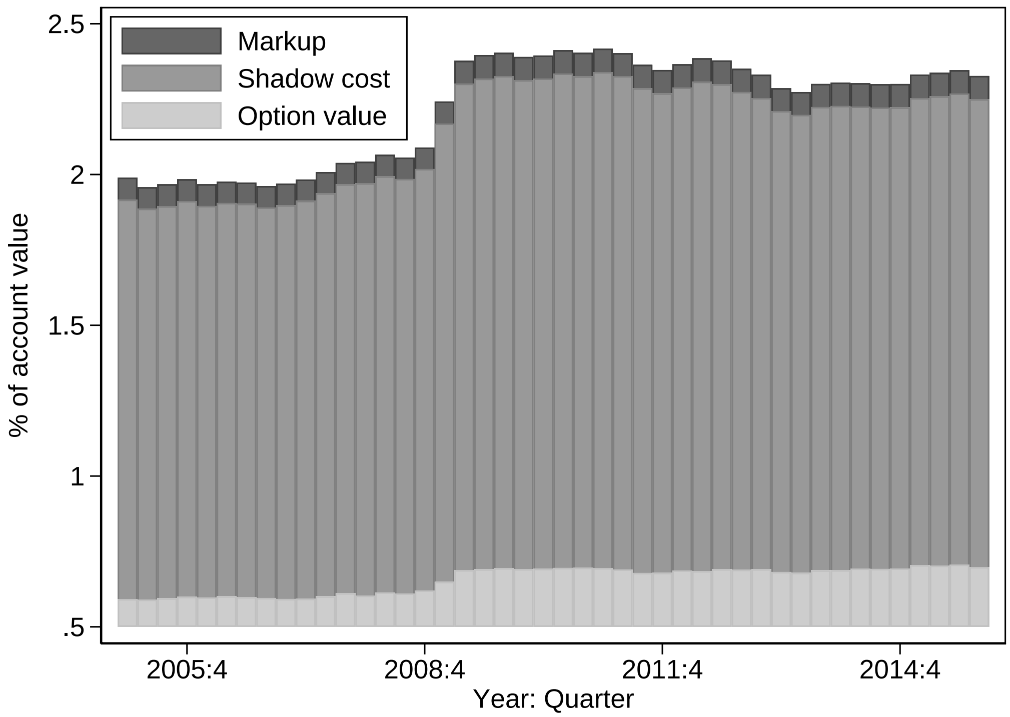

Figure 6. Decomposition of average fees. An estimated model of the variable annuity market is used to decompose the total fee into parts due to the option value, the shadow cost of capital, and the markup above marginal cost. This figure reports a decomposition of the total annual fee, averaged across contracts with sales weights. The sample includes all contracts with guaranteed living benefits from 2005:1 to 2015:4.

Appendix A. A Caution on Interpreting the Rollup Rate

The guaranteed amount at the end of the accumulation period can be written as a sum of the cumulative rollup rate and the payoff of a call option. Thus, we derive a lower bound on fees based only on the rollup rate to assess whether an annual fee such as 1.8% on MetLife Series VA with a GLWB is justified by a rollup rate of 5%. We show that the implied fee based on the rollup rate is actually negative because the time value of money during the withdrawal period more than offsets the high rollup rate during the accumulation period. Therefore, the high fees cannot be explained by the high rollup rate and must instead be attributed to the call option value, market power, or financial frictions.

Following the notation in the paper, let S_t be the mutual fund price per share at time t. Let $M_{t,t+s}$ be a strictly positive stochastic discount factor that discounts a payoff at time $t+s$ to its price at time t. Then the term structure of riskless interest rates is given by the usual pricing formula: $\mathbb{E}_t[M_{t,t+s}] = (1+y_{t,t+s})^{-s}$. That is, $y_{t,t+s}$ is the annually compounded yield on an s-year zero-coupon bond at time t.

Consider a GLWB with an annual fee P per dollar of account value, an annual rollup rate of r, an annual withdrawal rate of w, an accumulation period of T_a years, and a withdrawal period of T_w years. For simplicity, we assume that the withdrawal rate, the accumulation period, and the withdrawal period are all fixed. We also assume that there are no step-ups during the withdrawal period. For a contract issued at time t, the guaranteed amount at the end of the accumulation period at time $t+T_a$ is

$$X_{t,t+T_a} = \max\left\{(1+r)^{T_a}, \frac{S_{t+T_a}}{S_t}\right\} = (1+r)^{T_a} + \underbrace{\max\left\{0, \frac{S_{t+T_a}}{S_t} - (1+r)^{T_a}\right\}}_{\text{call option}}. \tag{A1}$$

For each dollar of account value, the zero-profit condition equates one plus the present value of fees to the present value of guaranteed income:

$$1 + \mathbb{E}_t\left[\sum_{s=1}^{T_a} M_{t,t+s} \frac{PS_{t+s}}{S_t}\right] = 1 + T_a P = \mathbb{E}_t\left[\sum_{s=1}^{T_w} M_{t,t+T_a+s} w X_{t,t+T_a}\right]. \tag{A2}$$

Because $X_{t,t+T_a} \geq (1+r)^{T_a}$, a lower bound on fees based only on the rollup rate is

$$P \geq \frac{1}{T_a}\left(\sum_{s=1}^{T_w} \frac{w(1+r)^{T_a}}{(1+y_{t,t+T_a+s})^{T_a+s}} - 1\right). \tag{A3}$$

This equation shows that the rollup rate in the numerator is offset by the time value of money in the denominator because the guaranteed amount is only payable as annual income over T_w years. We show the empirical relevance of this issue by computing the lower bound on fees, based on the historical zero-coupon Treasury yield curve (Gürkaynak et al., 2007).

Figure A.1 reports the lower bound on fees for an annual rollup rate of 5%, an annual withdrawal rate of 5%, and a withdrawal period of 20 years. To see the sensitivity of the results to the accumulation period, the figure reports the lower bound for an accumulation period of 10 and 20 years. The lower bound on fees is negative for most of the sample period and becomes positive only after 2011:4 for the 20-year accumulation period. This means that the high fees cannot be explained by a rollup rate of 5% and must instead be attributed to the call option value, market power, or financial frictions.

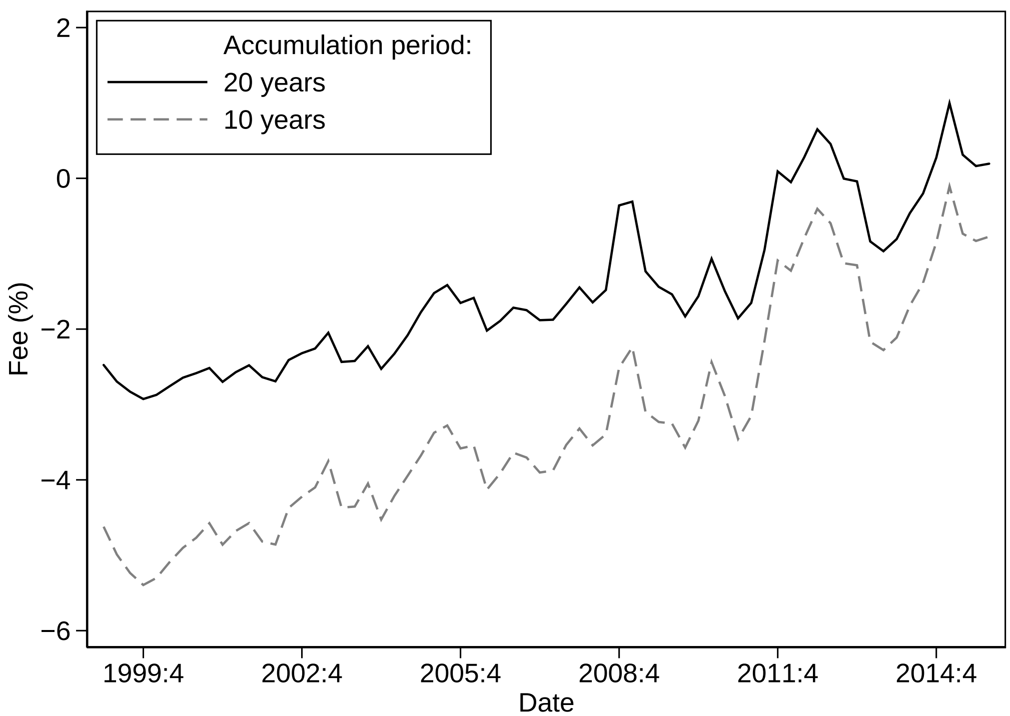

Figure A.1. A lower bound on fees based on the rollup rate. The lower bound is based on an annual rollup rate of 5%, an annual withdrawal rate of 5%, and a withdrawal period of 20 years. The calculation uses an average of the zero-coupon Treasury yield curve within each quarter from 1999:1 to 2015:4, assuming that the yield curve is flat beyond 30 years.

Appendix B. Portfolio of U.S. Life Insurers

We construct monthly returns on a value-weighted portfolio of publicly traded U.S. life insurers that have variable annuity liabilities, based on the following list.

Table B.I
Publicly Traded U.S. Life Insurers

This table reports the first observation for which monthly stock returns are available from January 1999 to December 2017.

Insurer	First observation	
AIG	January	1999
Allstate	January	1999
American National	January	1999
Ameriprise	November	2005
Assurant	March	2004
CIGNA	January	1999
Farm Bureau Life	January	1999
Genworth	June	2004
Hartford	January	1999
Horace Mann Life	January	1999
Kansas City Life	January	1999
Lincoln	January	1999
Metropolitan Life	May	2000
Nationwide	January	1999
Phoenix Life	July	2001
Principal Financial Group	November	2001
Protective Life	January	1999
Prudential	January	2002
Symetra Life	February	2010

Appendix C. Proofs

A. Proof of Proposition 1

This proof covers the case of a multi-product insurer that offers multiple contracts and chooses the fees and the rollup rates, accounting for demand elasticities across contracts. Let bold letters denote vectors corresponding to their scalar counterparts. Let $\mathbf{1}$ be a vector of ones, $\mathbf{I}$ be an identity matrix, and $\operatorname{diag}(\cdot)$ be a diagonal matrix (e.g., $\operatorname{diag}(\mathbf{1}) = \mathbf{I}$). The insurer sets a vector of fees $\mathbf{P}_t$ and rollup rates $\mathbf{r}_t$ to maximize firm value

$$J_t = (\mathbf{P}_t - \mathbf{V}_{t,t})'\mathbf{Q}_t - C_t, \tag{C1}$$

which generalizes equation (11). Substituting equations (7) and (8) into equation (9), the law of motion for statutory capital is

$$K_t = R_{K,t}K_{t-1} + (\mathbf{P}_t - \mathbf{V}_{t,t} - \phi_t\mathbf{V}_{t,t})'\mathbf{Q}_t, \tag{C2}$$

where

$$R_{K,t} = \frac{A_{t-1}}{K_{t-1}}R_{A,t} - \frac{(1+\phi_t)L_{t-1}}{K_{t-1}}\frac{\mathbf{V}'_{t-1,t}\mathbf{Q}_{t-1}}{\mathbf{V}'_{t-1,t-1}\mathbf{Q}_{t-1}} \tag{C3}$$

is the return on statutory capital.

The partial derivative of firm value with respect to the fee is

$$\begin{aligned}
\frac{\partial J_t}{\partial \mathbf{P}_t} =& \frac{\partial(\mathbf{P}_t - \mathbf{V}_{t,t})'\mathbf{Q}_t}{\partial \mathbf{P}_t} + c_t\frac{\partial K_t}{\partial \mathbf{P}_t} \\
=& \mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t}(\mathbf{P}_t - \mathbf{V}_{t,t}) + c_t\left(\mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t}(\mathbf{P}_t - \mathbf{V}_{t,t} - \phi_t\mathbf{V}_{t,t})\right) \\
=& (1+c_t)\mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t}((1+c_t)(\mathbf{P}_t - \mathbf{V}_{t,t}) - c_t\phi_t\mathbf{V}_{t,t}).
\end{aligned} \tag{C4}$$

The optimal fee satisfies

$$\frac{\partial J_t}{\partial \mathbf{P}_t} = 0 \Leftrightarrow \mathbf{P}_t + \left(\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t}\right)^{-1}\mathbf{Q}_t = \frac{1 + c_t(1+\phi_t)}{1+c_t}\mathbf{V}_{t,t}. \tag{C5}$$

Equation (13) follows from the definition of semi-elasticity of demand to the fee.

The partial derivative of firm value with respect to the rollup rate is

$$
\begin{aligned}
\frac{\partial J_t}{\partial \mathbf{r}_t} =& \frac{\partial (\mathbf{P}_t - \mathbf{V}_{t,t})'\mathbf{Q}_t}{\partial \mathbf{r}_t} + c_t \frac{\partial K_t}{\partial \mathbf{r}_t} \\
=& -\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}\mathbf{Q}_t + \frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t}(\mathbf{P}_t - \mathbf{V}_{t,t}) \\
&+ c_t \left(-(1+\phi_t)\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}\mathbf{Q}_t + \frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t}(\mathbf{P}_t - \mathbf{V}_{t,t} - \phi_t \mathbf{V}_{t,t}) \right) \\
=& -(1 + c_t(1+\phi_t))\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}\mathbf{Q}_t + \frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t}((1+c_t)(\mathbf{P}_t - \mathbf{V}_{t,t}) - c_t\phi_t \mathbf{V}_{t,t}) \\
=& -(1 + c_t(1+\phi_t))\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}\mathbf{Q}_t - (1+c_t)\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t}\left(\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{P}_t}\right)^{-1}\mathbf{Q}_t,
\end{aligned}
\tag{C6}
$$

where the last line follows from substituting equation (C4). At an interior optimum, the rollup rate satisfies

$$
\frac{\partial J_t}{\partial \mathbf{r}_t} = 0 \Leftrightarrow -\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t}\left(\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{P}_t}\right)^{-1}\mathbf{Q}_t = \frac{1 + c_t(1+\phi_t)}{1+c_t}\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}\mathbf{Q}_t. \tag{C7}
$$

Because $\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}$ is a diagonal matrix, we can rewrite this equation as

$$
-\text{diag}(\mathbf{Q}_t)^{-1}\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t}\left(\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{P}_t}\right)^{-1}\mathbf{Q}_t = \frac{1 + c_t(1+\phi_t)}{1+c_t}\frac{\partial \mathbf{V}_{t,t}'}{\partial \mathbf{r}_t}\mathbf{1}. \tag{C8}
$$

Equation (15) follows from the definition of semi-elasticities of demand to the fee and the rollup rate.

The left side of equations (C5) and (C8) correspond to the left side of equations (25) and (26) for a multi-product insurer. For the random coefficients logit model, we denote the vector of demand for all contracts that an insurer sells as

$$
\mathbf{Q}_t = \int \mathbf{q}_t(\alpha_P)\, dF(\alpha_P). \tag{C9}
$$

The partial derivative of demand with respect to the vector of fees is

$$
\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{P}_t} = \int \alpha_P(\text{diag}(\mathbf{q}_t(\alpha_P)) - \mathbf{q}_t(\alpha_P)\mathbf{q}_t(\alpha_P)')\, dF(\alpha_P). \tag{C10}
$$

The partial derivative of demand with respect to the vector of rollup rates is

$$
\frac{\partial \mathbf{Q}_t'}{\partial \mathbf{r}_t} = \int \alpha_r(\text{diag}(\mathbf{q}_t(\alpha_P)) - \mathbf{q}_t(\alpha_P)\mathbf{q}_t(\alpha_P)')\, dF(\alpha_P). \tag{C11}
$$

Thus, the estimated model of variable annuity demand in Table IV directly implies the left side of equations (C5) and (C8).

B. *A Constant Semi-Elasticity Demand Function*

We assume constant semi-elasticities of demand in Corollary 1. Before its proof, we give an example of a demand function with constant semi-elasticities of demand to show that the assumption is compatible with the oligopolistic market structure. Let the semi-elasticities of demand for contracts sold by insurer i to the fee and the rollup rate of insurer j be $\epsilon_P(i,j) = -\frac{\partial \log(Q_{i,t})}{\partial P_{j,t}}$ and $\epsilon_r(i,j) = \frac{\partial \log(Q_{i,t})}{\partial r_{j,t}}$, respectively. The demand function

$$\log(Q_{i,t}) = \alpha_i - \sum_{j=1}^{I} \epsilon_P(i,j) P_{j,t} + \sum_{j=1}^{I} \epsilon_r(i,j) r_{j,t} \tag{C12}$$

has constant semi-elasticities of demand. The budget constraint $\sum_{i=1}^{I} P_{i,t} Q_{i,t} = 1$ implies that the semi-elasticities of demand must satisfy the restrictions

$$Q_{i,t} - \sum_{j=1}^{I} P_{j,t} Q_{j,t} \epsilon_P(j,i) = 0, \tag{C13}$$

$$\sum_{j=1}^{I} P_{j,t} Q_{j,t} \epsilon_r(j,i) = 0. \tag{C14}$$

C. *Proof of Corollary 1*

The partial derivative of the shadow cost of capital with respect to the reserve valuation is

$$\begin{aligned} \frac{\partial \lambda_t}{\partial (1+\phi_t) V_{t-1,t}} &= - \frac{\phi_t L_{t-1}}{(1+c_t)^2 K_{t-1} V_{t-1,t-1}} \frac{\partial c_t}{\partial K_t} \\ &= \frac{\phi_t L_{t-1}}{(1+c_t)^2 K_{t-1} V_{t-1,t-1}} \frac{\partial^2 C_t}{\partial K_t^2} > 0. \end{aligned} \tag{C15}$$

Differentiating equation (13) for the optimal fee with respect to the reserve valuation, we have

$$\frac{\partial P_t}{\partial (1+\phi_t) V_{t-1,t}} = \frac{\partial \lambda_t}{\partial (1+\phi_t) V_{t-1,t}} V_{t,t} > 0. \tag{C16}$$

Differentiating equation (15) for the optimal rollup rate with respect to the reserve valuation, we have

$$0 = \frac{\partial \lambda_t}{\partial (1+\phi_t)V_{t-1,t}} \frac{\partial V_{t,t}}{\partial r_t} + \lambda_t \frac{\partial^2 V_{t,t}}{\partial r_t^2} \frac{\partial r_t}{\partial (1+\phi_t)V_{t-1,t}}. \tag{C17}$$

Rearranging, we have

$$\frac{\partial r_t}{\partial (1+\phi_t)V_{t-1,t}} = -\left(\lambda_t \frac{\partial^2 V_{t,t}}{\partial r_t^2}\right)^{-1} \frac{\partial \lambda_t}{\partial (1+\phi_t)V_{t-1,t}} \frac{\partial V_{t,t}}{\partial r_t} < 0. \tag{C18}$$

By the chain rule, the partial derivative of sales with respect to the reserve valuation is

$$\frac{\partial Q_t}{\partial (1+\phi_t)V_{t-1,t}} = \frac{\partial Q_t}{\partial P_t} \frac{\partial P_t}{\partial (1+\phi_t)V_{t-1,t}} + \frac{\partial Q_t}{\partial r_t} \frac{\partial r_t}{\partial (1+\phi_t)V_{t-1,t}} < 0. \tag{C19}$$