

The Fragility of Market Risk Insurance

RALPH S.J. KOIJEN and MOTOHIRO YOGO

ABSTRACT

Variable annuities, which package mutual funds with minimum return guarantees over long horizons, accounted for \$1.5 trillion or 35% of U.S. life insurer liabilities in 2015. Sales decreased and fees increased during the global financial crisis, and insurers made guarantees less generous or stopped offering guarantees to reduce risk exposure. These effects persist in the low-interest rate environment after the global financial crisis, and variable annuity insurers suffered large equity drawdowns during the COVID-19 crisis. We develop and estimate a model of insurance markets in which financial frictions and market power determine pricing, contract characteristics, and the degree of market completeness.

THE TRADITIONAL ROLE OF LIFE insurers is to insure idiosyncratic risk through products such as life annuities, life insurance, and health insurance. With the secular decline of private defined benefit and government pension

Ralph S.J. Koijen is at the University of Chicago Booth School of Business, NBER, and CEPR. Motohiro Yogo is at Princeton University, NBER, 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. We thank Adam Xu and Zhen Ye for assistance on constructing data from Morningstar Annuity Intelligence. We thank Don Noh, Jihong Song, and Haiyue Yu for proofreading. For comments and discussions, we thank Naoki Aizawa, Mark Flannery, Victoria Ivashina, Arvind Krishnamurthy, Emanuel Mönch, Borghan Narajabad, Theo Nijman, Anna Paulson, Richard Rosen, and Donghwa Shin. We also thank seminar participants at Banque de France, 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, the 2017 DNB/Riksbank Macroprudential Conference, the 2017 SITE Workshop on Financial Regulation, the 2017 NBER Conference on Financial Market Regulation, the 2017 ICPM-Netspar Discussion Forum, the 2017 IMF Conference on Monetary, Financial, and Prudential Policy Interactions in the Post-Crisis World, the 2017 NBER Insurance Meeting, and the 2018 Macroeconomics of Pensions and Retirement Financing Conference. This paper is based upon work supported by the National Science Foundation under grant 1727049 and the Julis-Rabinowitz Center for Public Policy and Finance. Koijen acknowledges financial support from the Center for Research in Security Prices at the University of Chicago and the Fama Research Fund at the University of Chicago Booth School of Business. The authors have no conflicts of interest to disclose as defined by *The Journal of Finance*'s disclosure policy.

Correspondence: Motohiro Yogo, Princeton University, Department of Economics, Julis Romo Rabinowitz Building, Princeton, NJ 08544; e-mail: myogo@princeton.edu.

DOI: 10.1111/jofi.13118

© 2022 the American Finance Association

plans around the world, life insurers are increasingly taking on the role of insuring market risk through minimum return guarantees. In the United States, 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 fixed 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 (Chan and Kogan (2002); Dumas (1989)) 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 maturities. 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 global financial crisis, many insurers including Aegon, Allianz, AXA, Delaware Life, John Hancock, and Voya suffered large increases in variable annuity reserves ranging from 27% to 125% of total equity. In June 2009, Hartford was bailed out by the Troubled Asset Relief Program 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, which pose a challenge for insurers in the low interest rate environment after the global financial crisis. Consequently, the stock returns of U.S. life insurers have significantly negative exposure to long-term bond returns after the global financial crisis (Hartley, Paulson, and Rosen (2017)).

The COVID-19 crisis has again exposed the fragility of variable annuity insurers, which suffered among the lowest stock returns in the overall financial sector that includes banking. From January 2 to April 2, 2020, the equity drawdown, which is the maximum decrease in the cumulative stock return, was -51% for a value-weighted portfolio of U.S. variable annuity insurers. AIG, Brighthouse Financial, and Lincoln National suffered the largest equity drawdowns, each exceeding -65% . The equity drawdown on variable annuity insurers was substantially larger than -34% for the S&P 500 index and -43% for the subset of financial sector stocks in the S&P 500 index. To put the pain

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)).

of variable annuity insurers into perspective, the equity drawdown on the airline industry was -62% .

Given their size and potential risk, variable annuities are essential to 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 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 then decreased 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 global financial crisis, insurers made the minimum return guarantees less generous or stopped offering guarantees to reduce risk exposure. The share of contracts with minimum return guarantees decreased from 36% in 2007:4 to 20% in 2011:4. In the cross section of insurers, sales decreased more for insurers who suffered larger increases in the valuation of existing liabilities. These insurers moved their variable annuity reserves 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 a model of insurance markets in which financial frictions and market power are important determinants of pricing, contract characteristics, and the degree of market completeness. Insurers compete in an oligopolistic market by choosing the fee and the rollup rate, where the latter 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 reduces risk exposure by not only raising the fee but also by reducing the rollup rate. When the shadow cost of capital is sufficiently high, the insurer stops offering minimum return guarantees to avoid additional risk exposure.

Variable annuity demand 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. The average demand elasticity to the fee decreased after the global financial crisis, while the average demand elasticity to the rollup rate remained nearly constant. However, the lower

demand elasticity to the fee cannot fully explain the higher fees after the global financial crisis.

Through the insurer's optimality conditions, we decompose the time variation in fees and rollup rates. The 34 basis point increase in the average annual fee from 2007:4 to 2009:2 reflects an increase of 1 basis point in the markup, 16 basis points in the option value, and 17 basis points in the shadow cost of capital. Thus, financial frictions are just as important as the option pricing channel for explaining the increase in fees during the global financial crisis. The increase in the shadow cost of capital partly explains the decrease in rollup rates. More importantly, contracts with generous guarantees became too capital intensive because the option value was more sensitive to the rollup rate, which explains why insurers stopped offering minimum return guarantees.

Previous research shows that equity and interest risk exposures from variable annuities are one of the most important sources of risk for U.S. life insurers. Variable annuity insurers became financially constrained and significantly reduced the prices of fixed annuities and life insurance during the global financial crisis (Koijen and Yogo (2015)). Variable annuity insurers hold less liquid bonds, and the common equity risk exposure through variable annuities makes these insurers more vulnerable to fire-sale dynamics in bond markets (Ellul et al. (2018)). Variable annuity insurers have negative duration and negative convexity, making them vulnerable to a low interest rate environment (Hartley, Paulson, and Rosen (2017)). Consequently, they continue to register high in systemic risk indicators long after the global financial crisis (Acharya, Philippon, and Richardson (2017)). These papers study the consequences of variable annuities 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 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 private and government pension plans. 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. Our theory of insurance markets explains how insurers choose prices and contract characteristics and why they may not offer insurance, addressing the same questions as Rothschild and Stiglitz (1976). However, we focus on financial frictions and market power instead of informational frictions as the important determinants of market equilibrium. Our theory could apply to other insurance markets in which insurers bear significant aggregate risk such as long-term care insurance (Cutler (1996)) and catastrophe insurance of climate or cyber risk.

Our work also relates to the mutual fund literature. Previous research shows 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)). Calvet et al. (2019) find that minimum return guarantees can increase participation in equity risk for Swedish households that do not hold stocks and mutual funds. 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 aggregate 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 differentiated product demand system for the variable annuity market at the contract level. Section V estimates a model of variable supply to quantify the importance of financial frictions. Section VI concludes.

I. Institutional Background

We start with an example of an actual product to illustrate 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 together 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 global financial crisis.

A. An Example of a Variable Annuity

A variable annuity is a mutual fund that is sold through an insurer with longevity insurance and a potential tax advantage. For an additional fee, the insurer offers an optional minimum return guarantee on the mutual fund. Thus, a variable annuity with a minimum return guarantee is a retail financial product that packages a mutual fund with a long-maturity put option on the mutual fund. To illustrate how variable annuities work, we start with an example of an actual product.

MetLife Investors USA Insurance Company (2008) offers a variable annuity 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 a Guaranteed Lifetime Withdrawal Benefit (GLWB). MetLife Series VA has an annual base contract expense of 1.3% of account value, and a GLWB has an annual fee of 0.5% of

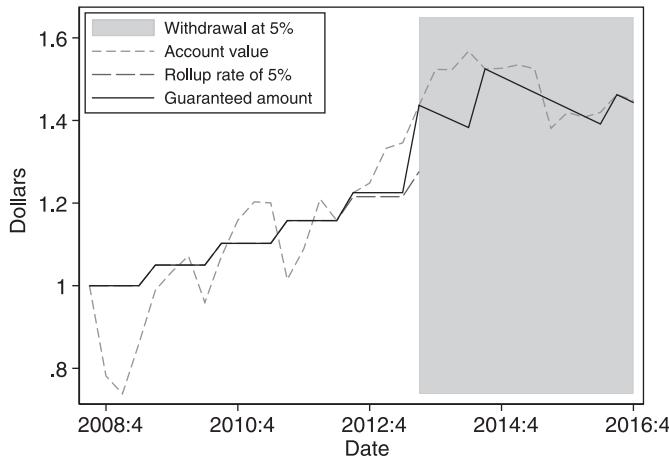


Figure 1. An example of a guaranteed living withdrawal benefit. The evolution of account value and the guaranteed amount are shown for MetLife Series VA with a GLWB from 2008:3 to 2016:4. The investment option is the American Funds Growth Allocation Portfolio. The investor is assumed to annually withdraw 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.

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.

To understand the GLWB, we first describe a stand-alone investment in the mutual fund and the withdrawals that it would enable for retirement income. 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. This behavior describes an investor who invests in a mutual fund five years before 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. A GLWB has an annual rollup rate of 5% before 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, a GLWB is a put option on the mutual fund that locks in each 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 market risk.

When 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. A 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 contract feature based on the historical term structure of interest rates.

A GLWB is the most common type of guaranteed living benefit. The three other types of guaranteed living benefits are a Guaranteed Minimum Withdrawal Benefit (GMWB), a Guaranteed Minimum Income Benefit (GMIB), and a 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 the 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, the insurer offers 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 do not go into the details of guaranteed death benefits in this paper.

Even without minimum return guarantees, variable annuities may be attractive to investors because of a potential tax advantage in nonqualified 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 the 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 variable annuity fees only if the accumulation period is sufficiently 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% pretax 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:

$$RBC = \frac{\text{Assets} - \text{Reserves}}{\text{Required capital}}. \quad (1)$$

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 by Junus and Motiwalla (2009), Actuarial Guideline 43 determines the reserve value of variable annuities since December 2009, and the C-3 Phase II regulatory standard determines the contribution of variable annuities to required capital since December 2005. Actuarial Guideline 43 is a higher reserve requirement than its precursor Actuarial Guideline 39, so insurers were given a phase-in period through 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 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 (called CTE 70). 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 (called CTE 90).

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, both reserves and required capital increase in an adverse scenario like the global financial crisis, which 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 contrast to the conditional tail expectation under Actuarial Guideline 43, generally accepted accounting principles (GAAPs) allow insurers to record variable annuity reserves at market value. As a result, variable annuity

reserves under the statutory accounting principles could increase relative to those under GAAP after a period of high volatility (Credit Suisse (2012)). Moreover, an insurer that implements a hedging program under GAAP could actually increase the volatility of accounting equity under the statutory accounting principles. For these reasons, insurers have an incentive for captive reinsurance with a less regulated and unrated reinsurer within the same insurance group (Koijen and Yogo (2016)). Through captive reinsurance, insurers could increase risk-based capital or implement a hedging program under GAAP.

The insurer's risk management could also have an economic risk constraint, which works similarly to a risk-based capital constraint. For example, let ϵ be a multiplicative shock to the leverage ratio with a cumulative distribution function $F(\epsilon)$, which arises from a risk mismatch between assets and liabilities. Consider a value-at-risk constraint under which the probability that assets cover liabilities must exceed a threshold:

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

We can rewrite this constraint as

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

which is similar to risk-based capital defined in equation (1). An insurer with more conservative risk management has higher $F^{-1}(\kappa)$ due to a higher κ or lower risk reflected in the cumulative distribution function.

As a consequence of the global 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 together with a risk-based capital constraint to affect variable annuity supply.

C. Reasons for Risk Mismatch

In theory, insurers could use derivatives to hedge uncertainty in the valuation of minimum return guarantees. In practice, insurers do not fully hedge variable annuity risk for various economic and institutional reasons (Drexler, Plestis, and Rosen (2017); Ellul et al. (2018); Koijen and Yogo (2017); Sen (2019)).

Insurers may not be able to fully hedge because the minimum return guarantees have longer maturities than traded options. Insurers are exposed to unexpected changes in implied volatility if they attempt to hedge the 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)). In addition to basis risk, derivatives could expose insurers to counterparty risk. Although collateral could reduce counterparty risk, it would increase the cost of the hedging programs (Berends and King (2015)). 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 a 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 associations (Lee, Mayers, and Smith (1997)). Another reason that insurers may not want to hedge is that existing regulation does not properly reward hedging of market value. Insurers report accounting equity under the statutory accounting principles at the operating company level and under 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 the statutory accounting principles or GAAP (Credit Suisse (2012)).

Whether insurers target market or accounting equity depends on whether economic frictions (e.g., value-at-risk constraint) or regulatory frictions (i.e., risk-based capital constraint) are more important. Captive reinsurance is more efficient than hedging for the purposes of reducing regulatory frictions. Section III shows that insurers used reinsurance to move variable annuity reserves off balance sheet during the global financial crisis, suggesting a role for regulatory frictions.

D. 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. The market value of equity decreases with unexpected decreases in interest rates, especially when interest rates are low. Consistent with this view, Hartley, Paulson, and Rosen (2017) find that U.S. life insurers' stock returns have significantly negative exposure to long-term bond returns in the low interest rate environment after the global financial crisis. In contrast, property and casualty insurers and U.K. life insurers (who do not have variable annuities) do not have such interest risk exposure.

We update the finding in Hartley, Paulson, and Rosen (2017) with a longer sample, which provides supportive evidence that variable annuity insurers have difficulty managing interest rate risk in the low interest rate environment. We construct monthly returns on a value-weighted portfolio of publicly traded U.S. variable annuity insurers, which are listed in Appendix B. We regress excess portfolio returns, relative to the one-month T-bill rate, on 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 1.36, and the 10-year bond beta is -0.01 and statistically insignificant. On average, insurers do not have significant interest risk exposure,

Table I
Risk Exposure of Variable Annuity Insurers

We construct monthly returns on a value-weighted portfolio of publicly traded U.S. variable annuity insurers, 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 one-month T-bill rate, on excess stock market returns and excess 10-year Treasury bond returns. Heteroskedasticity-robust standard errors are reported in parentheses. The sample period is January 1999 through December 2017.

Factor	By Subsample			
	1999 to 2007	2008 to 2009	2010 to 2017	
Stock market return	1.36 (0.19)	0.56 (0.15)	2.56 (0.22)	1.11 (0.08)
10-year bond return	-0.01 (0.32)	-0.38 (0.29)	1.14 (0.66)	-1.28 (0.16)
Alpha (%)	-0.22 (0.46)	0.35 (0.47)	-1.14 (1.70)	0.41 (0.29)
Observations	228	108	24	96

controlling for the stock market exposure. However, the 10-year bond beta varies over time when we break the sample into three subperiods: precrisis (1999 to 2007), financial crisis (2008 to 2009), and postcrisis (2010 to 2017). In the postcrisis subsample, the 10-year bond beta is -1.28 with a *t*-statistic greater than 7. Thus, unexpected decreases in interest rates are bad news for U.S. life insurers during this low interest rate environment. A coefficient near -1 implies that the negative duration gap is close to the duration of the 10-year Treasury bond.

Another possible source of interest rate risk is the franchise value, or the present value of future insurance business (Shi (2021)). If the franchise value falls with interest rates, an insurer's stock price could fall by more than what is implied by the interest risk mismatch on its current balance sheet. Nevertheless, a 10-year duration gap is surprisingly large. This raises the question of why insurers do not increase the duration of their assets to reduce the duration gap. Kojien and Yogo (2022b) argue that insurers optimally hold corporate bonds, which expose them to interest risk mismatch because corporate bonds have a shorter maturity distribution than Treasury bonds.

E. Stock Returns during the COVID-19 Crisis

The COVID-19 crisis has again exposed the fragility of variable annuity insurers, as we show in Panel A of Figure 2. From January 2 to April 2, 2020, the equity drawdown, which is the maximum decrease in the cumulative stock return, was -51% for a value-weighted portfolio of U.S. variable annuity insurers.² This equity drawdown was substantially larger than -34% for the S&P

² We compute the portfolio return as a buy-and-hold portfolio with fixed weights as of December 31, 2019. A continuously rebalanced portfolio would imply decreasing weights for insurers who suffered the lowest returns, even though the market value of their liabilities presumably increased.

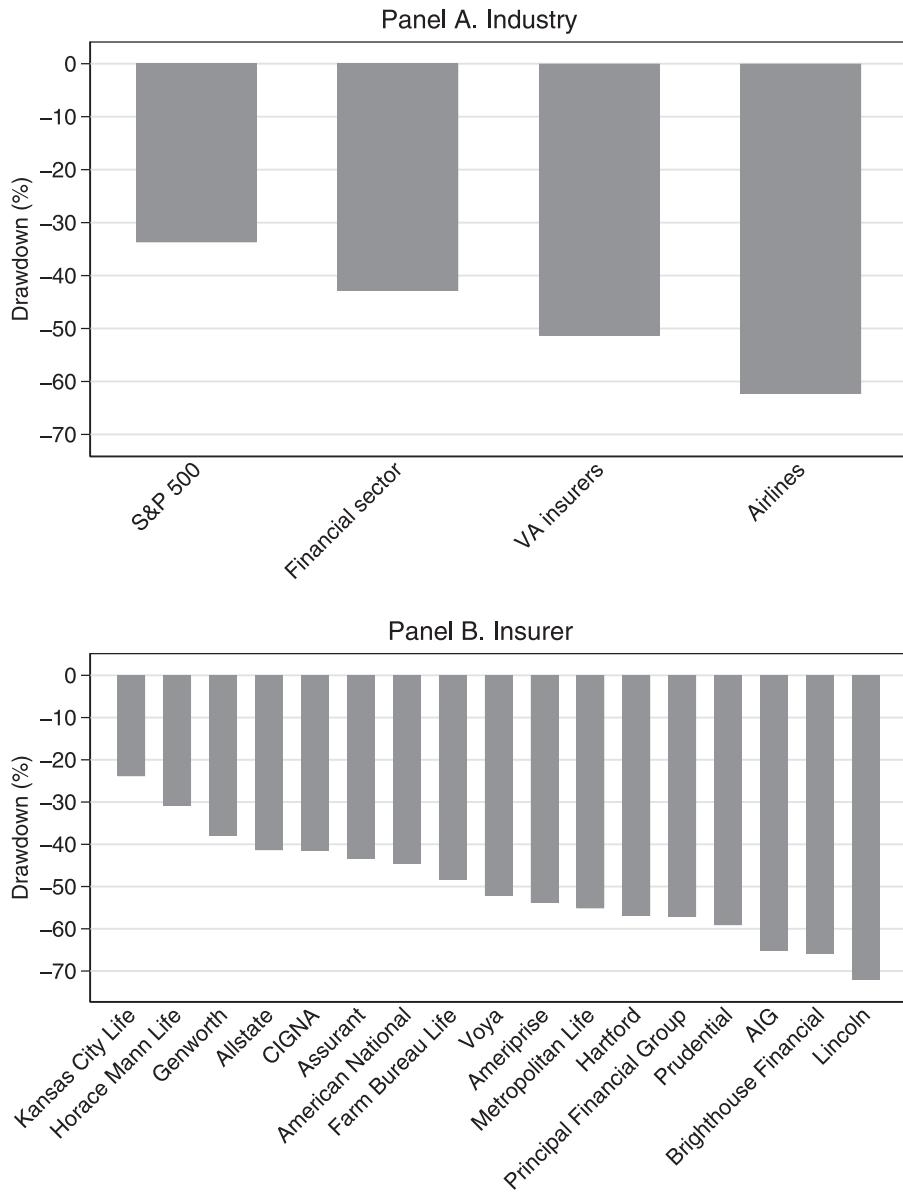


Figure 2. Equity drawdowns during the COVID-19 crisis. Panel A shows the equity drawdowns on the S&P 500 index, the Financial Select Sector SPDR Fund, a value-weighted portfolio of U.S. variable annuity insurers, and the U.S. Global Jets ETF. Panel B shows the equity drawdowns on individual insurers who make up the portfolio in Panel A. The equity drawdowns are based on stock returns from January 2 to April 2, 2020.

500 index and -43% for the Financial Select Sector SPDR Fund, which is the subset of financial sector stocks in the S&P 500 index. In fact, the equity drawdown on U.S. variable annuity insurers was only slightly smaller than -62% for the U.S. Global Jets exchange-traded fund (ETF), which tracks the U.S. airline industry. Panel B of Figure 2 shows the equity drawdowns on individual insurers who make up the portfolio in Panel A. AIG, Brighthouse Financial, and Lincoln National suffered the largest equity drawdowns, each exceeding -65% .

II. 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.³ For each guaranteed living benefit, the key characteristics are the 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 a 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 insurers' annual financial statements, which are filed with the NAIC (National Association of Insurance Commissioners (2005–2015)). General Interrogatories Part 2 of Table 9.2 of the financial statements reports the total related account value, the gross amount of variable annuity reserves, and the reinsurance reserve credit on variable annuities. The total related account value is the market value of the mutual funds.

³ We use assets under management by subaccount from Morningstar (2016b) to compute a measure of investment options that adjusts for the nonuniform distribution of assets across subaccounts within a contract. Our measure is the inverse of the Herfindahl index over the subaccount shares within each contract, which equals the number of investment options when the subaccounts are uniformly distributed.

The gross amount of variable annuity reserves is the accounting value of the minimum return guarantees net of hedging programs. We define *variable annuity liabilities* as total related account value plus gross amount of variable annuity reserves minus reinsurance reserve credit on variable annuities. For each insurer, we define its *reserve valuation* as the ratio of gross amount of variable annuity reserves to total related account value. The reserve valuation measures the value of the minimum return guarantees per dollar of underlying mutual funds. In the cross section, the reserve valuation is higher for insurers who have sold more generous guarantees. In the time series, the reserve valuation increases when the stock market falls, interest rates fall, or volatility rises. We define the *reinsurance share of variable annuities* as the ratio of reinsurance reserve credit on variable annuities to gross amount of variable annuity reserves.

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)).

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 the summary statistics that only require the Morningstar data, we use a longer sample that starts in 1999:1.

B. Aggregate Facts

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 varied between 34% and 41% of total liabilities as their value fluctuates with the market value of the mutual funds. 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 decreased from 44 in 2008 to 38 in 2015.

As we discuss above, the reserve valuation (i.e., the ratio of gross amount of variable annuity reserves to total related account value) measures the value of the minimum return guarantees per dollar of underlying mutual funds. 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 before the global financial crisis.

Table III lists the top insurers ranked by their variable annuity liabilities in 2007. Eight of these insurers (Aegon, Allianz, AXA, Delaware Life, Hartford,

Table II
Summary of the Variable Annuity Market

Variable annuity liabilities are total related account value plus gross amount of variable annuity reserves minus reinsurance reserve credit on variable annuities. The reserve valuation is the ratio of gross amount of variable annuity 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

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 ranges from 92% to 97%. For five of the eight insurers, the increases in variable annuity reserves are a significant share of total equity, ranging from 29% to 125%.

The nine insurers with the largest variable annuity liabilities in Table III coincide almost perfectly with the nine insurers who suffered the largest equity drawdowns in Figure 2. AXA and John Hancock (part of Manulife Financial) in Table III are foreign insurers who are not part of Figure 2, which focuses on U.S. life insurers. Brighthouse Financial was spun off from Metropolitan Life in 2017, so it was part of Metropolitan Life at the time of Table III in 2007. Thus, Principal Financial Group is the only insurer that breaks the perfect correspondence of the top nine between Table III and Figure 2. Because the COVID-19 crisis began three years after the first draft of this paper, we view Figure 2 as out-of-sample evidence. The long maturity of the minimum return guarantees means that variable annuities continue to be an important source of risk and that the insurance sector remains fragile.

Figure 3 shows quarterly sales of variable annuities across all contracts from 1999:1 to 2015:4. Sales grew robustly from \$22 billion in 2005:1 to a peak of \$41 billion in 2007:4. Sales then decreased during the global financial crisis to \$27 billion in 2009:2, partly rebounded to \$34 billion in 2011:2, and decreased again to \$20 billion in 2015:4. For comparison, the figure also 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 hence

Table III
Top Insurers by Variable Annuity Liabilities

Variable annuity liabilities are total related account value plus gross amount of variable annuity reserves minus reinsurance reserve credit on variable annuities. The reserve valuation is the ratio of gross amount of variable annuity 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.

Insurer	VA Liabilities in 2007 (Billion \$)	Change from 2007 to 2008	
		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

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 decrease in variable annuity sales after 2008 is partly explained by insurers who stopped offering minimum return guarantees. Figure 4 shows the number of insurers and contracts offering minimum return guarantees from 1999:1 to 2015:4. Eleven insurers stopped offering minimum return guarantees from 2008 to 2015, during which six insurers stopped selling variable annuities altogether (see Table II). Thus, some insurers chose to remain in the variable annuity market but stopped offering minimum return guarantees. Without minimum return guarantees, variable annuities are essentially mutual funds with longevity insurance and a potential tax advantage.

Panel A of Figure 5 shows the average annual fee on open (for sale) minimum return guarantees from 1999:1 to 2015:4. The increase in fees during

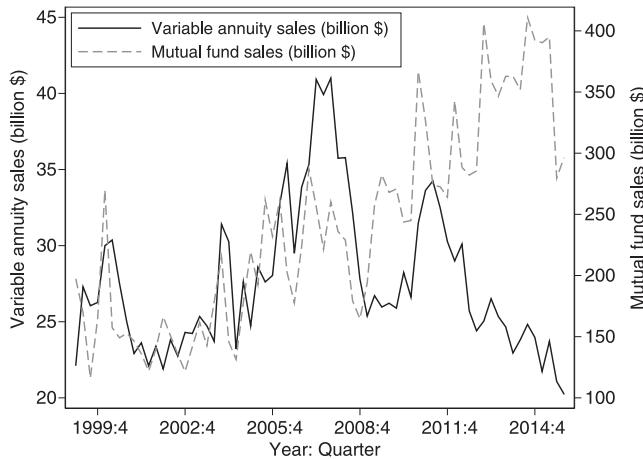


Figure 3. Variable annuity sales. The left axis shows quarterly sales of variable annuities across all contracts from 1999:1 to 2015:4. The right axis shows the aggregate sales of U.S. open-end stock and bond mutual funds (excluding money market funds and funds of funds).

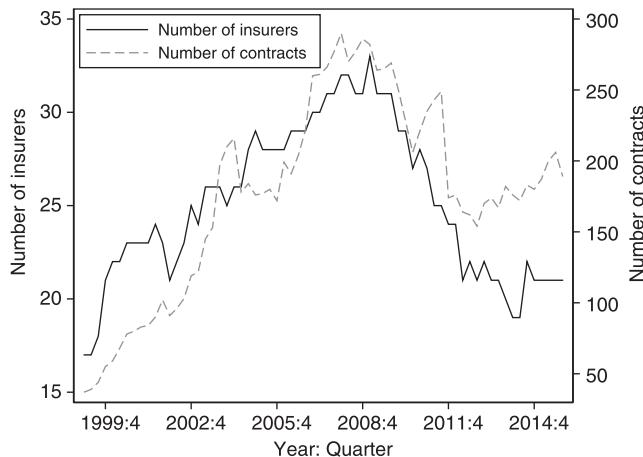


Figure 4. Number of insurers and contracts with minimum return guarantees. The sample includes all contracts with minimum return guarantees from 1999:1 to 2015:4.

the global financial crisis coincides with the decrease in sales, suggesting an important role for a supply shock. The average annual fee on minimum return guarantees 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 minimum return guarantees was 1.08% (2.33% including the base contract expense) in 2015:4.

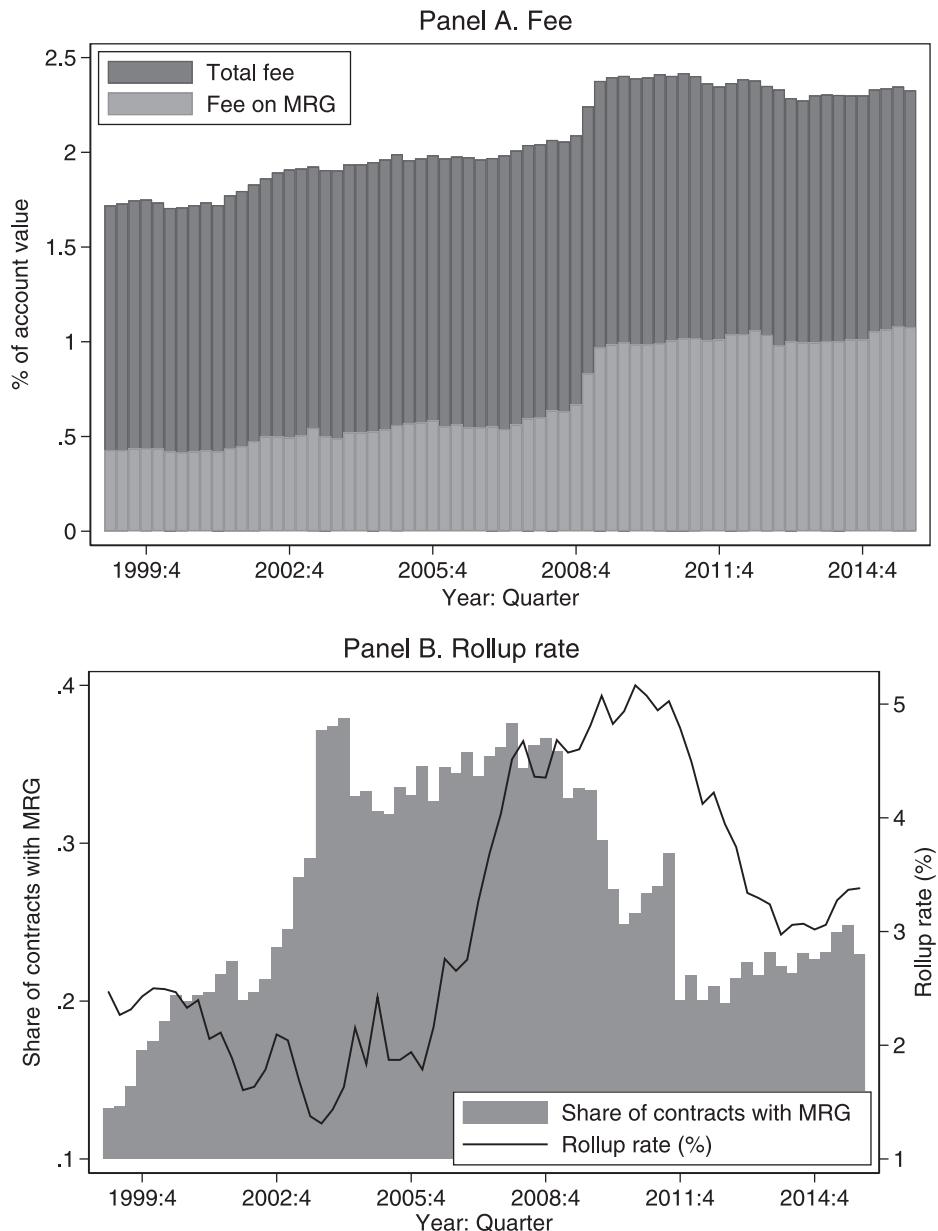


Figure 5. Fees and rollup rates on minimum return guarantees. Panel A shows the annual fee on open minimum return guarantees, averaged across contracts with sales weighting. The total annual fee includes the base contract expense. Panel B shows the rollup rate on open minimum return guarantees, averaged across contracts with sales weighting, and the share of contracts with minimum return guarantees. The sample includes all contracts with minimum return guarantees from 1999:1 to 2015:4.

Panel B of Figure 5 summarizes the rollup rates on open contracts from 1999:1 to 2015:4. Conditional on offering a minimum return guarantee, 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 global financial crisis and decreased only after 2011. However, the share of contracts with minimum return guarantees decreased after the global financial crisis from 36% in 2007:4 to 20% in 2011:4, consistent with Figure 4. Thus, many insurers responded to the global financial crisis through the extensive margin by not offering minimum return guarantees rather than through the intensive margin by reducing rollup rates.

To summarize Figures 3 to 5, variable annuity sales decreased, fees increased, and many insurers stopped offering minimum return guarantees in response to the global financial crisis. This evidence is consistent with a supply shock as a consequence of tightening risk-based capital and economic risk constraints. Following the discussion in Section I, two factors could explain why variable annuity supply did not fully recover long after the global financial crisis. First, insurers may have been more cautious because of the difficulty of managing interest rate risk in the low interest rate environment. Second, Actuarial Guideline 43 increased the capital requirements for variable annuities. Despite the enormous attention that Actuarial Guideline 43 received in the industry, its impact 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 discuss 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 global 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 reduced 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, which explains the decrease in sales.

To formalize this narrative, we extend the insurance pricing model in Kojen and Yogo (2015) to contract design. The insurer competes in an oligopolistic market by choosing the fee and the rollup rate subject to a risk-based capital or an economic risk constraint. The insurer could reduce the rollup rate or stop offering minimum return guarantees to reduce risk exposure. From the investor's perspective, the insurance market becomes incomplete when insurers stop offering minimum return guarantees. Thus, we develop a more complete theory of the supply side of insurance markets that explains pricing, contract characteristics, and the degree of market completeness.

To simplify the exposition, 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.

A. Variable Annuity Market

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 . Therefore, 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 as a lump sum 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

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

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 the variable annuity market, so the insurance market is incomplete when $r_t = -1$.

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

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

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 is 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 assume that the partial derivatives of option value have the conventional signs. Furthermore, we assume that the first two partial derivatives of option value with respect to the rollup rate are positive.

ASSUMPTION 1: *The option value decreases in the mutual fund price, decreases in the riskless interest rate, and increases in volatility. The first two partial derivatives of option value with respect to the rollup rate are positive (i.e., $\partial V_{t,t}/\partial r_t > 0$ and $\partial^2 V_{t,t}/\partial r_t^2 > 0$).*

We also do not need parametric assumptions about variable annuity demand. We assume that demand is continuous, continuously differentiable, strictly decreasing in the fee, and strictly increasing in the rollup rate. In an

oligopolistic market, the demand for a contract depends on the fees and the rollup rates of all competing contracts. To simplify the notation, we denote the demand for a contract as Q_t with the understanding that it depends on the fees and the rollup rates of all competing contracts.

An institutional feature of the variable annuity market is that the rollup rate is positive (i.e., $r_t \geq 0$) or $r_t = -1$ in the case of mutual funds without minimum return guarantees. Insurers do not offer a variable annuity with a negative rollup rate presumably because investors have a psychological aversion to “negative interest rates.” To model this institutional feature, we assume that the rollup rate is constrained to be in the set $\mathcal{R} = \{-1\} \cup (-0.01, \infty)$. A lower bound slightly less than zero ensures that $r_t = 0$ is an interior solution and that $r_t = -1$ is the only corner solution.⁴

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. The insurer’s equity fluctuates because of risk mismatch between assets and minimum return guarantees in the general account.

We 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

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

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

⁴ By making the rollup rate a continuous choice, the model is not designed to match the discreteness of the empirical distribution of rollup rates. Among contracts with a GLWB, the frequency of a 0% rollup rate is similar to that of a 5% rollup rate. This fact motivates our assumption that a 0% rollup rate is an interior solution.

The general account assets evolve according to

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

where $R_{A,t}$ is an exogenous gross asset return in period t . Current assets are the gross return on previous assets plus the fees on new contracts. As we discuss in Section I, insurers do not fully hedge variable annuity risk for various economic and institutional reasons. 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

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

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 as equity minus required capital that is proportional to liabilities:

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

For simplicity, we assume that $\phi_t > 0$ is an exogenous parameter that does not depend on the fee or the rollup rate.⁵ 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. Through equation (8), 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. Moreover, financial frictions make equity issuance costly. We model the cost of financial frictions through a cost function:

$$C_t = C(K_t). \quad (10)$$

⁵ This formulation of statutory capital as a difference rather than a ratio is for mathematical convenience in the derivations that follow. Kojien 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 $(A_t - L_t)/(\phi_t L_t) \geq 1$.

This cost function 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.⁶

An alternative interpretation of the cost function is that the insurer has an economic risk constraint, such as the value-at-risk constraint in Section I. As a consequence of the global 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 are consistent with the evidence in Section II.

D. Optimal Fee and Rollup Rate

The insurer chooses the fee P_t and the rollup rate $r_t \in \mathcal{R}$ to maximize firm value, which is the profit minus the cost of financial frictions:

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

We could have specified firm value as the present value of profits as in Kojien and Yogo (2015), but we opt for the simpler presentation because the key insights do not depend on dynamics.

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 semielasticity of demand to the fee as $\epsilon_t^P = -\partial \log(Q_t)/\partial P_t$ and to the rollup rate as $\epsilon_t^r = \partial \log(Q_t)/\partial r_t$. We also define the marginal cost of capital as

$$c_t = -\frac{\partial C_t}{\partial K_t} > 0. \quad (12)$$

The marginal cost of capital 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*

$$P_t = \frac{1}{\epsilon_t^P} + \underbrace{\lambda_t V_{t,t}}_{marginal\ cost}, \quad (13)$$

⁶ See Ellul et al. (2015) and Kojien 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.

where the shadow cost of capital is

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

At an interior optimum, the rollup rate satisfies

$$\frac{\epsilon_t^r}{\epsilon_t^P} = \lambda_t \frac{\partial V_{t,t}}{\partial r_t}. \quad (15)$$

Otherwise, $r_t = -1$ is optimal.

The optimal fee (13) is the sum of two terms. The first term is the markup that is inversely related to the demand elasticity 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 risk weight ϕ_t .

We clarify two potential points of confusion in relation to Kojien and Yogo (2015). First, equation (13) in this paper is identical to Kojien and Yogo (2015, equation (21)). The reason that they may appear different is that ϵ_t^P is the semielasticity of demand in this paper, while it is the full elasticity of demand in the earlier work. Semielasticity is the natural formulation in this paper because the fee is expressed as a percentage of account value. Second, equation (13) implies that marginal cost *decreases* in the shadow cost of capital if $\phi_t < 0$. In Kojien and Yogo (2015), the prices of fixed annuities decreased during the global financial crisis because statutory reserve regulation effectively made $\phi_t < 0$.

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 slope of the option value with respect to the rollup rate, which increases in the rollup rate by Assumption 1.

Equation (C.6) in Appendix C clarifies why the demand elasticity to the fee ϵ_t^P appears in equation (15). The intuition is that the insurer earns a higher markup when the demand elasticity to the fee is lower, so variable annuities are less capital intensive per unit sold. This slackness in statutory capital allows the insurer to offer a higher rollup rate to increase demand and profits.

The intuition for the right side of equation (15) is simple for the logit model of demand, in which case the left side is constant. 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 reducing the rollup rate on new contracts. When the shadow cost of capital is sufficiently high, the insurer offers mutual funds without 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.

Proposition 1 provides a narrative for the aggregate facts in Figures 3 to 5. Insurers suffered an adverse shock to risk-based capital during the global financial crisis and 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 reduced 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, which explains the decrease in sales.

E. Evidence from the Cross Section of Insurers

We provide reduced-form evidence from the cross section of insurers that supports Proposition 1. A measurement challenge is that the shadow cost of capital is not directly observed. However, the reserve valuation is a relevant empirical proxy that we describe 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:

$$\begin{aligned} \frac{\partial \lambda_t}{\partial (1 + \phi_t)V_{t-1,t}} &= - \frac{\phi_t L_{t-1}}{(1 + c_t)^2 K_{t-1} V_{t-1,t-1}} \frac{\partial c_t}{\partial K_t} \\ &= \frac{\phi_t L_{t-1}}{(1 + c_t)^2 K_{t-1} V_{t-1,t-1}} \frac{\partial^2 C_t}{\partial K_t^2} > 0. \end{aligned} \quad (16)$$

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 to Section V.

Depending on the contract characteristics of existing liabilities, different insurers experienced different shocks to the reserve valuation during the global financial crisis. Insurers who sold more generous guarantees before the global financial crisis suffered larger increases in the reserve valuation than those that sold less generous guarantees. Moreover, insurers who sold more generous guarantees could have made risk management more conservative after the global 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.

Panel A of Figure 6 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 are six insurers (Fidelity Investments, MassMutual, New York Life, Northwestern, Ohio National, and Thrivent Financial) that did not offer a GLWB in 2007, which tends to be the most generous guarantee among guaranteed living benefits. The reserve valuation did not change much for these insurers because they sold less generous guarantees.

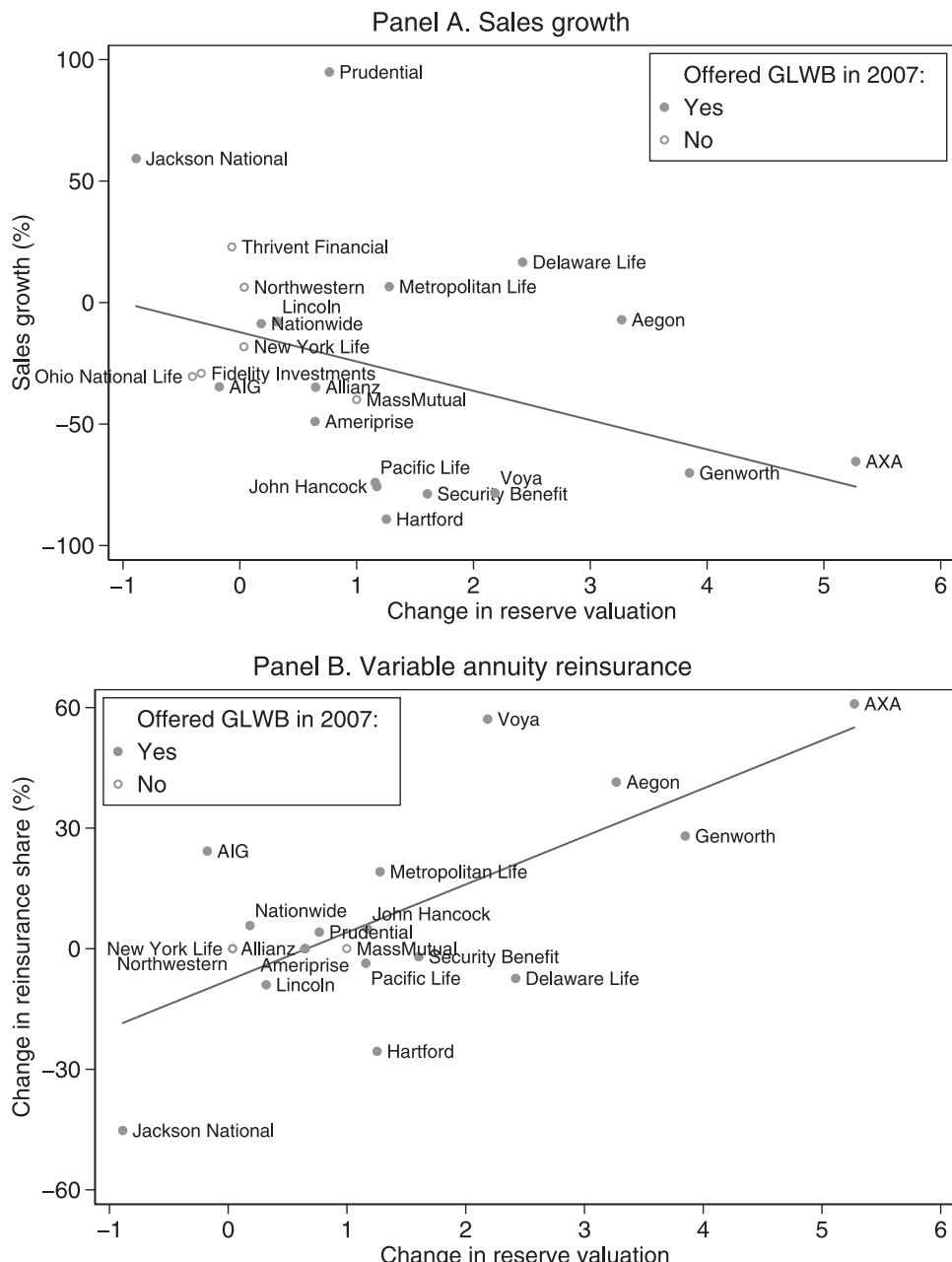


Figure 6. Cross section of insurers during the global financial crisis. Panel A is a scatter plot of sales growth versus the change in the reserve valuation from 2007 to 2010. Panel B is a scatter plot of the change in the reinsurance share of variable annuities versus the change in the reserve valuation 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.

As we discuss in Section I, insurers could increase risk-based capital through captive reinsurance. If insurers who suffered large increases in the reserve valuation were in fact constrained, they had an incentive to move variable annuity reserves off balance sheet through reinsurance. Panel B of Figure 6 is a scatter plot of the change in the reinsurance share of variable annuities versus the change in the reserve valuation from 2007 to 2010. The linear regression line shows that the change in the reinsurance share of variable annuities is positively related to the change in the reserve valuation. On the top right are insurers like AXA and Genworth that reinsured more variable annuity liabilities as they suffered large increases in the reserve valuation. This evidence suggests an important role for a risk-based capital constraint rather than an economic risk constraint.

IV. Estimating Variable Annuity Demand

Variation in fees and rollup rates 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 the fee into the markup and 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, Nijman, and Werker (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 who 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, Levinsohn, and Pakes (1995)), which is a tractable and microfounded 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 observed 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 . For an investor with a realized coefficient α_P on the fee, the indirect utility from buying contract i in period t is

$$U_{i,t} = \alpha_P P_{i,t} + \alpha_r r_{i,t} + \beta' \mathbf{x}_i + \xi_{i,t} + \varepsilon_{i,t}, \quad (17)$$

where $\varepsilon_{i,t}$ is a logit error drawn from a type 1 extreme value distribution. The probability that the investor buys contract i in period t is

$$q_{i,t}(\alpha_P) = \frac{\exp(-\alpha_P P_{i,t} + \alpha_r r_{i,t} + \beta' \mathbf{x}_{i,t} + \xi_{i,t})}{1 + \sum_{j=1}^I \exp(-\alpha_P P_{j,t} + \alpha_r r_{j,t} + \beta' \mathbf{x}_{j,t} + \xi_{j,t})}, \quad (18)$$

where I is the total number of contracts across all insurers. The denominator of equation (18) captures how demand for a contract depends on the fees and characteristics of all 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, which is independently and identically distributed over time. The coefficient on the fee α_P is lognormally distributed, ensuring a positive demand elasticity. Integrating equation (18) over the distribution of investors, the market share for contract i in period t is

$$Q_{i,t} = \int q_{i,t}(\alpha_P) dF(\alpha_P). \quad (19)$$

The semielasticity of demand to the fee for contract i in period t is

$$\epsilon_{i,t}^P = \frac{1}{Q_{i,t}} \int \alpha_P q_{i,t}(\alpha_P) (1 - q_{i,t}(\alpha_P)) dF(\alpha_P). \quad (20)$$

Through equation (13), the markup is inversely related to the demand elasticity to the fee.

The demand elasticity to the fee varies over time through changing contract characteristics, interacting with the distribution of random coefficients on the fee. When the fee increases or the rollup rate decreases, the more price elastic investors substitute into competing contracts with lower fees or the outside asset. Thus, the demand elasticity to the fee decreases because the remaining investors are less price elastic on average. In Panel A of Figure 5, fees could have increased after the global financial crisis not only because of an increasing shadow cost of capital but also because of a decreasing demand elasticity.

The estimation sample comprises 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 minimum return guarantee. We assign a type of minimum return guarantee to each contract following the procedure in Section II. The rollup rate is 0% for contracts with minimum return guarantees but no step ups and -100% for contracts without minimum return guarantees. This treatment of the rollup rate is consistent with the model of variable annuity supply in Section III, in which we assume 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, a dummy for whether the contract offers a GLWB, and

share class fixed effects. The number of investment options captures the menu or the complexity of options within contracts (Célérier and Vallée (2017)). A GLWB is the most common type and tends to be the most generous guarantee among guaranteed living benefits. The share class (i.e., A, B, C, I, L, O, or X) determines whether there is an initial sales charge or a surrender charge for early withdrawal. For example, B is the most common share class, and it has a surrender charge but no sales charge. The share class also determines the commission schedule for the investment advisor who sells the variable annuity. For example, investment advisors do not earn a commission on the I share class.

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 the outside asset (i.e., mutual funds) if they are concerned about the stability of the insurance sector. The unobserved characteristic $\xi_{i,t}$ in equation (18) 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 the outside asset.

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 reinsurance share of variable annuities 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 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 equation (16), the reserve valuation $(1 + \phi_t)V_{t-1,t}$ is a relevant instrument that is positively related to the shadow cost of capital. 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 $V_{t,t}$. Even if the option value of existing liabilities were collinear with the option value of new contracts because the contract characteristics happen to be identical, $1 + \phi_t$ is another source of variation in the reserve valuation that could break the collinearity. Recall that $1 + \phi_t$ represents the ratio of reserve to market value under Actuarial Guideline 43. Therefore, the reserve valuation is an accounting value that does not coincide with the market value or the investors' valuation that enters demand. As we discuss 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 the contract characteristics in our specification.

We have a similar motivation for the reinsurance share of variable annuities as an instrument. Koijen and Yogo (2016) show that most reinsurance is with less regulated and unrated reinsurers within the same insurance group, which relaxes regulatory constraints and reduces tax liabilities. Thus, insurers reinsurance a higher share of variable annuity reserves when marginal cost is high, leading to a positive relation between the reinsurance share of variable annuities and marginal cost for a given insurer. We assume that the reinsurance share of variable annuities 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.

C. Estimation Methodology

We estimate the random coefficients logit model of variable annuity demand by two-step generalized method of moments. Let $\mathbf{z}_{i,t}$ be a vector of instruments that includes the reserve valuation, the reinsurance share of variable annuities, and the square of these instruments to help identify the variance of the random coefficients on the fee. The moment condition is

$$\mathbb{E}[\xi_{i,t} | \mathbf{z}_{i,t}, \mathbf{x}_{i,t}] = 0. \quad (21)$$

Because $\xi_{i,t}$ is not analytical, we compute it numerically as follows.

We rewrite the market share (19) as

$$Q_{i,t}(\mu_P, \sigma_P^2, \delta_t) = \int \frac{\exp(-e^\nu P_{i,t} + \delta_{i,t})}{1 + \sum_{j=1}^I \exp(-e^\nu P_{j,t} + \delta_{j,t})} dF(\nu), \quad (22)$$

where

$$\delta_{i,t} = \alpha_r r_{i,t} + \beta' \mathbf{x}_{i,t} + \xi_{i,t}, \quad (23)$$

$$\nu \sim \mathbb{N}\left(\mu_P - \frac{\sigma_P^2}{2}, \sigma_P^2\right), \quad (24)$$

and $\delta_t = (\delta_{1,t}, \dots, \delta_{I,t})'$. Starting with an initial guess of the parameters $(\mu_P, \sigma_P^2, \alpha_r, \beta)$ and the vector of mean utility $\delta_t(1)$, the estimation proceeds as follows.

1. Iterate on the following equation until convergence:

$$\delta_t(n+1) = \delta_t(n) + \log(\mathbf{Q}_t) - \log(\mathbf{Q}_t(\mu_P, \sigma_P^2, \delta_t(n))), \quad (25)$$

computing the market share (22) numerically through a simulation with 500 draws. Let $\delta_t(N)$ be the converged vector of mean utility.

2. Compute $\xi_{i,t} = \delta_{i,t}(N) - \alpha_r r_{i,t} - \beta' \mathbf{x}_{i,t}$ and evaluate the objective function corresponding to moment condition (21).
3. Stop if the objective function is minimized. Otherwise, update the parameters $(\mu_P, \sigma_P^2, \alpha_r, \beta)$ by Newton's method and go back to step 1.

D. Estimated Model of Variable Annuity Demand

Table IV reports the estimated parameters for the random coefficients logit model of variable annuity demand. The estimate of the mean parameter is $\mu_P = 3.37$ with a standard error of 0.13. The estimate of the standard deviation parameter is $\sigma_P = 0.30$ with a standard error of 0.05. The coefficient on the rollup rate is 0.18 with a standard error of 0.01. The signs of the coefficients confirm that demand decreases in the fee and increases in the rollup rate.

Demand also increases in the number of investment options, the GLWB dummy, and the A.M. Best rating. The coefficient on the number of investment options is 0.18 with a standard error of 0.01. The coefficient on the GLWB dummy is 17.02 with a standard error of 2.64. The coefficient on the I share class is -13.82 with a standard error of 2.34, which means that the I share class has lower demand than the B share class. The coefficient on the A.M. Best rating, which is standardized, is 0.73 with a standard error of 0.10. This means that demand increases by 73% per one-standard-deviation increase in the rating.

We compute the semielasticity of demand for each contract through equation (20). For contracts with minimum return guarantees, the semielasticity of demand to the fee has a mean of 16.4 and a standard deviation of 0.8 across contracts in 2007:4. A semielasticity of 16.4 means that demand decreases by 16.4% per one basis point increase in the fee. The dynamics of sales and fees during the global financial crisis are especially important for identifying the demand elasticity. Sales decreased and fees increased sharply during the global financial crisis, especially for contracts with minimum return guarantees. The average semielasticity of demand to the fee falls to 15.4 in 2009:2 and ultimately to 14.8 in 2015:4. As we discuss above, the demand elasticity to the fee varies over time through changing contract characteristics, interacting with the distribution of random coefficients on the fee.

Table IV
Estimated Model of Variable Annuity Demand

The random coefficients logit model of variable annuity demand is estimated by two-step generalized method of moments. The random coefficient on the fee is parameterized as $\log(\alpha_P) \sim N(\mu_P - \sigma_P^2/2, \sigma_P^2)$, where the table reports estimates of μ_P and σ_P . For reporting purposes, the rollup rate is orthogonalized with respect to the number of investment options, the GLWB dummy, and share class fixed effects. B is the omitted category for the share class fixed effects. The specification includes insurer fixed effects whose coefficients are not reported for brevity. The instruments are log reserve valuation, the reinsurance share of variable annuities, and the squares of these variables. Heteroskedasticity-robust standard errors are reported in parentheses. The sample includes all contracts from 2005:1 to 2015:4.

Variable	Mean	SD
Fee	3.37 (0.13)	0.30 (0.05)
Rollup rate	0.18 (0.01)	
Investment options	0.11 (0.01)	
GLWB	17.02 (2.64)	
Share class		
A	-9.01 (1.60)	
C	2.01 (0.62)	
I	-13.82 (2.34)	
L	4.99 (1.05)	
O	-5.60 (1.03)	
X	3.86 (0.82)	
A.M. Best rating	0.73 (0.10)	
Observations	32,419	

The semielasticity of demand to the fee is 6 for S&P 500 index funds in 2000 (Hortaçsu and Syverson (2004)).⁷ The demand for variable annuities may be more elastic than that for index funds for several reasons. First, variable annuity investors are wealthier and less risk averse than the average household (Brown and Poterba (2006)). Second, variable annuity investors may be more inclined to shop around because a variable annuity is a large investment that is costly to reverse. Third, financial frictions could interact with broker incentives to increase the demand elasticity. After the global financial crisis,

⁷ This semielasticity is implied by an asset-weighted average fee of 32.2 basis points in 2000 (Hortaçsu and Syverson (2004, table II)) and a marginal cost of 16 basis points (Hortaçsu and Syverson (2004, column D of table III)).

insurers made the minimum return guarantees less generous or stopped offering guarantees to reduce risk exposure, which changed the composition of new contracts toward those for which brokers earn lower commissions. Thus, brokers may have had weaker incentives to sell variable annuities after the global financial crisis. The role of brokers in determining demand elasticities is an important area for future research, which recent work explores using new data on broker commissions (Bhattacharya, Illanes, and Padi (2020); Barbu (2021); Egan, Ge, and Tang (2020)).

For contracts with minimum return guarantees, the semielasticity of demand to the rollup rate has a mean of 0.18 across contracts in 2007:4. A semielasticity of 0.18 means that demand increases by 18% per one-percentage-point increase in the rollup rate. The demand elasticity to the rollup rate is nearly constant across contracts and over time. The coefficient on the rollup rate is constant in our specification, which implies that the demand elasticity to the rollup rate is not sensitive to changing contract characteristics.

Our baseline specification limits the random coefficients to the fee. For robustness, we estimate 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 coefficients converges to zero or has large standard errors that indicate that the richer model is poorly identified. Thus, we find evidence for heterogeneity across investors in the demand elasticity to the fee but not to the rollup rate. The identification problem arises from the fact that the variation in market shares can only identify a limited covariance structure for the random coefficients.

V. Estimating Variable Annuity Supply

A. Empirical Specification

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

$$\log \left(P_{i,n,t} - \frac{1}{\epsilon_{i,n,t}^P} \right) = \log(V_{i,n,t}) + \log(\lambda_{n,t}). \quad (26)$$

This equation decomposes 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

$$\log \left(\frac{\epsilon_{i,n,t}^r}{\epsilon_{i,n,t}^P} \right) - \log(\lambda_{n,t}) - \log \left(\frac{\partial V_{i,n,t}}{\partial r_{i,n,t}} \right) = \omega_{i,n,t} \geq 0. \quad (27)$$

At an interior solution (i.e., $\omega_{i,n,t} = 0$), the marginal benefit of a higher rollup rate through demand is equal to the marginal cost of a higher rollup rate

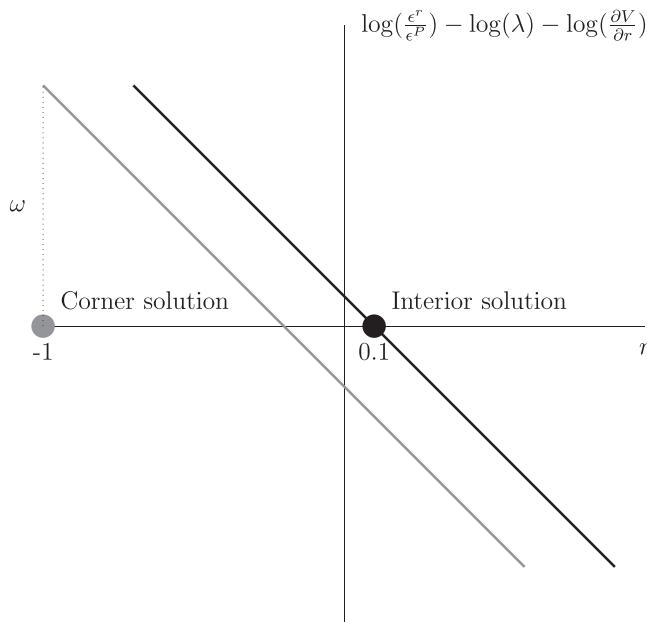


Figure 7. Interior versus corner solution for the rollup rate. When the relative demand elasticities minus the shadow cost of capital is high, the optimal rollup rate is at an interior solution (i.e., $r = 0.1$) as illustrated by the black line. When the relative demand elasticities minus the shadow cost of capital is low, the optimal rollup rate is at a corner solution (i.e., $r = -1$) as illustrated by the gray line.

through financial frictions. The black line in Figure 7, which represents the left side of equation (27) as a function of the rollup rate, illustrates an interior solution. The line is downward sloping because of the third term on the left side of equation (27). The negative of the slope of the option value with respect to the rollup rate decreases in the rollup rate by Assumption 1.

At a corner solution, the marginal benefit of a higher rollup rate through demand is greater than the marginal cost of a higher rollup rate through financial frictions. The gray line in Figure 7 illustrates a corner solution when the relative demand elasticities minus the shadow cost of capital is low. The insurer would like to reduce risk exposure by reducing the rollup rate, but the constraint prevents the insurer from choosing a negative rollup rate that would be the unconstrained optimum. Therefore, the insurer offers mutual funds without minimum return guarantees, so that equation (27) holds at $r_{i,n,t} = -1$ with $\omega_{i,n,t} > 0$.

Because the same shadow cost of capital enters both equations (26) and (27), 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 high fees and low rollup rates across all contracts that an insurer offers.

We take three steps to transform equations (26) and (27) into estimation equations. First, we parameterize the option value to depend on the rollup

rate and a vector $\mathbf{y}_{i,n,t}$ of other contract characteristics, which are the number of investment options, the GLWB dummy, and share class fixed effects. The option value of contract i in period t is

$$V_{i,n,t} = \exp(\delta' \mathbf{y}_{i,n,t} + \exp(\Delta' \mathbf{y}_{i,n,t}) r_{i,n,t} + \nu_{i,n,t}). \quad (28)$$

The vector δ contains coefficients that determine the level of option value. The residual $\nu_{i,n,t}$ represents unobserved (to the econometrician) contract characteristics that affect the level of option value. The slope of the option value with respect to the rollup rate is

$$\frac{\partial V_{i,n,t}}{\partial r_{i,n,t}} = \exp(\Delta' \mathbf{y}_{i,n,t}) V_{i,n,t}. \quad (29)$$

The vector Δ contains the coefficients that determine the slope of the option value with respect to the rollup rate.

An alternative to our econometric approach in equation (28) is an option pricing model (Bauer, Kling, and Russ (2008); Milevsky and Salisbury (2006)). However, long-horizon volatility is a key input in option pricing, which we cannot estimate from traded options that have much shorter maturity than the minimum return guarantees. An insight that we offer for this literature is that the fee actually includes a markup and the shadow cost of capital, according to equation (13). Therefore, one should not estimate an option pricing model directly on the fee without first taking out the markup and the shadow cost of capital. An interesting area for future research is to estimate long-horizon implied volatility using our estimates of option value for a large cross section of contracts.

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 reinsurance share of variable annuities. Let $\mathbf{1}_n$ be a vector whose n th element is one and other elements are zero. The shadow cost of capital for insurer n in period t is

$$\lambda_{n,t} = \exp(\Gamma' \mathbf{z}_{n,t} + \gamma' \mathbf{1}_n). \quad (30)$$

The vector Γ contains the coefficients on insurer characteristics. The vector γ contains the coefficients on the insurer fixed effects, which capture permanent differences in the shadow cost of capital across insurers. For example, stock companies have a more complex financial structure than mutual companies that allows for captive reinsurance (Koijen and Yogo (2016)). We do not have time fixed effects, so the only time variation in the shadow cost of capital comes from insurer characteristics. This assumption is conservative in that we do not overattribute time variation in marginal cost to the shadow cost of capital rather than the option value.

Third, we derive 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 multiproduct insurer and describe how to compute semielasticities of demand from the estimated model of variable annuity demand.

Taking these three steps, equations (26) and (27) become

$$v_{i,n,t} = \log \left(P_{i,n,t} - \frac{1}{\epsilon_{i,n,t}^P} \right) - \delta' \mathbf{y}_{i,n,t} - \exp(\Delta' \mathbf{y}_{i,n,t}) r_{i,n,t} - \Gamma' \mathbf{z}_{n,t} - \gamma' \mathbf{1}_n, \quad (31)$$

$$\omega_{i,n,t} = \log \left(\frac{\epsilon_{i,n,t}^r}{\epsilon_{i,n,t}^P} \right) - \log \left(P_{i,n,t} - \frac{1}{\epsilon_{i,n,t}^P} \right) - \Delta' \mathbf{y}_{i,n,t}. \quad (32)$$

After subtracting marginal cost, equation (32) for the optimal rollup rate identifies only the slope of the 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.

B. Estimation Methodology

Let $\mathbf{0}$ be a vector of zeros, and $\mathbf{1}$ be a vector of ones. Let $\text{diag}(\cdot)$ be a diagonal matrix, so that $\text{diag}(\mathbf{1})$ is an identity matrix. The moment condition for the optimal fee is

$$\mathbb{E} \left[v_{i,n,t} \begin{pmatrix} \mathbf{y}_{i,n,t} \\ \mathbf{y}_{i,n,t} r_{i,n,t} \\ \mathbf{z}_{n,t} \\ \mathbf{1}_n \end{pmatrix} \right] = \mathbf{0}. \quad (33)$$

Since the vector of contract characteristics satisfies $\mathbf{y}_{i,n,t} \geq \mathbf{0}$ in our specification, the moment inequality for the optimal rollup rate is

$$\mathbb{E}[\omega_{i,n,t} \mathbf{y}_{i,n,t}] \geq \mathbf{0}. \quad (34)$$

We transform this equation into a moment equality as

$$\mathbb{E}[\text{diag}(\omega_{i,n,t} \mathbf{1} - \Omega) \mathbf{y}_{i,n,t}] = \mathbf{0}, \quad (35)$$

where $\Omega \geq \mathbf{0}$ is a vector of auxiliary parameters that captures the slackness of the inequality.

We plug in point estimates of demand elasticities in equations (31) and (32). We then estimate equations (33) and (35) by two-step generalized method of moments. Moon and Schorfheide (2009) discuss identification issues related to our model that has both moment equalities and inequalities. If $\Omega = \mathbf{0}$, the model is overidentified, and the moment condition for the optimal rollup rate is informative about the slope of the option value with respect to the rollup rate. If $\Omega > \mathbf{0}$, the model is exactly identified, and the moment condition for the optimal rollup rate is uninformative about the slope of the option value

with respect to the rollup rate. Between these extreme cases, only a subset of the moments in equation (35) may be informative about the slope of the option value with respect to the rollup rate.

The intercept in equation (31) 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 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 on the option value and a lower bound on the shadow cost of capital for each contract.

C. Estimated Model of Variable Annuity Supply

Table V reports the estimated model of variable annuity supply. 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 reinsurance share of variable annuities. These estimates also validate the “first stage” of the demand estimation in Table IV, which relies on log reserve valuation and the reinsurance share of variable annuities as relevant instruments for fees and rollup rates.

The average value of the shadow cost of capital is $\lambda_t = 1.52$ for the cross section of insurers in 2008:4. We check the internal consistency of this estimate to the estimated marginal cost of capital in Koijen and Yogo (2015) through a back-of-the-envelope calculation. In equation (14), the shadow cost of capital depends on the marginal cost of capital c_t and the risk weight ϕ_t , which is an unknown value that depends on the overall risk profile of minimum return guarantees. Based on a relevant case study that provides a sense of magnitudes, the required capital for a GLWB under the C-3 Phase II regulatory standard is 8.2% of account value, while the required capital under Actuarial Guideline 43 is 4.4% of account value (Junus and Motiwalla (2009, table 9)). Assuming that the reserve value under Actuarial Guideline 43 is conservative and greater than the market value, a lower bound on the risk weight is $\phi_t = 8.2/4.4 - 1 = 0.86$. This implies that the marginal cost of capital is

$$c_t = \frac{\lambda_t - 1}{\phi_t - \lambda_t + 1} = 1.53 \quad (36)$$

when $\lambda_t = 1.52$ and $\phi_t = 0.86$. According to Koijen and Yogo (2015, figure 7), the average value of the marginal cost of capital is $c_t = 0.96$ for the cross section of insurers in 2008:4. Moreover, 1.53 is well within the range of estimates for the cross section of variable annuity insurers (Koijen and Yogo (2015, table IV)).

Through the estimated model for the optimal fee, we decompose the total annual fee for contracts with minimum return guarantees, averaged across contracts with sales weighting. Figure 8 reproduces the total annual fee from

Table V
Estimated Model of Variable Annuity Supply

A model of variable annuity supply is estimated by two-step generalized method of moments. For reporting purposes, the rollup rate is orthogonalized with respect to the number of investment options, the GLWB dummy, and share class fixed effects. B is the omitted category for the share class fixed effects. In Panel C, the shadow cost of capital depends on the A.M. Best rating, log reserve valuation, the reinsurance share of variable annuities, and insurer fixed effects whose coefficients are not reported for brevity. In Panel D, the auxiliary parameters for the moments corresponding to the number of investment options, the GLWB dummy, and share class fixed effects for C and I are set to zero because their inequality constraints are binding. Heteroskedasticity-robust standard errors are reported in parentheses. The sample includes all contracts from 2005:1 to 2015:4.

Variable	Coefficient	Standard Error
Panel A: Level of Option Value		
Investment options	0.44	(0.02)
GLWB	46.05	(0.25)
Share class		
A	-44.01	(0.93)
C	11.53	(0.36)
I	-78.83	(1.39)
L	18.78	(0.26)
O	-19.08	(1.13)
X	16.06	(0.33)
Panel B: Slope of Option Value		
Investment options	2.90	(0.13)
GLWB	-21.67	(0.91)
Share class		
A	29.41	(5.15)
C	5.09	(1.31)
I	89.48	(1.85)
L	-28.94	(2.51)
O	13.92	(10.15)
X	-16.88	(3.45)
Panel C: Shadow Cost of Capital		
A.M. Best rating	-2.08	(0.25)
Reserve valuation	0.48	(0.25)
Variable annuities reinsured	0.97	(0.20)
Panel D: Constraint on the Rollup Rate		
Investment options	0.00	
GLWB	0.00	
Share class		
A	26.73	(4.63)
C	0.00	
I	0.00	
L	25.44	(1.89)
O	15.50	(9.17)
X	12.81	(2.73)
Constant	14.14	(0.60)
Observations	32,419	

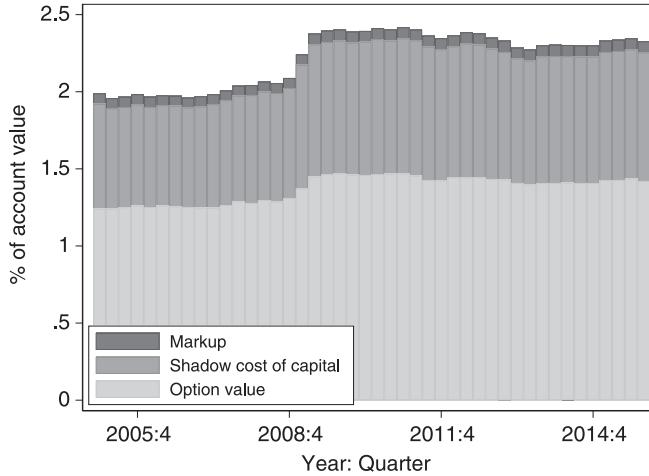


Figure 8. Decomposition of fees. The total annual fee, averaged across contracts with sales weighting, is decomposed into the markup, the option value, and the shadow cost of capital. The sample includes all contracts with minimum return guarantees from 2005:1 to 2015:4.

Panel A of Figure 5 and shows its decomposition into the markup, the option value, and the shadow cost of capital. The total annual fee was 2.04% of account value in 2007:4, which is the sum of 0.06% for the markup, 1.29% for the option value, and 0.69% for the shadow cost of capital. Thus, the profit was 0.75% of account value, which are mostly from financial frictions (0.69%) rather than market power (0.06%). 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 reflects an increase of 1 basis point in the markup, 16 basis points in the option value, and 17 basis points in the shadow cost of capital. Thus, financial frictions are just as important as the option pricing channel for explaining the increase in fees during the global financial crisis.

Rearranging equation (27) for the optimal rollup rate and taking the expectation across contracts,

$$\mathbb{E} \left[\log \left(\frac{\epsilon_{i,n,t}^r}{\epsilon_{i,n,t}^P} \right) - \log(\lambda_{n,t}) \right] = \mathbb{E} \left[\log \left(\frac{\partial V_{i,n,t}}{\partial r_{i,n,t}} \right) + \omega_{i,n,t} \right]. \quad (37)$$

We use this equation to decompose the time variation in the determination of optimal rollup rates. Panel A of Figure 9 shows the two variables on the left side of equation (37), which is the relative demand elasticities minus the shadow cost of capital. Both variables are reported as deviations from their time-series mean. As we discuss in Section IV, the demand elasticity to the fee decreased, while the demand elasticity to the rollup rate remained nearly constant. Thus, the relative demand elasticities increased after the global financial crisis, which increases the optimal rollup rate. As we discuss in Section III, the intuition is that insurers earn higher markups when the demand

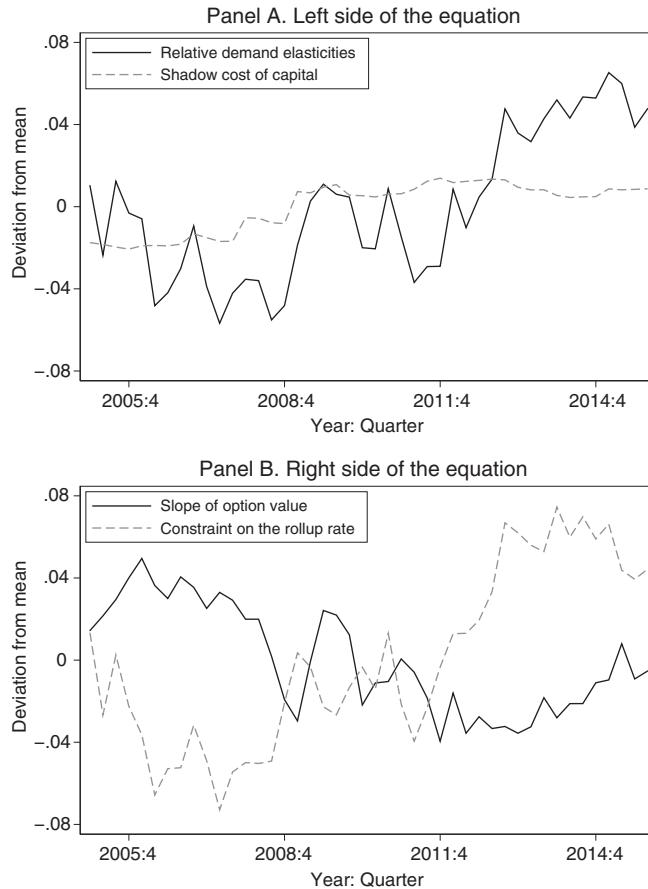


Figure 9. Decomposition of rollup rates. Panel A shows the two variables on the left side of equation (37), which is the relative demand elasticities minus the shadow cost of capital. Panel B shows the two variables on the right side of equation (37), which is the slope of the option value with respect to the rollup rate plus the constraint on the rollup rate. All variables are reported as deviations from their time-series mean. The sample includes all contracts from 2005:1 to 2015:4.

elasticity to the fee is lower, so variable annuities are less capital intensive per unit sold. The shadow cost of capital also increased during the global financial crisis, which reduces the optimal rollup rate. Between the two effects, the relative demand elasticities have larger time variation and play a more important role in the determination of rollup rates.

Panel B of Figure 9 shows the two variables on the right side of equation (37), which is the slope of the option value with respect to the rollup rate plus the constraint on the rollup rate. Both variables are reported as deviations from their time-series mean. The slope of the option value with respect to the rollup rate has a downward trend after the global financial crisis. The contracts with generous guarantees are too capital intensive after the global financial crisis

because the option value is more sensitive to the rollup rate. At an interior optimum, insurers optimally reduce rollup rates to the point where the slope of the option value is sufficiently low to satisfy equation (15). The optimal rollup rate could be so low that the constraint on the rollup rate is binding, which explains why insurers stopped offering minimum return guarantees. The relative demand elasticities in Panel A closely track the constraint on the rollup rate in Panel B. Thus, insurers stopped offering minimum return guarantees despite the fact that the relative demand elasticities increased.

As we discuss in Section II, two factors could explain why variable annuity supply did not fully recover long after the global financial crisis. First, insurers may have been more cautious because of the difficulty of managing interest rate risk in the low interest rate environment. Second, Actuarial Guideline 43 increased the capital requirements for variable annuities. These two factors ultimately drive up the shadow cost of capital, which is consistent with the evidence in Figures 8 and 9.

VI. 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 United States because of data availability, insurance products with minimum return guarantees are important globally and represent a major share of life insurer liabilities in Austria, Denmark, France, Germany, the Netherlands, and Sweden (European Systemic Risk Board (2015); Hombert and Lyonnet (2017); Kojen and Yogo (2022a)). 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 without 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 underfunding 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 underfunded 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.

Barbu, Alexandru, 2021, Ex-post loss sharing in consumer financial markets, Working paper, London Business School.

Bauer, Daniel, Alexander Kling, and Jochen Russ, 2008, A universal pricing framework for guaranteed minimum benefits in variable annuities, *Astin Bulletin* 38, 621–651.

Berends, Kyal, and Thomas B. King, 2015, Derivatives and collateral at U.S. life insurers, *Federal Reserve Bank of Chicago Economic Perspectives* 39, 21–37.

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.

Bhattacharya, Vivek, Gaston Illanes, and Manisha Padi, 2020, Fiduciary duty and the market for financial advice, Working paper, Northwestern University.

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).

Calvet, Laurent, Paolo Sodini, Claire Célérier, and Boris Vallée, 2019, Can financial innovation solve household reluctance to take risk? Working paper, EDHEC Business School.

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.

Egan, Mark, Shan Ge, and Johnny Tang, 2020, Conflicting interests and the effect of fiduciary duty: Evidence from variable annuities, Working paper, Harvard University.

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).

Koijen, Ralph S. J., and Motohiro Yogo, 2022a, Global life insurers during a low interest rate environment, *AEA Papers and Proceedings* 112, forthcoming.

Koijen, Ralph S. J., and Motohiro Yogo, 2022b, Understanding the ownership structure of corporate bonds, *American Economic Review: Insights* forthcoming.

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.

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

Milevsky, Moshe A., and Thomas S. Salisbury, 2006, Financial valuation of guaranteed minimum withdrawal benefits, *Insurance: Mathematics and Economics* 38, 21–38.

Moon, Hyungsik Roger, and Frank Schorfheide, 2009, Estimation with overidentifying inequality moment conditions, *Journal of Econometrics* 2, 136–154.

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 University.

Shi, Yu, 2021, The nonfinancial value of financial firms, Working paper, UCLA.

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.

Appendix 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: $\mathbb{E}_t[M_{t,t+s}] = (1 + y_t(s))^{-s}$. That is, $y_t(s)$ is the annually compounded zero-coupon yield at maturity of s years and 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

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

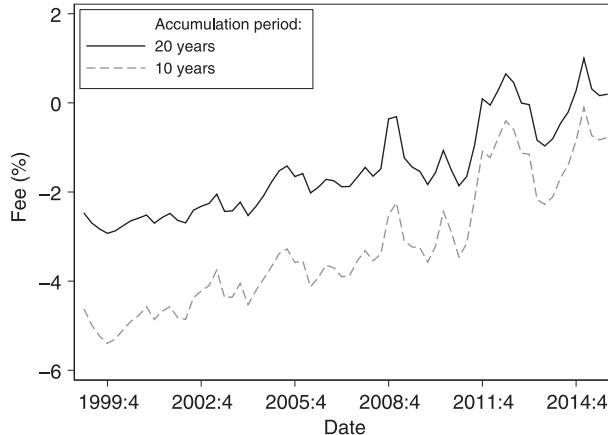


Figure A.1. A lower bound on fees based on the rollup rate. The lower bound is based on an annual rollup rate of 5%, an annual withdrawal rate of 5%, and a withdrawal period of 20 years. The calculation uses an average of the zero-coupon Treasury yield curve within each quarter from 1999:1 to 2015:4,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

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:

$$1 + \mathbb{E}_t \left[\sum_{s=1}^{T_a} M_{t,t+s} \frac{PS_{t+s}}{S_t} \right] = 1 + T_a P = \mathbb{E}_t \left[\sum_{s=1}^{T_w} M_{t,t+T_a+s} w X_{t,t+T_a} \right]. \quad (\text{A.2})$$

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

$$P \geq \frac{1}{T_a} \left(\sum_{s=1}^{T_w} \frac{w(1+r)^{T_a}}{(1+y_t(T_a+s))^{T_a+s}} - 1 \right). \quad (\text{A.3})$$

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, Sack, and Wright (2007)).

Figure A.1 shows 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 shows 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.

Appendix B: Portfolio of U.S. Life Insurers

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

Table B.1
Publicly Traded U.S. Life Insurers

This table lists 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
Brighthouse Financial	September 2017
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
Voya	June 2013

Appendix C: Proof of Proposition 1

This proof covers the case of a multiproduct 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. Let $\text{diag}(\cdot)$ be a diagonal matrix, so that $\text{diag}(\mathbf{1})$ is an identity matrix.

Generalizing equation (11), the insurer chooses a vector of fees \mathbf{P}_t and rollup rates \mathbf{r}_t to maximize firm value:

$$J_t = (\mathbf{P}_t - \mathbf{V}_{t,t})' \mathbf{Q}_t - C_t. \quad (\text{C.1})$$

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

$$K_t = R_{K,t} K_{t-1} + (\mathbf{P}_t - (1 + \phi_t) \mathbf{V}_{t,t})' \mathbf{Q}_t, \quad (\text{C.2})$$

where

$$R_{K,t} = \frac{A_{t-1}}{K_{t-1}} R_{A,t} - \frac{(1 + \phi_t) L_{t-1}}{K_{t-1}} \frac{\mathbf{V}'_{t-1,t} \mathbf{Q}_{t-1}}{\mathbf{V}'_{t-1,t-1} \mathbf{Q}_{t-1}} \quad (\text{C.3})$$

is the return on statutory capital.

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

$$\begin{aligned} \frac{\partial J_t}{\partial \mathbf{P}_t} &= \frac{\partial (\mathbf{P}_t - \mathbf{V}_{t,t})' \mathbf{Q}_t}{\partial \mathbf{P}_t} + c_t \frac{\partial K_t}{\partial \mathbf{P}_t} \\ &= \mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} (\mathbf{P}_t - \mathbf{V}_{t,t}) + c_t \left(\mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} (\mathbf{P}_t - (1 + \phi_t) \mathbf{V}_{t,t}) \right) \\ &= (1 + c_t) \mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} ((1 + c_t) (\mathbf{P}_t - \mathbf{V}_{t,t}) - c_t \phi_t \mathbf{V}_{t,t}). \end{aligned} \quad (\text{C.4})$$

The optimal fee satisfies

$$\frac{\partial J_t}{\partial \mathbf{P}_t} = \mathbf{0} \Leftrightarrow \mathbf{P}_t + \left(\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} \right)^{-1} \mathbf{Q}_t = \frac{1 + c_t (1 + \phi_t)}{1 + c_t} \mathbf{V}_{t,t}. \quad (\text{C.5})$$

Equation (13) follows from the definition of semielasticity of demand to the fee.

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

$$\begin{aligned} \frac{\partial J_t}{\partial \mathbf{r}_t} &= \frac{\partial (\mathbf{P}_t - \mathbf{V}_{t,t})' \mathbf{Q}_t}{\partial \mathbf{r}_t} + c_t \frac{\partial K_t}{\partial \mathbf{r}_t} \\ &= - \frac{\partial \mathbf{V}'_{t,t}}{\partial \mathbf{r}_t} \mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} (\mathbf{P}_t - \mathbf{V}_{t,t}) \\ &\quad + c_t \left(-(1 + \phi_t) \frac{\partial \mathbf{V}'_{t,t}}{\partial \mathbf{r}_t} \mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} (\mathbf{P}_t - (1 + \phi_t) \mathbf{V}_{t,t}) \right) \\ &= -(1 + c_t (1 + \phi_t)) \frac{\partial \mathbf{V}'_{t,t}}{\partial \mathbf{r}_t} \mathbf{Q}_t + \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} ((1 + c_t) (\mathbf{P}_t - \mathbf{V}_{t,t}) - c_t \phi_t \mathbf{V}_{t,t}) \\ &= -(1 + c_t (1 + \phi_t)) \frac{\partial \mathbf{V}'_{t,t}}{\partial \mathbf{r}_t} \mathbf{Q}_t - (1 + c_t) \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} \left(\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} \right)^{-1} \mathbf{Q}_t, \end{aligned} \quad (\text{C.6})$$

where the last line follows from substituting equation (C.4). At an interior optimum, the rollup rate satisfies

$$\frac{\partial J_t}{\partial \mathbf{r}_t} = \mathbf{0} \Leftrightarrow - \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} \left(\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} \right)^{-1} \mathbf{Q}_t = \frac{1 + c_t (1 + \phi_t)}{1 + c_t} \frac{\partial \mathbf{V}'_{t,t}}{\partial \mathbf{r}_t} \mathbf{Q}_t. \quad (\text{C.7})$$

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

$$- \text{diag}(\mathbf{Q}_t)^{-1} \frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} \left(\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} \right)^{-1} \mathbf{Q}_t = \frac{1 + c_t (1 + \phi_t)}{1 + c_t} \frac{\partial \mathbf{V}'_{t,t}}{\partial \mathbf{r}_t} \mathbf{1}. \quad (\text{C.8})$$

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

The left sides of equations (C.5) and (C.8) correspond to the first terms inside the logarithm in equations (26) and (27), respectively. For the random coefficients logit model, we denote the vector of demand for all contracts that an insurer sells as

$$\mathbf{Q}_t = \int \mathbf{q}_t(\alpha_P) dF(\alpha_P). \quad (\text{C.9})$$

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

$$\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{P}_t} = \int -\alpha_P (\text{diag}(\mathbf{q}_t(\alpha_P)) - \mathbf{q}_t(\alpha_P) \mathbf{q}_t(\alpha_P)') dF(\alpha_P). \quad (\text{C.10})$$

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

$$\frac{\partial \mathbf{Q}'_t}{\partial \mathbf{r}_t} = \int \alpha_r (\text{diag}(\mathbf{q}_t(\alpha_P)) - \mathbf{q}_t(\alpha_P) \mathbf{q}_t(\alpha_P)') dF(\alpha_P). \quad (\text{C.11})$$

Thus, the estimated model of variable annuity demand in Table IV directly implies the left sides of equations (C.5) and (C.8).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Replication Code.