Selection through Lapsation in Life Insurance Markets*

Cameron M. Ellis† & Johannes G. Jaspersen‡

January 7, 2026

Abstract

Life insurance is often believed to feature advantageous, rather than adverse, selection since policyholders have lower mortality than non-policyholders. We introduce policy lapses as an alternative selection channel. We demonstrate theoretically that lapsing can induce adverse selection, even among naïve consumers. We then empirically validate our theory using administrative data and confirm that lapse rates respond significantly to price changes, leading to adverse selection and a downward-sloping cost curve. No comparable effect emerges in mortality rates. This adverse selection reduces overall demand by 14.2% and reduces welfare by $1.3 billion annually. Furthermore, lapse-based selection exacerbates inequity, disproportionately benefiting wealthy consumers.

JEL Codes: **D91** · **G22** · **G52**

Keywords: Adverse Selection · Insurance Demand · Household Finance · Insurance Markets

*Acknowledgments: We thank Daniel Bauer, Benjamin Collier, Keith Ericson, Shan Ge, Daniel Gottlieb, David McCarthy, Casey Rothschild, Adam Solomon, Elisabeth Stöckl, and Justin Sydnor as well as seminar participants at the University of Iowa, The University of Georgia, LMU Munich, Goethe University Frankfurt, The National Association of Insurance Commissioners (NAIC) annual meeting, and the MRIC/CEAR Behavioral Insurance Workshop for helpful comments. Barbara Reißer provided excellent research assistance. Ellis received financial support from the NAIC through the research fellows program. We thank the NAIC for their help in understanding the data.

†**Corresponding Author** Tippie College of Business, University of Iowa, S284, Pappajohn Business Building, 21 E Market St, Iowa City, IA, 52246 (cameron-ellis@uiowa.edu)

‡LMU Munich School of Management, LMU Munich (jaspersen@lmu.de)

1 Introduction

Asymmetric information is perhaps the most studied topic in insurance markets and life insurance is the largest private insurance market in the world, with $3.3 trillion in premiums paid in 2024 (Holzhausen et al., 2025). Despite this, we don't actually know much about selection in life insurance markets. A few studies find adverse selection, where individuals with higher mortality risk are more likely to purchase life insurance (e.g., Giesbert et al., 2011; He, 2009), yet other, equally high quality, studies show advantageous selection, where life insurance holders tend to have lower mortality risk (e.g., Cawley and Philipson, 1999; McCarthy and Mitchell, 2010). Many studies fail to find either (e.g., Harris and Yelowitz, 2014; Wuppermann, 2017) and some even find both (e.g., Ai and Steinorth, 2024) along different points of the health distribution.

Seemingly the only common feature in the existing literature on selection in life insurance is the use of mortality data as a proxy for costs. Yet, Gottlieb and Smetters (2021) show that mortality is not the only, or even primary, driver of life insurance costs. Instead, it is lapse rates that truly drive costs and lapse risk is not individually underwritten (Koijen et al., 2024). However, lapsing has not been examined as an alternative channel for adverse selection.

This paper fills that gap. We adopt the framework of Einav et al. (2010) and show that mortality and lapsation likely have opposing effects on selection in life insurance, regardless of the level of knowledge households have about their own risks. We then use data from US life insurance companies' statutory filings to estimate the degree of selection for the individual life insurance market. We find a strong and robust degree of selection with a 1% increase in premiums leading to a 0.13% increase in average future costs. We show that this selection is driven entirely by a reduction in future lapse rates with a tightly estimated null effect on future mortality. Using our causal estimates, we next calculate that this lapse-induced adverse selection decreases overall life insurance demand by 14.2%, leading to an annual welfare loss of $1.3 billion. Finally, we estimate a structural demand model to examine the equity consequences of "fixing" this adverse selection. We find that solutions are progressive, with welfare gains concentrated in low-wealth households.

Life insurance premiums have two primary drivers: mortality rates and lapse rates. Higher

mortality risk increases an individual's expected value of the insurance contract, because payouts are more likely, while higher risk of lapsing decreases the contract's expected value (Gottlieb and Smetters, 2021). Since the policyholder's mortality risk increases with age and premiums are typically flat, premiums tend to be higher than the actuarial value early in the policy and lower than the actuarial value later in the policy. If policyholders lapse before reaching the later stages of a policy they will have "overpaid" the entire time. Lapse rates are enormous. Between 1991 and 2010, $24 trillion of $29.7 trillion in new individual life insurance coverage issued in the United States lapsed (Gottlieb and Smetters, 2021).[1]

Studies showing advantageous selection in life insurance typically explain the counter-intuitive result through a correlation of both demand and mortality risk with a third factor.[2] Wealth, income, and/or financial sophistication are the typical candidates since they are highly correlated with both mortality and life insurance demand (e.g. Chetty et al., 2016; Gropper and Kuhnen, 2025). These third factors are also highly correlated with the the decisions to lapse (Fang and Kung, 2021).[3]

To explore these complex dynamics, we build upon and extend the framework of Einav et al. (2010) to life insurance and incorporate mortality risk, lapse risk, and a third factor, which we call consumer type. Consumer type proxies for wealth and financial sophistication and influences income, mortality risk, and lapse probability. Our theoretical model yields two key insights. First, we demonstrate that both adverse and advantageous selection can emerge in the life insurance market, depending on the relative strength of opposing effects. "High-type" individuals (with higher wealth) tend to have lower mortality risk but also lower lapse probabilities. These factors have countervailing effects on insurers' costs and consumers' willingness to pay for insurance. Consequently, the direction of selection in the market is theoretically ambiguous and depends on

[1] Analysis of our own data indicates that a policy is more than twice as likely to lapse than to cease for all other reasons combined. Of all policies ceasing in a year, 68% are due to lapsation, 15% to surrender, 12% to death, and 5% to expiration.

[2] Similar third factor correlations exist in other markets, for example Fang et al. (2008) and Cutler et al. (2008).

[3] The literature identifies several reasons for lapsing, including preference shocks (Fang and Kung, 2010; Liebenberg et al., 2012; Fei et al., 2015), income shocks (Linton, 1932; Outreville, 1990; Kuo et al., 2003), policy replacement (Outreville, 1990; Carson and Forster, 2000), changes in interest rates (Schott, 1971; Pesando, 1974; Kuo et al., 2003), and non-expected utility explanations (Shefrin, 2002; Mulholland and Finke, 2014; Gottlieb and Smetters, 2021). For a comprehensive review, see Eling and Kochanski (2013).

which effect dominates. Second, we show that both forms of selection can occur even if consumers are naïve and do not know their own risk. The type of selection (adverse or advantageous) that emerges will depend on if influence of consumer type on costs is stronger through the mortality channel or the lapse channel.

We next empirically test our theoretical predictions and quantify the extent of selection in the life insurance market. We leverage comprehensive, administrative data compiled from the National Association of Insurance Commissioners (NAIC) regulatory reports. Our data feature firm-level data for every operating life insurance company in the U.S. from 1993 to 2017 and includes both national-level and state-specific data. To address potential endogeneity concerns in our empirical analysis, we follow an extensive recent tradition of instrumenting insurance prices with prices/costs by the same firm in other markets (e.g., Koijen and Yogo, 2016, 2015; Ge, 2022; Gao et al., 2023; Wenning, 2024). Specifically, we use a firm's group-market life insurance premiums in a year as an instrument for their individual-market life insurance premiums in the same year. We combine this instrument with extensive fixed effects to ensure that our identifying variation only comes from firm-year level cost shocks in the group market spilling over into individual market pricing.

We estimate the causal impact of a change in premiums on (1) average future costs, (2) contemporaneous demand, (3) future lapse rates, (4) future surrender rates, and (5) future death rates. We find strong and robust evidence of adverse selection in the life insurance market, with a 1% increase in premiums associated with a 0.13% increase in average future costs for insurers. Our results also highlight the dominance of the lapse channel in driving this selection, with a 1% increase in premiums reducing lapse rates by 0.15%. We also are able to tightly estimate a null effect of premiums on death rates, indicating that mortality risk is not the primary channel through which selection is occurring in this market.[4] Our findings underscore the importance of considering lapse behavior, rather than focusing solely on mortality, when analyzing selection in life insurance markets and provide empirical validation for our theoretical model of the mechanisms driving selection in this massive financial market.

[4]We additionally examine the impact of premiums on surrender rates. Our primary specification finds a negative impact (around a quarter of the lapse channel), but we do not find this result to be robust.

We next use our empirical results to estimate the efficiency losses due to lapse-based adverse selection. We find that, on average, annual new life insurance business is 14.2% smaller in equilibrium than in the efficient market outcome. This implies an annual net present value for the dead-weight loss of $1.29 billion for the companies in our estimation sample. We examine the robustness of this welfare result in four ways. First, we use a block-bootstrap procedure to show robustness to sampling error. Second, we consider the alternative explanation of quantity rationing and show that controlling for it does not change our results. Third, we use the specification curves of Simonsohn et al. (2020) to show robustness to research design choices. Finally, we develop a novel method, in the spirit of Kang and Vasserman (2025), to show robustness to functional form assumptions.

Given our estimated welfare losses and the documented correlation between life insurance demand, lapse behavior, and wealth, our results raise concerns about economic inequality. To analyze the equity consequences of the documented selection, we also estimate a Berry et al. (1995, BLP) style demand model in which we combine with micro-moments on purchase and lapse behavior by wealth from the Health and Retirement Survey. After estimating the model, we examine the distributional impacts of eliminating the selection through individually-underwritten lapse risk. We find that, in addition to the overall reduction in deadweight-loss from the removal of selection, there would also be a transfer from the top half of the wealth distribution to the bottom half, with the poorest receiving the largest gain.[5]

We make three primary contributions. Our first contribution is to provide a comprehensive theoretical and empirical analysis of selection in the US life insurance market, considering both mortality risk and lapse behavior. Our theoretical model demonstrates that adverse and advantageous selection can coexist within the same market, depending on the relative strength of opposing effects related to consumer type, mortality risk, and lapse probabilities. We also show that knowledge of lapse rates is not necessary for this selection channel to occur (Carson et al., 2020; Gottlieb and Smetters, 2021). This helps explain divergent results in previous studies (e.g., Cawley and Philipson, 1999; McCarthy and Mitchell, 2010; Giesbert et al., 2011; He, 2009) and calls for a

[5]This cross-subsidy was identified, but not calculated, by Koijen et al. (2024).

more nuanced approach to modeling selection in long-term insurance markets. We use these theoretical insights to take the model to the data, providing an estimate of the welfare consequences of adverse selection in the entire US life insurance market. This contributes to the growing literature applying the Einav et al. (2010) equilibrium to various markets such as health insurance, unemployment insurance, consumer credit and others (Hackmann et al., 2015; Landais et al., 2021; DeFusco et al., 2022).

Our second contribution is to empirically demonstrate that lapsing, and not mortality, leads to adverse selection in life insurance markets. Gottlieb and Smetters (2021) shows that lapse rates have a first order effect on life insurance costs, yet these costs are only priced in the aggregate (Koijen et al., 2024), which leads to large selection issues. A growing body of literature also examines what is termed "dynamic" adverse selection, where the decision to lapse a policy is correlated with the insured's mortality risk (e.g., He, 2011; Lai et al., 2021; Ai and Steinorth, 2024). We add to this literature by showing that adverse selection in life insurance is also driven by the correlation of demand and ex ante lapse probabilities, rather that ex post lapse behavior and mortality. Our results highlight the how lapse behavior shapes the risk pool and challenges prior models that focus predominantly on mortality risk as the key driver of adverse selection.

Finally, we contribute through our novel extension of Kang and Vasserman's (2025) method of examining functional form robustness in analysis of welfare losses to Akerlovian selection markets (such as Einav et al., 2010). While Kang and Vasserman's (2025) method is more general, it cannot be applied to Einav et al. (2010) equilibria because the welfare losses in them are determined by two separate but interdependent functions (demand and average cost), whereas our parametric application can. The Kang and Vasserman (2025) procedure applies to situations where only the demand function determines welfare consequences, while our procedure is a simplified and complimentary tool for applied researchers examining any Akerlovian section market.

Our findings have significant policy implications, particularly regarding the pricing of life insurance policies. There are three potential solutions to lapse-driven selection: incorporating individual-based lapse estimates into pricing models, individual policy reserves, or removing flat premiums and instead have premiums reflect contemporaneous mortality risk. All of these

changes would benefit low-income consumers, making life insurance more accessible while reducing subsidies for high-income groups who maintain their policies long-term. Unfortunately, lapsed-based pricing would reshape the secondary life insurance market by creating new arbitrage opportunities for life settlements, especially within stranger-originated life insurance (STOLI) schemes. Contemporaneous-risk premiums would reduce such dynamics, but result in rising premiums over time which are unpopular with consumers. Individual policy reserves, which would pay out the front-loading portion of premium payments in case of lapsing, maintain the flat premium schedule and affect secondary life insurance markets only by making lapsing a viable alternative to selling one's life insurance.

2 An Illustrative Model of Selection in Life Insurance

In this section, we create an illustrative model of selection in life insurance markets with both mortality and lapsing. We show how introducing an unpriced consumer "type" (think wealth or financial literacy) that is correlated with demand, mortality, and lapse probability can induce either adverse and advantageous selection. We first establish a general result that the type of selection present in the Einav et al. (2010, henceforth EFC) equilibrium is determined solely by the overall correlation of consumer type with costs and with demand. This result implies that when the correlation of consumer type and lapse rate is strong, and lapse rates have a large impact on costs, it is possible to observe a negative correlation of life insurance take-up and mortality (which much of the previous literature has interpreted as advantageous selection) even if the market actually features adverse selection.

We then provide microfoundations for life insurance demand and producer costs incorporating both lapses and mortality. In the spirit of Gottlieb and Smetters (2021), we provide two different foundations for policy lapses: consumers forgetting to pay their premium and persistent income shocks. For both models, we consider both sophisticated consumers and those who are naïve about their own lapse probability or mortality.

2.1 Determining the Market Equilibrium

The EFC equilibrium of an insurance market is defined by the intersection of the demand curve and the average cost curve in the quantity/ price plane. For this illustration, we assume that consumer types can be summarized by a single (scalar-valued) type, which we denote ψ. This simplification allows for analytically tractable results and is sufficient for our purpose. Other assumptions about consumer heterogeneity are possible but would likely require parametric assumptions to render solutions. Let types be distributed according to $g(\psi)$ and define the quantity demanded at price Π as

$$Q(\Pi) = \int \mathbf{1}(\bar{\Pi}(\psi) \geq \Pi) g(\psi) d\psi \tag{1}$$

where $\bar{\Pi}(\psi)$ is the maximum willingness to pay for consumers of type ψ.[6] Marginal costs at price Π are then given as

$$MC(\Pi) = E[C(\psi)|\bar{\Pi}(\psi) = \Pi] \tag{2}$$

where $C(\psi)$ are the costs of consumers of type ψ. In order to guarantee an analytical solution, we make the simplifying assumption that both C and $\bar{\Pi}$ are monotonic in ψ. While this assumption may not be realistic for the entire support of $g(\psi)$, we would argue that it can feasibly appear for the relevant area of the support where the equilibrium solution lies. We can calculate average costs as

$$AC(\Pi) = \frac{1}{Q(\Pi)} \int C(\psi) \mathbf{1}(\bar{\Pi}(\psi) \geq \Pi) g(\psi) d\psi. \tag{3}$$

From Equation (3) it is apparent that slopes of the average cost and marginal cost curves have the same sign in the quantity/ price plane. To determine the type of the market equilibrium, it is relevant whether the cost curves are increasing or decreasing in price.

[6] Note that the demand curve is always falling in price because there is only a single insurance contract and consumers are assumed to purchase one of these contracts at most.

Proposition 1. *If $C(\psi)$ and $\bar{\Pi}(\psi)$ are monotonic in ψ, the resulting equilibrium is one of adverse selection if $sgn\left(\frac{\partial C}{\partial \psi}\right) = sgn\left(\frac{\partial \bar{\Pi}}{\partial \psi}\right)$ and one of advantageous selection if $sgn\left(\frac{\partial C}{\partial \psi}\right) \neq sgn\left(\frac{\partial \bar{\Pi}}{\partial \psi}\right)$, where sgn denotes the strict sign of an expression (excluding 0).*

All proofs are provided in Appendix A. Below, we will use this result in combination with microfoundations to illustrate the possibility of both adverse and advantageous selection in life insurance markets and the dependence of this result on the channels which determine demand and costs.

2.2 Lapsing by Forgetting to Make a Payment

Gottlieb and Smetters (2021) show that a major determinant for policy lapses is consumers forgetting to make a payment. In this first model, we the take this perspective and assume that the probability to lapse a policy is exogenous to the rest of the model and only influenced by the type of the consumer.[7] We assume that consumers receive income Υ_i and have the option to purchase a life insurance contract similar to term life contracts commonly observed in practice: a flat premium P for all periods, a payout B in case of death and no premiums or payout in case the policy lapses. As is common in this type of analysis (see e.g. Einav et al., 2010), we treat the payout as fixed to avoid the appearance of quantity rationing. Note, however, that we consider the robustness of our empirical findings to quantity rationing in Section 4.4.1 and find that there is no discernible effect on our finding of adverse selection through lapsation.

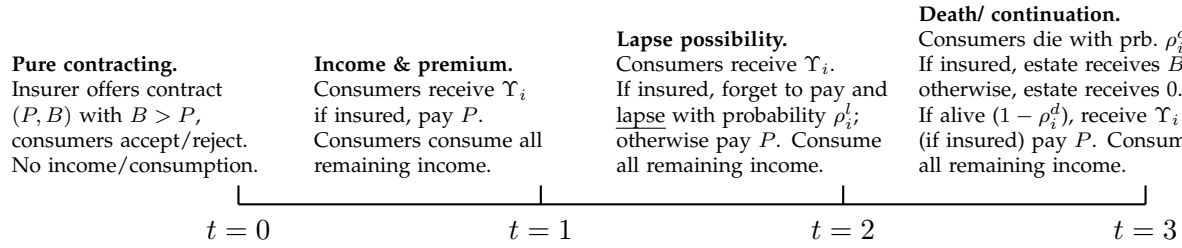
Timing: The time line of the model is given in Figure 1. To simplify the exposition, we follow the literature on life insurance lapses and assume that all assets outside of the insurance contract are perfectly illiquid and there is no savings.[8] Lapse and death probabilities (ρ_i^l and ρ_i^d) are exogenous to the model and only influenced by consumer type.

Consumer utility and costs: As long as they are alive, consumers derive utility $u_t(x)$ from con-

[7] Since we are focused on the demand side of the market, we do not model the supply of life insurance products.

[8] Gottlieb and Smetters (2021), Fang and Kung (2010) and Daily et al. (2008) all do not consider savings in their models featuring life insurance lapsation.

Figure 1: Time Line of Model when Lapses Originate from Forgotten Payments



Note: We assume $B > P$ such that the insurance contract is not first-order stochastically dominated. All assets outside the insurance contract are perfectly illiquid.

sumption in three periods $t \in \{1, 2, 3\}$.[9] In all periods, u is twice continuously differentiable with $u_t' > 0$ and $u_t'' < 0$. Dead consumers obtain utility from bequest motives. Utility from money in death is denoted $v_3(\kappa + x)$ with properties $v_3' > 0$ and $v_3'' < 0$ similar to utility from consumption when alive, but with an additional luxury good factor κ.[10]

In the baseline model, we consider sophisticated consumers who are informed about their probability of dying and their probability of lapsing, and cover the case of naïve consumers below. The expected utility of consumer i with parameters $\{\Upsilon_i, \rho_i^l, \rho_i^d\}$ when buying a life insurance contract (indicated LI) for premium P is given by

$$EU^{LI}(\Upsilon_i, \rho_i^l, \rho_i^d, P) = u_1(\Upsilon_i - P) + (1 - \rho_i^l) \underbrace{\left[u_2(\Upsilon_i - P) + (1 - \rho_i^d)u_3(\Upsilon_i - P) + \rho_i^d v_3(\kappa + B - P)\right]}_{\text{Expected utility without lapsing}}$$
$$+ \rho_i^l \underbrace{\left[u_2(\Upsilon_i) + (1 - \rho_i^d)u_3(\Upsilon_i) + \rho_i^d v_3(\kappa)\right]}_{\text{Expected utility when lapsed}}. \tag{4}$$

We are interested in how the maximum willingness to pay for life insurance changes in consumer type and thus the underlying characteristics Υ_i, ρ_i^l and ρ_i^d. We first need to define the maximum premium consumer i is willing to pay, $\bar{P}_i$. Assuming that the outside option is not purchas-

[9] Note that by time-indexing the utility function, we subsume all additive preference models, including the commonly used additive exponential discounting model.

[10] This factor and concavity of v_3 ensure that if an individual could freely allocate x between consumption when alive and bequest when dead, there is a critical level of x below which the individual would choose to consume 100% of the amount (see Lockwood, 2018, for a similar model set-up). The model is thus in line with the prevailing empirical evidence that bequests only appear in households above a certain level of wealth (Ameriks et al., 2011).

ing insurance, we can define this maximum acceptable premium through $EU^{LI}(\Upsilon_i, \rho_i^l, \rho_i^d, \bar{P}_i) - EU^0(\Upsilon_i, \rho_i^d) = 0$; where

$$EU^0(\Upsilon_i, \rho_i^d) = u_1(\Upsilon_i) + u_2(\Upsilon_i) + (1 - \rho_i^d)u_3(\Upsilon_i) + \rho_i^d v_3(\kappa). \tag{5}$$

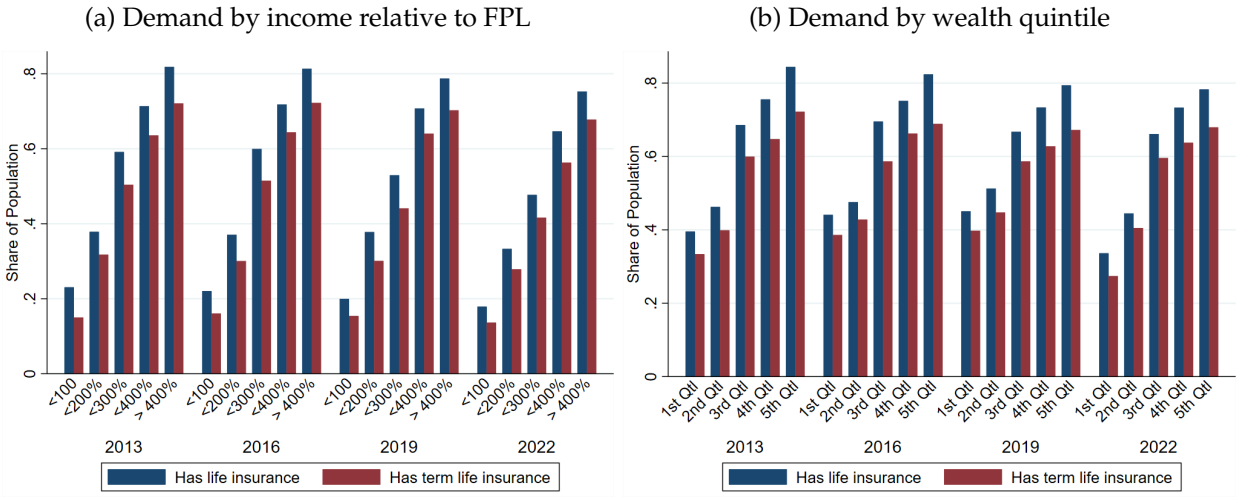
Costs of individual i are determined by the consumer's individual probability of death and lapse probability. Since the insurance company pays the fixed indemnity B if the consumer dies and has not lapsed the policy, costs are $C_i = \rho_i^d(1 - \rho_i^l)B$.

Consumer heterogeneity: The above definitions of $\bar{P}_i$ and C_i allow us to show that $\frac{\partial \bar{P}_i}{\partial \Upsilon_i} > 0$, $\frac{\partial \bar{P}_i}{\partial \rho_i^d} > 0$, $\frac{\partial \bar{P}_i}{\partial \rho_i^l} < 0$, $\frac{\partial C_i}{\partial \rho_i^d} > 0$, and $\frac{\partial C_i}{\partial \rho_i^l} < 0$. The positive influence of income on $\bar{P}$ stems from the concavity of the utility function. As the income of the consumer increases, the marginal cost of the premium decreases, while the benefit from bequests stays constant. This effect is in line with empirical evidence, as can be seen in Panel (a) of Figure 2 that shows the effect of income on (term-) life insurance demand using data from the Survey of Consumer Finances. We can clearly see a strong relationship between income and insurance demand. This mirrors results from the literature that wealthier and higher income individuals purchase more life insurance (Truett and Truett, 1990; Bernheim, 1991; Chen et al., 2001; Gropper and Kuhnen, 2025).

The influences of ρ^d and ρ^l on both $\bar{P}$ and C are intuitive. The maximum willingness to pay and costs both increase in ρ^d because the increased likelihood to die makes the bequest case, and thus the payout from the policy, more likely and thus consumers are willing to pay a higher premium for it. If consumers have a higher chance of lapsing, there is less chance of a policy payout, decreasing costs. If this is not reflected in individual premiums, then consumers are more likely to pay a premium without having a potential payout from the policy and $\bar{P}$ declines.

To illustrate the potential magnitude of these effects in a real life insurance contract, Figure 3 shows the annual premiums for a 20-year term life insurance policy and for repeated 1-year policies. In the beginning of the 20-year contract, policyholders pay over three times the risk-aligned annual premium. This equalizes with later premiums, beginning in contract year 12. However, any policyholder who cancels early will have paid (substantially) more than the aggregate 1-year

Figure 2: Life Insurance Demand by Income and Wealth

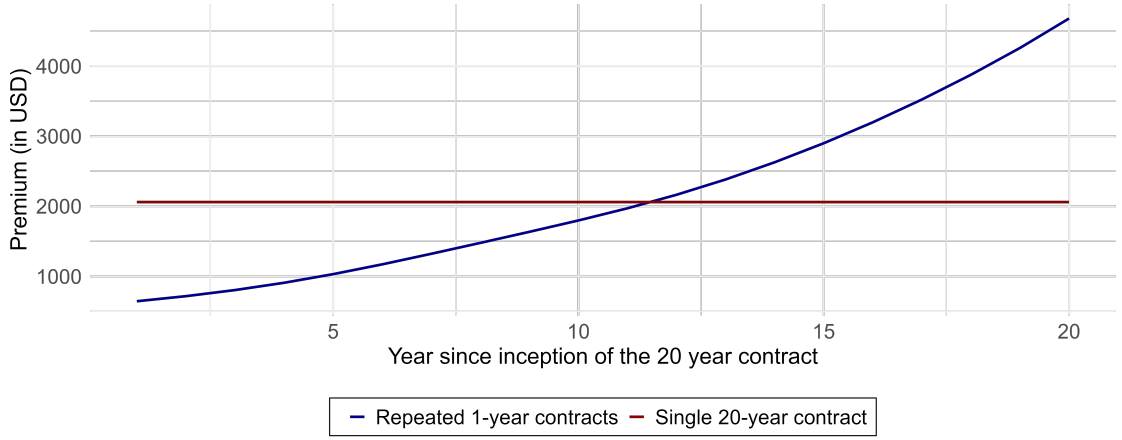


Note: Panel (a) shows the amount of households with 18-64 year-old household heads owning any type of life insurance coverage (blue) or term life (red) in the US. Demand is differentiated by year and income relative to the federal poverty line of the corresponding year. Panel (b) shows the same data, but differentiated by wealth quintile with wealth defined as the sum of all assets minus the sum of all debt. Data is taken from four waves of the Survey of Consumer Finances, see Appendix C for details on the variable construction.

premiums. While the sales process itself does lead to fixed costs, which could make very early policy lapses increase costs rather than decrease them, Koijen et al. (2024) show that in practice the "mortality" effect outweighs this "sales" effect. Further, the life insurance policies are often backdated far enough at the time of the sale to recoup all sales cost in the initial premium (Carson et al., 2020).

Maximum willingness to pay and type: Models of multiple periods are necessary when considering life insurance demand, especially when analyzing lapse behavior (e.g., Gottlieb and Smetters, 2021). However, using a three-period model requires clarification of the price that is used to determine insurance demand in the EFC equilibrium. It is not equal to the premium P in Equation (4), because the consumers pay this premium multiple times. To clear the disambiguation, we will determine the market equilibrium using Π to denote the price. Consumers always pay the premium in the first period and pay it in the second and third periods, if they do not lapse such that $\Pi_i = P(3 - 2\rho_i^l)$. The corresponding maximum price consumer i is willing to pay is given by $\bar{\Pi}_i = \bar{P}_i(3 - \rho_i^l)$.

Figure 3: Annual payments in 20-year and annual term life contracts



Note: Figure displays the premiums for a 20 year term life insurance contract of a 40-year old German male with a death benefit of $500,000 and the premiums of repeated annual contracts renewed each year. Mortality data is taken from the 2008 DAV German Endowment mortality table.

We now collapse the three consumer characteristics into a single, scalar-valued type ψ_i to facilitate an analytically tractable solution. Define ψ as a factor which has a positive influence on income ($\frac{\partial \Upsilon_i}{\partial \psi_i} \geq 0$) and, at the same time, reflects the other empirical regularities which are relevant for our analysis. ψ_i can be seen as a combination of wealth, financial literacy, and sophistication. This interpretation allows us to find clear indications from the empirical literature, on how ψ_i influences ρ_i^d and ρ_i^l. The "health-wealth gradient" is a long-standing result in public health that good financial standing and longevity are correlated positively, such that we can assume $\frac{\partial \rho_i^d}{\partial \psi_i} \leq 0$ (see, e.g., Chetty et al., 2016; Kinge et al., 2019). Analyses of life insurance lapse behavior also show financial literacy and wealth as important determinants (Nolte and Schneider, 2017; Fang and Kung, 2021). We can thus also assume that $\frac{\partial \rho_i^l}{\partial \psi_i} \leq 0$.

Market outcome: Given these assumptions, we can express both maximum willingness to pay and costs as a function of type, which lets us state the following result.

Proposition 2. *When policy lapses result from consumers forgetting to pay their premiums and consumers are sophisticated about their type, then*

1. *both adverse and advantageous selection can appear in the market,*

2. *when* $\frac{\partial \rho_i^l}{\partial \psi_i} = 0$ *and* $\frac{\partial \rho_i^d}{\partial \psi_i} < 0$*, adverse selection results if* $\frac{\partial \bar{P}_i}{\partial \Upsilon_i} \frac{\partial \Upsilon_i}{\partial \psi_i} < \frac{\partial \bar{P}_i}{\partial \rho_i^d} \frac{\partial \rho_i^d}{\partial \psi_i}$ *and advantageous selection results when the opposite holds, and*

3. *when* $\frac{\partial \rho_i^l}{\partial \psi_i} < 0$ *and* $\frac{\partial \rho_i^d}{\partial \psi_i} = 0$*, adverse selection results.*

Importantly for our empirical analysis, both adverse and advantageous selection are possible, but the channel through which consumer type affects consumer characteristics is important. When policy lapses are important for the market outcome, adverse selection is more likely. Advantageous selection can only appear if mortality affects the equilibrium.

Consumer naïveté: It is unclear whether assuming sophisticated consumers on life insurance markets is valid (Cawley and Philipson, 1999). Gottlieb and Smetters (2021) make the theoretical case that insurance contracts would not look like they do if consumers were considering their lapse probability when making the purchasing decision. Empirically, they conclude from a survey of life insurance policyholders that consumers are unlikely to have a realistic assessment of their lapse probability, often assuming that it is 0. A similar argument can be made about mortality risk. Subjective mortality estimates of survey respondents have substantial differences to their objective counterparts (Bissonnette et al., 2017).

It is thus possible that consumers do vary in their insured costs (through both ρ^d and ρ^l), but are not aware of this variation and do not take it into account in their insurance choices. Naïvety in consumers makes adverse selection less likely. After all, adverse selection traditionally arises from consumers knowing their underlying risk and act accordingly. We model consumer naïveté by assuming that they determine their maximum willingness to pay without regard to their idiosyncratic values of ρ_i^d and ρ_i^l and instead simply assume themselves to be equal to the average in probability of death and that their lapse probability is zero.[11] Formally, if average values are denoted by index b (for baseline), consumers with type ψ_i have a naively determined maximum acceptable premium, $\bar{P}_i^n$, defined implicitly by

$$EU^{LI}(\Upsilon(\psi_i), 0, \rho_b^d, \bar{P}_i^n) - EU^0(\Upsilon(\psi_i), \rho_b^d) = 0 \tag{6}$$

[11]The assumption on the lapse probability follows Gottlieb and Smetters (2021) in their model of consumer naïveté. The model arrives at similar conclusions if consumers assume themselves to have an average lapse probability.

while costs remain as they were before.

While the determinants of life insurance demand change, the consequences for the market equilibrium are comparable to the case of sophisticated consumers. When the influence of lapsing on costs is stronger than that of mortality, adverse selection results. In the opposite case, advantageous selection appears. This is summarized in the following proposition.

Proposition 3. *When policy lapses result from consumers forgetting to pay their premiums and consumers are naïve about their probabilities to lapse or die then adverse selection results when* $\rho^d B \frac{\partial \rho_i^l}{\partial \psi_i} < (1-\rho^l) B \frac{\partial \rho_i^d}{\partial \psi_i}$ *and advantageous selection results when* $\rho_i^d B \frac{\partial \rho_i^l}{\partial \psi} > (1-\rho_i^l) B \frac{\partial \rho_i^d}{\partial \psi_i}$.

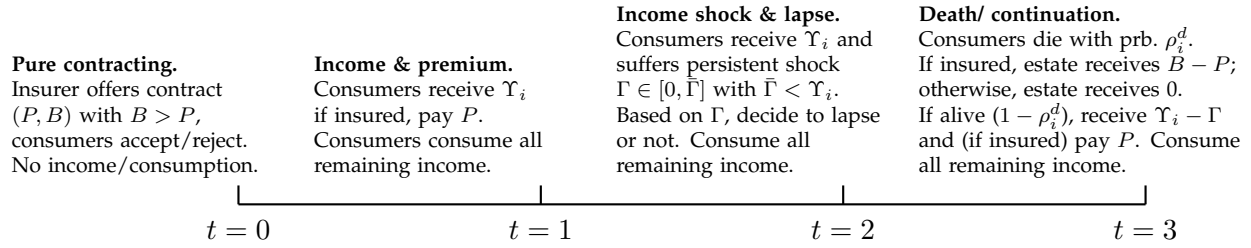
The naïve consumer model acts as a robustness check of our result that both adverse and advantageous selection can appear in the live insurance market. Whether the model is realistic is unclear. There are also a variety of results which oppose the view of naïve consumers when it comes to lapse and death probabilities. Carson et al. (2020) show that life insurance policyholders in their sample select themselves into contracts which are more suited to their own lapse behavior, thus indicating sophistication about their lapse probabilities. Hurd and McGarry (2002) find that subjective mortality estimates predict actual mortality and various studies have found that people make economic choices based on their own mortality (e.g. Jayachandran and Lleras-Muney, 2009).

2.3 Lapsing due to Income Shocks

The second commonly cited reason for consumers to lapse is income shocks leading to liquidity needs, which is often called the "emergency fund" hypothesis (Fier and Liebenberg, 2013; Gottlieb and Smetters, 2021). In this second model, we endogenize the lapse decision by including a persistent income shock that can make lapsing the rational choice for consumers. Insurance contracts and all other aspects of the model are kept as before. We again start with a model in which consumers are sophisticated, being fully informed about their potential income shock and treat the case of naïve consumers afterwards.

Timing: The time line of the model is given in Figure 4. The decision to lapse is now the optimal reaction to the realization of the income shock (denoted Γ).

Figure 4: Time Line of Model when Lapses Originate from Income Shocks



Note: As before, we assume $B > P$ such that the insurance contract is not first-order stochastically dominated. All assets outside the insurance contract are perfectly illiquid.

Consumer utility, costs and type: We assume the shock is distributed with density $h(\Gamma)$ and cumulative distribution function $H(\Gamma)$. The consumers make their lapsation decision optimally given the circumstances. That is, they lapse if $\Gamma \geq \hat{\Gamma}_i$ whereas $\hat{\Gamma}_i$ is defined as that loss amount for which the consumer is indifferent between lapsing and not lapsing:

$$u_2(\Upsilon_i - \hat{\Gamma}_i) + \rho_i^d v_3(\kappa) + (1-\rho_i^d) u_3(\Upsilon_i - \hat{\Gamma}_i) = u_2(\Upsilon_i - \hat{\Gamma}_i - P) + \rho_i^d v_3(\kappa + B - P) + (1-\rho_i^d) u_3(\Upsilon_i - \hat{\Gamma}_i - P). \tag{7}$$

Taken together, the expected utility of having life insurance for the consumer is given by

$$\begin{aligned} EU^{LI}(\Upsilon_i, \rho_i^d, P) = {} & u_1(\Upsilon_i - P) + (1 - H(\hat{\Gamma}_i)) E[u_2(\Upsilon_i - \Gamma) + \rho_i^d v_3(\kappa) + (1 - \rho_i^d) u_3(\Upsilon_i - \Gamma) | \Gamma \geq \hat{\Gamma}_i] \\ & + H(\hat{\Gamma}_i) E[u_2(\Upsilon_i - \Gamma - P) + \rho_i^d v_3(\kappa + B - P) + (1 - \rho_i^d) u_3(\Upsilon_i - \Gamma - P) | \Gamma < \hat{\Gamma}_i]. \end{aligned} \tag{8}$$

Further, the expected utility without life insurance is given by

$$EU^0(\Upsilon_i, \rho_i^d) = u_1(\Upsilon_i) + E[u_2(\Upsilon_i - \Gamma) + \rho_i^d v_3(\kappa) + (1 - \rho_i^d) u_3(\Upsilon_i - \Gamma)]. \tag{9}$$

These definitions naturally extend themselves to the maximum willingness to pay and costs. Specifically, $\bar{\Pi}_i = \bar{P}_i(3 - 2(1 - H(\hat{\Gamma}_i))) = \bar{P}_i(1 + 2H(\hat{\Gamma}_i))$ and $C_i = \rho_i^d H(\hat{\Gamma}_i) B$.

Consumer heterogeneity now has two dimensions Υ_i and ρ_i^d which are collapsed into the scalar-valued type ψ_i. By the same arguments as above, we set $\frac{\partial \Upsilon_i}{\partial \psi_i} \geq 0$ and $\frac{\partial \rho_i^d}{\partial \psi_i} \leq 0$.

Market outcome: Expressing costs and maximum willingness to pay as a function of type, we state the result below. We again see that both adverse and advatangeous selection are possible. The proof in Appendix A further shows that the relative strength of the lapse / income channel is decisive for the market outcome. The proposition, however, also highlights that when there is no influence of type on one of these two dimensions, there can be only adverse selection.

Proposition 4. *When policy lapses are the result of persistent income shocks and consumers are sophisticated about their type, then*

1. *both adverse and advantageous selection can appear in the market and*
2. *when either* $\frac{\partial \Upsilon_i}{\partial \psi_i}$ *or* $\frac{\partial \rho_i^d}{\partial \psi_i}$ *is zero and the other is non-zero, the market features adverse selection.*

Consumer naïveté: There is again a case against the assumption of sophisticated consumers. Respondents in Gottlieb and Smetters (2021) seemingly do not consider their own beliefs about future income shocks when forming a subjective lapse probability. Thus we also consider a model in which consumers in $t = 0$ assume that they will not have an income shock. As before, consumers also assume that their probability of death equals the baseline probability and is unaffected by their type.

In this model, expected utility with and without life insurance are given by

$$EU^{LI}(\Upsilon_i, \rho_b^d, P) = u_1(\Upsilon_i - P) + u_2(\Upsilon_i - P) + \rho_b^d v_3(\kappa + B - P) + (1 - \rho_b^d)u_3(\Upsilon_i - P) \text{ and} \quad (10)$$

$$EU^0(\Upsilon_i, \rho_i^d) = u_1(\Upsilon_i) + E[u_2(\Upsilon_i) + \rho_i^d v_3(\kappa) + (1 - \rho_i^d)u_3(\Upsilon_i)]. \quad (11)$$

Such that the definition of the maximum willingness to pay coincides with that for naïve consumers in the first model. Costs are unaffected by consumer naïveté and are the the same as when consumers are sophisticated regarding their income shocks.

Market outcomes again mirror the results for naïve consumers in the first model. Both adverse and advantageous selection are possible and the market result depends on the relative strength of the lapse and mortality channels on cost. Formally

Proposition 5. *When policy lapses are the result of persistent income shocks and consumers are naïve about their income shocks and their probability of death then adverse selection results when* $\rho^d h(\hat{\Gamma}) \frac{\partial \hat{\Gamma}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} > -\frac{\partial \rho^d}{\partial \psi} \left(H(\hat{\Gamma}) + \rho^d h(\hat{\Gamma}) \frac{\partial \hat{\Gamma}}{\partial \rho^d} \right)$ *and advantageous selection results when the opposite relationship holds.*

2.4 Summary of Findings

Whether policy lapses are caused by consumers forgetting to pay their premiums or by them suffering income shocks, our results show that both adverse and advantageous selection are possible market outcomes. This is also not affected by wether consumers are naïve or sophisticated about their chances of lapsing or dying. The results also make clear that simply measuring the correlation between mortality and insurance demand is insufficient to determine the market outcome. It is possible that this correlation is positive as documented in Cawley and Philipson (1999) or Hedengren and Stratmann (2016), and yet the market outcome is one of adverse selection. A full characterization of the market outcome requires measuring total costs of the insurance company.

3 Data

Data Source: Our primary data source are the 1993-2017 annual statutory reports that life insurers file with the National Association of Insurance Commissioners (NAIC National Association of Insurance Commissioners, 2017). These reports, which cover every licensed life insurer in the U.S., offer comprehensive information on premiums, face amounts, death and surrender benefits, and other financial details at both the state and national levels. We focus on the U.S. individual life insurance market because our framework and empirical strategy target long-term contractual features and related lapse behavior.

We use two exhibits from these filings: the "State Page", which provides state-by-state data on total premiums, the quantity (number of policies and total face amount) of in-force, newly issued, and newly ceased policies as well as data on surrender benefits and death benefits paid to policyholder. We also use the "Exhibit of Life Insurance", which offers similar data on a nationwide

basis, with additional details on the premiums earned for newly issued policies and a detailed breakdown of why policies ceased.

Variable Construction: For our analysis, we will be estimating both a demand function and a cost function. For the demand function, we define quantity as the face value of newly issued policies for each year, which we observe at the firm-state-year level. Prices are more complicated. While we observe total premiums, total face value of policies in force, and the face value of newly written policies at the state level for every firm, we only observe the premiums associated with the newly issued policies at the national level. However, prices for individual life insurance are largely set nationally (Wenning, 2024). Thus, using prices on the national firm-year basis is not a large limitation for our analysis.

Estimating costs is more challenging. The NAIC reports have detailed, firm-state-year level data on cashflow, but there is no way to connect payouts in one period to when that policy was initially purchased. As a workaround, we calculate a (discounted) "future average costs per ceased" variable for each firm-state-year. For firm j in state s in year t, we calculate

$$AC_{jst} = \frac{1}{\sum_{\tau=t+1}^{T} \varrho^{\tau}} \sum_{\tau=t+1}^{T} \varrho^{\tau} \frac{B^{D}_{js\tau} + B^{R}_{js\tau}}{N_{js\tau}} \tag{12}$$

Where $B^{D}_{js\tau}$ and $B^{R}_{js\tau}$ are the death and surrender benefits incurred in year τ, $N_{js\tau}$ is the total face value of policies that "cease" in year τ, and ϱ is the discount factor. We calculate future death, lapse, and surrender rates in a similar manner.

Sample Selection: We begin from the sample of all insurer-state-years with positive issued individual (or "ordinary") life insurance. We then sequentially apply filters to ensure data quality and analytical suitability, requiring: (1) valid positive values for key variables (policy costs, premiums (group and individual), and in-force amounts), (2) plausible rate values between 0 and 1 for lapse, surrender, and mortality rates, and (3) a minimum 5-year observation window to assess long-term outcomes. The impact of each filter on our sample is shown in Table 1. The final sample retains around 25% of original observations, representing 77,652 firm-state-years and $9.2 trillion of issued face value. The largest sample restriction is the requirement that the insurer sells both

individual and group insurance (which is required for our instrument).

Table 1: Sample Selection Steps

Sample Selection Step	Observations	% of Full	Total Issued ($B)	% of Total
Full Sample	316,819	100.0	37,065.34	100.0
Valid Lapse Rate	286,080	90.3	34,941.97	94.3
Valid Death Rate	285,684	90.2	34,935.99	94.3
Valid Surrender Rate	283,567	89.5	34,891.2	94.1
Valid In-Force Amounts	280,073	88.4	34,824.21	94.0
Valid Ind. Premiums	275,204	86.9	34,786.84	93.9
Valid Grp. Premiums	98,519	31.1	12,877.18	34.7
Valid Future Costs	97,852	30.9	12,871.16	34.7
>5 Years Observed	77,652	24.5	9,296.34	25.1

Notes: This table shows the sequential impact of data filters on the sample size. The full sample includes all firm-state-year observations from the NAIC data that issue individual life insurance. Subsequent rows show observations remaining after applying each filter criterion. "Valid" means non-missing, positive, and finite.

Final Sample Descriptive Statistics: Table 2 summarizes key descriptive statistics. The average insurance company issues life insurance coverage with a face value of $101 million per firm-state-year. In contrast, the median life insurer only issues $8 million per year, showing the skewness in size in this industry. It appears because insurers often concentrate their business in a few states, but also sell a small number of policies in several other states. The coverage has a mean future costs when ceased of $25 cents (with a median of $10 cents) per $1 of face value. This is contrasted with a mean individual premium rate of about $2–3 cents per $1 of face value (with a median of 1.14 cents).

The premium income and cost parameters differ by about an order of magnitude. To have profitable life insurance companies, coverage thus needs to cease at a rate of 10% or less, on average. This is is supported by the data. On average, 5% of policies lapse annually, 2% terminate due to surrender and 1.5% due to death. Figure 11 shows these parameters and the expiration rate for our sample, weighted by policies in-force, over time. We can see that the order of the different reasons to cease is the same over time. While the death rate is approximately constant over the whole time period, the surrender rate decreases over time, ending at about half of its 1993 value in 2017. The lapse rate has a slight downward trend and shows a temporary spike during the

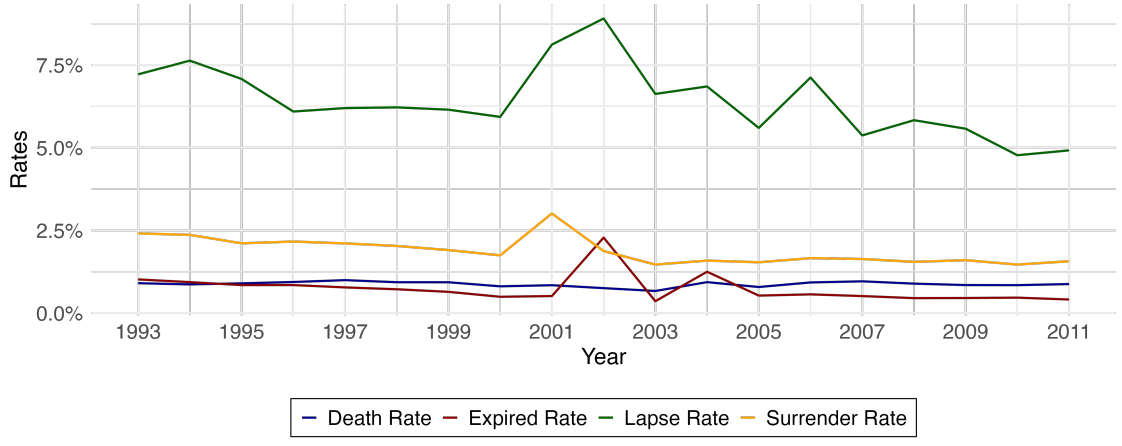
Table 2: Summary Statistics

Variable	Mean	SD	P25	Median	P75
Life Insurance Issued (000,000s)	101.41	292.36	0.8	8.33	55.15
Avg. Future Cost when Ceased	0.25	0.47	0.04	0.1	0.22
Avg. Ind. Life Premium	0.02	0.03	0.01	0.01	0.02
Avg. Grp. Life Premium	0.03	0.06	0.00	0.01	0.02
Future Lapse Rate (%)	4.69	3.87	2.02	3.69	6.03
Future Surrender Rate (%)	2.23	1.23	1.39	1.98	2.89
Future Death Rate (%)	1.52	1.37	0.56	1.1	1.9
In-force Prior Year (000,000s)	758.47	2,239.95	12.23	77.94	407.2
Years Observed	14.59	5.44	10.00	14.00	19.00
Unique:					
Firms:	287				
Firm-Years:	2,446				
Firm-States:	9,868				
Firm-State-Years:	77,652				

Note: This table presents the summary statistics for subsample used in the analysis. All variables (except years observed) have been Winsorized at the 1% and 99% level. Future Cost, and Premiums are all per $ of Face Value. Amount Issued, and Amount In-Force Prior are all in units of $ of Face Value. Future Costs are the average costs per dollar of face value incurred for policy surrender and policyholder death over the next 10 years weighted by policies in force and discounted using a 3% discount rate.

economic downturn of the dot-com bubble.

Figure 5: Policy Cease Rates Over Time



Note: Figure displays the timeline of the average lapse, surrender, death, and expiry rate for our sample, weighted by policies in-force.

4 Estimation Strategy and Results

4.1 Methods

In the reduced form, we start by estimating firm-level demand curve and average cost curves.[12] Assuming (log) linearity, we define the these two curves as follows:

$$ln(Q_{jst}) = \alpha_{jst} + \beta ln(P_{jst}) \tag{13}$$

$$ln(AC_{jst}) = \gamma_{jst} + \delta ln(P_{jst}) \tag{14}$$

where Q_{jst} is the demand for life insurance at firm j in state s and year t, P_{jst} is the premium, AC_{jst} is the average cost curve, where "costs" are defined as the present value of future cash outflows of the associated policy. β is the price-elasticity of demand and δ is the price elasticity of the average cost curve and represents the combination of mortality and lapse-induced selection. The intercept terms α and γ are specific to the firm-state-year observation to reflect the presence of control variables and fixed effects in our estimation. Note that the assumption of log-linearity translates into an assumption of constant price elasticities in both the demand and the cost function.

Suppressing indexes for legibility, we can then express the inverse demand function, that is the function mapping demand to prices, and average cost as a function of demand as

$$P(Q) = exp\left(-\frac{\alpha}{\beta}\right) \times Q^{\frac{1}{\beta}} \tag{15}$$

$$AC(Q) = exp\left(\gamma - \frac{\alpha\delta}{\beta}\right) \times Q^{\frac{\delta}{\beta}} \tag{16}$$

In their initial model, Einav et al. (2010) assume a zero profit equilibrium which would imply

[12] While we estimate firm-level demand curves (which is where our identifying variation is), the welfare costs from adverse selection come from low-risks exiting the market in equilibrium and not from cross-firm moves (Einav and Finkelstein, 2011). The market demand curve will be less elastic than the individual firm demand curves, since it doesn't include cross-price elasticities, but it is ambiguous which direction this will impact welfare estimates. More elastic demand removes more people from the market, but assigns less individual welfare loss while less elastic demand removes fewer, but has a larger welfare loss associated with each individual that loses coverage. In Section 5, we estimate a structural demand model that does account for cross-price elasticities, and find a slightly larger overall welfare loss than the one we find in this section.

that $P = AC(P)$. This assumption, however, is less realistic in the market for life insurance where we know markups to exist (Brown and Goolsbee, 2002). We thus adopt the model of Hackmann et al. (2015) which allows for profits of the firm. Further, our average costs are discounted total outflows of cash, while our premiums are measured annually. To adjust for both profits and the difference in time frame, we introduce a multiplicative firm-state-year specific factor ζ on average costs. The equilibrium average cost curve of firm i in state s at time t then has the intercept $\tilde{\gamma} = \gamma - ln(\zeta)$. Here, and below, the tilde indicates the adjusted version of the cost curve. To obtain its value, we set the market equilibrium for a given firm in a given state-year as the price and quantity observed in the data. Because this point defines both demand and adjusted average cost, we can use it to get the value of α through Equation (13). For this, we use the fact that we know β and δ, which are not firm-state-year specific, from our estimations. Leveraging Equation (16), we obtain $\tilde{\gamma}$ through[13]

$$\tilde{\gamma} = ln\left(\frac{P}{Q^{\frac{\delta}{\beta}}}\right) + \frac{\alpha\delta}{\beta}. \tag{17}$$

This allows us to express the adjusted average cost function as

$$\widetilde{AC}(q) = exp\left(\tilde{\gamma} - \frac{\alpha\delta}{\beta}\right) \times Q^{\frac{\delta}{\beta}} \tag{18}$$

To obtain welfare estimates, we further require the marginal cost function. Again, we use an adjusted version. We know from Einav et al. (2010) that

$$\widetilde{MC}(P) = \left(\frac{\partial Q(P)}{\partial P}\right)^{-1} \frac{\partial \widetilde{AC}(P) Q(P)}{\partial P} \tag{19}$$

$$\widetilde{MC}(P) = \frac{\delta + \beta}{\beta} \times \widetilde{AC}(P) \tag{20}$$

We substitute Equation (18) and obtain the adjusted marginal cost curve as

$$\widetilde{MC}(Q) = \frac{\delta + \beta}{\beta} \times exp\left(\tilde{\gamma} - \frac{\alpha\delta}{\beta}\right) \times Q^{\frac{\delta}{\beta}}. \tag{21}$$

[13] An alternative way to approach this is that in equilibrium it holds that $\widetilde{AC}_{\text{eq}} = P(Q_{\text{eq}})$. One can thus calculate a factor $\iota = \frac{AC(Q_{\text{eq}})}{P(Q_{\text{eq}})}$ which implies that $\widetilde{AC} = exp(\gamma)P^{\delta}/\iota$, such that $\tilde{\gamma}$ can be found from $\tilde{\gamma} = \gamma - ln(\iota)$.

We have set the equilibrium demand quantity Q_{eq} such that it implies $\widetilde{AC}(Q_{\text{eq}}) = P(Q_{\text{eq}})$ mechanically. The efficient level of demand, Q_{eff}, is given by $\widetilde{MC}(Q_{\text{eff}}) = P(Q_{\text{eff}})$, which renders

$$Q_{\text{eff}} = \left(\frac{\beta}{\delta+\beta}\right)^{\frac{\beta}{\delta-1}} \times Q_{\text{eq}} \tag{22}$$

The dead-weight welfare loss is then given by

$$DWL = \int_{Q_{\text{eq}}}^{Q_{\text{eff}}} P(Q)dQ - \int_{Q_{eqm}}^{Q_{\text{eff}}} MC(Q)dQ \tag{23}$$

$$DWL = \left[exp\left(-\frac{\alpha}{\beta}\right)\frac{\beta}{\beta+1}Q^{\frac{\beta+1}{\beta}} - exp\left(\tilde{\gamma} - \frac{\alpha\delta}{\beta}\right)Q^{\frac{\delta+\beta}{\beta}}\right]_{Q_{\text{eq}}}^{Q_{\text{eff}}} \tag{24}$$

Now that we have defined the dead-weight loss (DWL) as a function of parameters, we next directly estimate β and δ, which we can then combine along with observed values of Q_{eq} and P_{eq} for each market to back out α and $\tilde{\gamma}$ and calculate the firm-state-year specific value of DWL.

The central challenge to these estimations are unobserved product characteristics that jointly influence price and cost/demand. To account for these unobserved factors, we use a "Hausman"-inspired instrument (Hausman, 1996). The instrument we use is group-insurance premiums for plans offered by the same firm in the same year. Because our analysis focuses on the individual market, the goal of this instrument is to capture firm-level cost shocks that are external to the individual market.[14] These cost shocks could either be at the general firm level, such as wages or taxes, or they could be at the group-insurance level, such as unexpectedly high costs at a few large clients, and spill over. This instrument follows an extensive recent tradition of instrumenting insurance prices with prices/costs by the same firm in other markets (e.g., Koijen and Yogo, 2016, 2015; Ge, 2022; Gao et al., 2023; Wenning, 2024).

We exploit this instrument using two-stage least squares (2SLS). The first stage regresses the endogenous variable, individual life premiums, on the instrument, group life premiums. For-

[14] Any local mortality shocks that impact both group and individual insurance will be picked up by the state-year fixed effects.

mally, we estimate

$$\ln(P_{jt}) = \pi \ln(P_{jt}^{\text{group}}) + \theta_p \mathbf{X}_{jt} + \phi_{p,js} + \eta_{p,st} + \varepsilon_{p,jt} \tag{25}$$

where $\ln(P_{jt})$ is the log of individual life premiums for firm j in year t, $\ln(P_{jt}^{\text{group}})$ is the log of group life premiums, which serves as the instrument, for the same firm-year and $\mathbf{X}_{jt}$ includes firm-year-specific control variables.[15] Firm-state fixed effects (ϕ_{js}) and state-year fixed effects (η_{st}) are also included to isolate identifying variation to the firm-year level to prevent biasing the estimates through general, firm-level, unobserved quality (Nevo, 2001). The error term $\varepsilon_{p,jt}$ then captures the portion of premiums that is unexplained by the instrument and the included controls / fixed effects. Note that we estimate this first stage at the firm-state-year level because that is the level of our second stage. This explains why the fixed effects are defined state-wise even though all variation of the other variables in Equation (25) is on the firm-year level.

The predicted values from this first stage, $\widehat{\ln(P_{jt})}$, are then used in the second stage to estimate the causal relationship between premiums and the outcome variables of interest. Formally, the second-stage regressions are

$$\ln(Q_{jst}) = \beta \widehat{\ln(P_{jt})} + \theta_q \mathbf{X}_{jt} + \phi_{q,js} + \eta_{q,st} + \varepsilon_{q,jst} \tag{26}$$

$$\ln(AC_{jst}) = \delta \widehat{\ln(P_{jt})} + \theta_c \mathbf{X}_{jt} + \phi_{c,js} + \eta_{c,st} + \varepsilon_{c,jst} \tag{27}$$

where $\ln(Q_{jst})$ and $\ln(AC_{jst})$ represent (logged) quantity and the (logged) present value of future average costs, respectively, for firm j in state s in year t, and $\widehat{\ln(P_{jt})}$ is the predicted log of individual life premiums from the first stage. Because we use a log-linear specification, we can use discounted future average costs in Equation (27) rather than the time-frame and profit adjusted cost and still obtain the same estimate for δ.

As in the first stage, the model also includes firm-year controls $\mathbf{X}_{jt}$, firm-state fixed effects

[15] Firms set life insurance premiums largely at the national level rather than separately for each state that they operate in. See Wenning (2024) for more discussion. Because our fixed effects do not pick up firm-specific characteristics that change over time and because many life insurance firms expand or contract their business over our long sample period, our firm-year-specific control variables need to include firm size for proper identification.

(ϕ_{js}), and state-year fixed effects (η_{st}). The error terms ε_{jst} capture idiosyncratic shocks and are uncorrelated with the error term from the first stage. Following convention in demand estimation, we cluster our standard errors at the "market" level, which here is state-year (e.g., Gentzkow et al., 2014; Martin and Yurukoglu, 2017; Miller and Weinberg, 2017; Azar et al., 2018).[16] Our price instrument also allows to separately test the three potential selection channels through the same set up. We estimate each channel through

$$\ln(L_{jst}) = \lambda\widehat{\ln(P_{jt})} + \theta_l \mathbf{X}_{jt} + \phi_{l,js} + \eta_{l,st} + \varepsilon_{l,jst} \tag{28}$$

$$\ln(D_{jst}) = \omega\widehat{\ln(P_{jt})} + \theta_d \mathbf{X}_{jt} + \phi_{d,js} + \eta_{d,st} + \varepsilon_{d,jst} \tag{29}$$

$$\ln(R_{jst}) = \upsilon\widehat{\ln(P_{jt})} + \theta_r \mathbf{X}_{jt} + \phi_{r,js} + \eta_{r,st} + \varepsilon_{r,jst} \tag{30}$$

Where $\ln(L_{jst})$, $\ln(D_{jst})$, and $\ln(R_{jst})$ the (logged) average future lapse, death, and surrender rates. λ, ω, and υ then represent how random shocks in firms' prices influence the primary drivers of firms' costs. If one of the coefficients is zero, then that is evidence that consumers do not select along that dimension.

4.2 Demand and Cost Estimation Results

Table 3 present the results of our two-stage least squares (2SLS) analysis. Column (1) shows the first-stage regression results, while columns (2) and (3) display the second-stage results examining the impact of premiums on future costs and demand in the individual life insurance market.

The first-stage results demonstrate the strength of our instrument, group life insurance premiums. A 1% increase in group premiums is associated with a 0.05% increase in individual premiums, indicating a small, but significant, spillover of costs between markets. The strength of this relationship is evidenced by the high Kleibergen-Paap first-stage F-statistics (>650), supporting the relevance condition of our instrumental variable.

Column (2) of Table 3 estimates the demand model and reveals a price elasticity β of -0.904,

[16] Our results are largely robust to clustering at the "unit of random variation" (firm-year, as recommended by Abadie et al., 2023) with the only exception being that the estimate on future surrender values becomes insignificant.

Table 3: Estimation Results: Demand and Cost

	log(P)	log(Q)	log(AC)
log(Avg. Group Premium)	0.047***	–	–
	(0.002)	–	–
log(P) (fitted)	–	-0.904***	0.125***
	–	(0.093)	(0.017)
Firm Size Control	Yes	Yes	Yes
Firm-State FE	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes
Years Observed FE	Yes	Yes	Yes
Observations	77,652	77,652	77,652
KP First-stage F	–	668.9	668.9

* p < 0.1, ** p < 0.05, *** p < 0.01.

Notes: The instrumental variable for all models is *ln(Average Group Premium)*, which is the premium for the group market. In the second stage, the main dependent and independent variables are on the firm-state-year level. The estimation equations are given in Equation (25) for the first stage and Equations (26) and (27) for the second stage. P denotes the average individual life premium per $ of face value. Current demand (Q) is given by the face value of new policies sold. Future average costs (AC) are discounted monetary outflows in future periods as described in Equation (12). The size control is the face vale of in-force life insurance prior to the current year. Data is taken from annual reports submitted to the National Association of Insurance Commissioners. Standard errors, clustered by State-Year, are in parentheses.

indicating that a 1% increase in premiums leads to about a 0.9% decrease in the quantity of life insurance demanded. Column (3) shows the cost function estimation and our estimates of δ. We find that a 1% increase in premiums is associated with approximately a 0.125% increase in future costs for insurers. The positive coefficient suggests that as premiums rise, the insurer's risk pool changes in a way that increases expected future payouts, indicating adverse selection.

Table 4 isolates the two potential channels for selection. In column (1), we find that higher premiums are associated with lower lapse rates. A 1% increase in premiums corresponds to approximately 0.145% decrease in future lapse rates and a 0.034% decrease in future surrender rates. This result suggests that lapse and surrender behavior is a channel through which selection operates in this market. As detailed in Section 2 it does not necessarily indicate that people incorporate knowledge of their future lapse rates into their current purchasing decisions. In contrast, we find no statistically significant relationship between premiums and death rates with a point estimate close to zero. This tightly estimated null result is not a placebo test, but rather indicates that mortality risk is not a primary channel through which selection is occurring in this market. The

absence of a relationship between premiums and death rates, coupled with the significant relationship between premiums and lapse and surrender rates, suggests that the observed adverse selection is primarily driven by lapse and surrender behavior rather than mortality risk.

Table 4: Estimation Results: Channels

	log(Lapse)	log(Death)	log(Surr.)
log(P) (fitted)	-0.145***	0.003	-0.034***
	(0.014)	(0.009)	(0.008)
Firm Size Control	Yes	Yes	Yes
Firm-State FE	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes
Years Observed FE	Yes	Yes	Yes
Observations	77,652	77,652	77,652
KP First-stage F	668.9	668.9	668.9

* p < 0.1, ** p < 0.05, *** p < 0.01.
Notes: The instrumental variable for all models is *log(Average Group Premium)*, which is the premium for the group market. The main dependent and independent variables are on the firm-year level and are the (discounted) observed future rates at which existing policies either lapse, surrender, or end in death. The size control is the face vale of in-force life insurance prior to the current year. Data is taken from annual reports submitted to the National Association of Insurance Commissioners. Standard errors, clustered by State-Year, are in parentheses.

Our estimate of a single company's life insurance demand elasticity is informative in its own right. In a perfectly competitive market, an exogenous price increase by one firm should lead to a strong decline in the demand for products of that firm and a corresponding increase in demand for the products of other firms. However, while we do see decreasing demand, it is slightly inelastic. This implies significant frictions on life insurance markets. The nature of these frictions is unclear and demands further analysis. Previous evidence from other insurance markets shows that the role of distribution channels (Collier and Ragin, 2020) and advertising (Aizawa and Kim, 2018) could be significant, but there is also conflicting evidence for the latter (Shapiro, 2020). The magnitude of these (and potential other) frictions are as of yet not explored for life insurance markets, but our research implies that they do exist.

4.3 Welfare Results

We now use our estimated coefficients from Table 3 to calculate the welfare consequences of adverse selection in the US life insurance market according to the method laid out in Section 4.1. The resulting parameters provide us with demand, average cost, and marginal cost curves for each firm-state-year observation.[17] One such observation is displayed in panel (a) of Figure 6. It shows the dead-weight loss as a consequence of the marginal cost curve lying below the average cost curve which is a consequence of average costs being increasing in price. We can see in this picture that adverse selection leads to less policies sold in equilibrium than would be the efficient outcome in a world without information asymmetries. Due to the structure of our estimation, the relative decrease in the number of policies due to adverse selection is the same in every market at about 14.2%. This number is constant across all observations due to our assumption of constant elasticities in the demand and cost curves.

The exact amount of dead-weight loss measured for each observation varies considerably. As can be seen in the second row of Table 5, the average firm-state observation shows a dead-weight loss of about $34,900, but this value can be essentially 0 or higher than $1 million per firm.[18] This is not surprising, given that many firms have states in which they write relatively few policies. Aggregated across all firms, annual dead-weight losses in the initial year of purchase, shown in the third row of Table 5, have an average value of $142.79 million and are not skewed very strongly, showing that this mean value is not driven by outliers.

The fourth and fifth rows of Table 5 are a back-of-the-envelope estimate for the total NPV of the DWL and is calculated by projecting the purchase-year DWL into the future and discounting by both a time discount rate as well as the project rate of policy "ceasing".[19] The total NPV estimates are almost 10 times as large as for the purchase-year, with the average annual aggregate now at about 1.29 billion dollars a year.

Panel (b) of Figure 6 shows the development of the inflation adjusted dead-weight loss in pur-

[17] Note that this is the welfare loss for the year that the policy is purchased. We extend the estimates for the full policy timeline below.

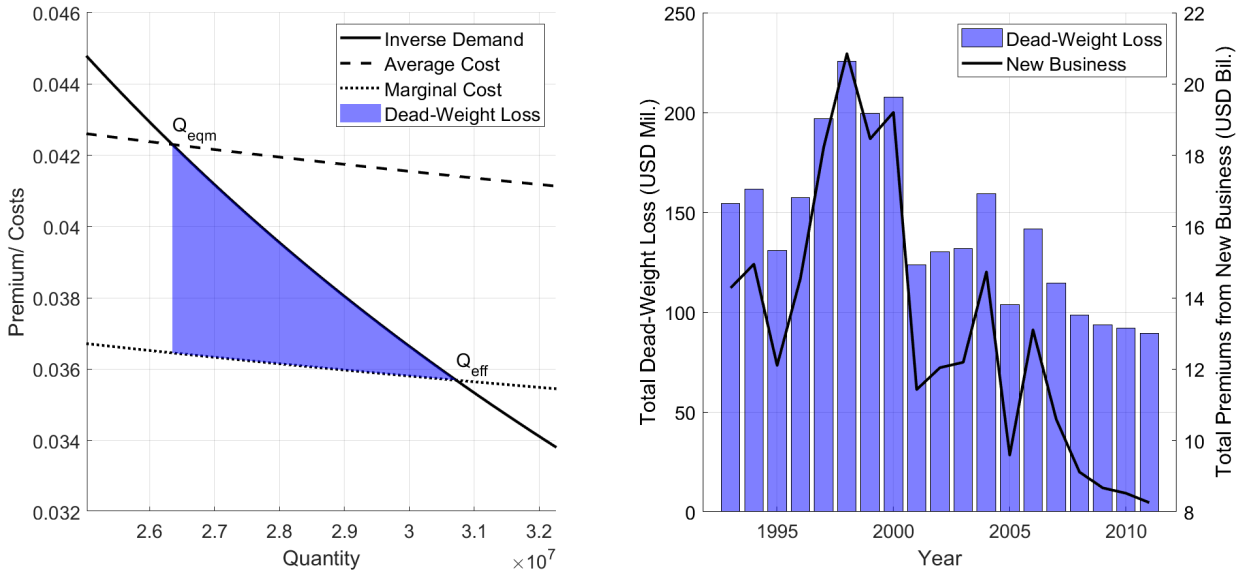
[18] Numbers are inflation adjusted to reflect 2023 values and are skewed highly across firm-state observations.

[19] See Appendix B for full details on the calculation.

Figure 6: Graphical Display of Welfare Results

(a) Market situation of a sample firm: Aetna Life, in the state of New York in year 1998

(b) Development of annual DWL implied by new business over time



Note: Panel (a) of the figure illustrates the demand, average cost, and marginal cost curves for a sample firm-state-year observation. Curves are calculated according to the assumptions and methods laid out in Section 4.1. Quantities are expressed in US$ policy face values, while premiums and costs are expressed in US$ per US$ policy face value. Panel (b) shows the development of the estimated aggregate dead-weight loss over time. On the secondary y-axis, the panel also reports the development of new business in the form of total annual premiums earned.

chase year over time. The figure shows a slightly declining trend in the absolute numbers. However, the figure shows an even stronger decline in the total premiums from new business, mirroring the development of fewer US households having life insurance coverage over the years. Relatively speaking, the problem of adverse selection in life insurance thus seems to have increased rather than decreased.

$1.29 billion in annual welfare losses shows how problematic the issue of lapsation-based adverse selection in life insurance is. This is the case even though Einav et al. (2010) equilibria commonly lead to relatively small estimates of absolute efficiency losses.[20] Moreover, our reduced-form estimates likely underestimate the welfare costs of moral hazard. This is due to the nature of how we calculate costs on the firm-state-year level. Our estimation implicitly assumes that a

[20]In the original study proposing the equilibrium, the authors themselves only report annual efficiency losses of about $10 per policyholder in health insurance.

Table 5: Results of Dead-Weight Loss calculations

	Mean	25th Pctl	Median	75th Pctl
Share of Eff. Q (%)	85.76 [83.57, 89.63]			
DWL in Purchase Year:				
Firm-State-Year ($000s)	34.9 [18.4, 48.7]	0.4 [0.2, 0.5]	3.3 [1.7, 4.5]	19.1 [10.0, 26.6]
Annual Aggregate ($000,000s)	142.79 [75.18, 198.94]	109.15 [57.46, 152.07]	131.88 [69.43, 183.73]	160.47 [84.48, 223.57]
NPV of Total DWL:				
Firm-State-Year ($000s)	316.3 [166.5, 440.6]	3.3 [1.7, 4.6]	29.5 [15.5, 41.1]	172.5 [90.8, 240.3]
Annual Aggregate ($000,000s)	1292.55 [680.49, 1800.79]	988.03 [520.16, 1376.52]	1193.73 [628.46, 1663.10]	1452.58 [764.73, 2023.73]

Note: Aggregate results from the dead-weight loss calculations according to the methods laid out in Section 4.1. Dollar values are inflation adjusted to reflect prices in 2023. The first row displays the share of policies sold in equilibrium when compared to the efficient market solution: $Q_{\text{eq}}/Q_{\text{eff}}$. There are no summary statistics other than the mean, because this value is the same across all observations due to our assumption of constant elasticities in the demand and cost curves. The second row reports the mean and quartiles for the dead-weight loss in the year of purchase across firms, averaged over all years that they appear in the data. The second row reports the mean and quartiles for the dead-weight loss in the year of purchase of the aggregated welfare losses on the annual level. The last two rows report the estimated NPV of the total dead-weight loss over the life of the contract at again the firm-state-year level and then aggregated to an annual level. 90% confidence intervals for each statistic, derived from block-bootstrapping at the Firm-Year Level, are in brackets.

change in price affects the costs of all insurance contracts in the firm's portfolio. In reality, however, it will only affect the costs of new contracts from this point on. The inertia of the insurance portfolio is not modeled appropriately in our reduced-form model, and thus the price elasticity of costs is likely understated in Table 3 leading to lower estimates of the dead-weight loss. Even with all these associated caveats, the NPV of efficiency losses in new business exceeds one billion dollars a year. In relative terms, the adverse selection reported here leads the market to shrink by about one sixth compared to the first-best solution. This underlines the magnitude of the effect of adverse selection in life insurance.

Table 5 provides standard errors for the dead-weight loss estimates which are derived from block-bootstrapping the estimation on the Firm-Year level and using the coefficients from each iteration to calculate the efficiency losses. While the confidence intervals are relatively large, they do not include 0, mirroring the result from Table 3 that average cost are robustly estimated to be

increasing in premiums.

4.4 Robustness

The confidence intervals provided in Table 5 illustrate the robustness of our dead-weight loss estimates against estimation errors. In this section, we explore the robustness of our welfare estimates to quantity rationing, research design choices, and functional form assumptions.

4.4.1 Robustness to Quantity Rationing

Could insurers mitigate adverse selection through non-linear pricing schedules that effectively screen customers by income? In practice, life insurance premiums do not scale linearly with face value, with larger policies having a lower per dollar-of-face-value price. This non-linear pricing structure could allow insurers to partially screen customers based on their intensive-margin demand for coverage, which correlates with wealth and income. To address this concern, we examine whether our main results are robust to controlling for the average policy size purchased by each firm's customers.

Table 6 presents our main specifications from Tables 3 and 4 with log(Policy Size) as an additional control variable. Since our identification strategy relies on cost shock spillovers from the group insurance market to the individual market, controlling for average policy size should not materially affect our core estimates except through the demand channel, where larger policies mechanically represent more coverage purchased. This is precisely what we find. The coefficient on log(Policy Size) in the demand equation is large and positive (0.634), capturing the extensive margin effect. However, the inclusion of this control changes our main price elasticity estimates minimally—demand elasticity increases by 27.2% (from -0.904 to -0.658), while the cost elasticity changes by only 2.5% and the lapse channel only changes by 2.1%.[21]

While we abstract from the complex contract design mechanisms insurers use to screen cus-

[21]The mortality coefficient changes by a large -47.8%, but this represents a movement from an already insignificant 0.004 to 0.002, both economically and statistically indistinguishable from zero.

Table 6: Robustness to Quantity Rationing: Combined Estimation Results

	log(P)	log(Q)	log(AC)	log(Lapse)	log(Death)	log(Surr.)
log(Group Prem)	0.045***	–	–	–	–	–
	(0.002)	–	–	–	–	–
log(P) (fitted)	–	-0.658***	0.128***	-0.142***	0.002	-0.032***
	–	(0.085)	(0.017)	(0.015)	(0.009)	(0.008)
log(Policy Size)	-0.086***	0.634***	0.008***	0.008***	-0.004***	0.005***
	(0.004)	(0.012)	(0.002)	(0.002)	(0.001)	(0.001)
% Δ in Elasticity	-3.3%	27.2%	2.5%	2.1%	-47.8%	5.8%
Firm Size Control	Yes	Yes	Yes	Yes	Yes	Yes
Firm-State FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Years Observed FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	77,652	77,652	77,652	77,652	77,652	77,652
KP First-stage F	–	610.8	610.8	610.8	610.8	610.8

* p < 0.1, ** p < 0.05, *** p < 0.01. Notes: This table presents robustness results combining all six models in a single table. Potential quantity rationing is accounted for by controlling for log(Policy Size). The instrumental variable for all models is log(Average Group Premium). P denotes the average individual life premium per $ of face value. Current demand (Q) is given by the face value of new policies sold. Future average costs (AC) are discounted monetary outflows in future periods as described in Equation (12). Standard errors are clustered by market (State-Year) and reported in parentheses.

tomers in our main analysis, this represents a rich area for future research. The literature on optimal insurance contract design under asymmetric information (e.g., Azevedo and Gottlieb, 2017; Geruso et al., 2019) suggests that insurers have multiple tools beyond simple linear pricing to address selection problems. Understanding how these mechanisms interact with lapse-based selection in life insurance markets would provide valuable insights.

4.4.2 Robustness to Research Design

To check the robustness of our results to various research design choices, we estimate a specification curve (Simonsohn et al., 2020). A specification curve visually illustrates how coefficient estimates shift when the researcher adopts different modeling decisions. Figure 7 plots our estimated Q_{eq} as a percent of Q_{eff}, with 90% confidence intervals, from 384 sets of regressions featuring different combinations of modeling choices. We present the impacts of our choices of (1) the

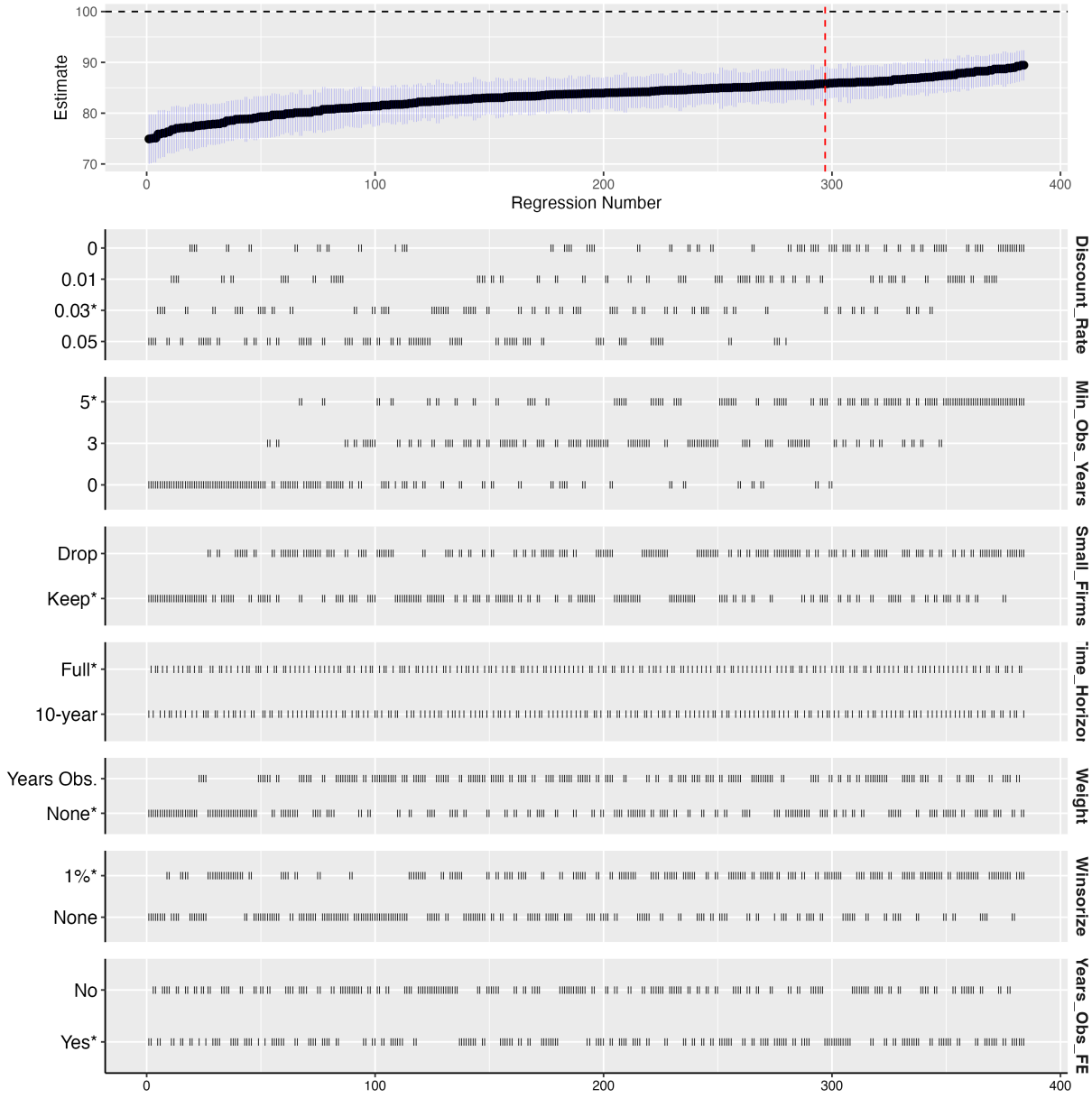
discount rate for future costs; (2) the minimum observed future years to include; (3) whether or not to include small firms; (4) using the full observed future values or cutting off after 10 years; (5) whether or not to weight the regressions based on how long firms are observed (with the idea that a longer observation period allows us to more accurately estimate future costs changes); (6) whether or not to Winsorize the data; and (7) whether or not to include fixed effects for how many years we observe each firm.

The figure can be interpreted in two parts. The top portion plots the estimated Q loss from the regression with the most amount lost on the left to the regression and the least amount lost on the right. This can be interpreted as a confidence interval of sorts: our "strongest" specification features an equilibrium Q of around 75% of the efficient one while our "weakest" specification features an equilibrium Q of nearly 90% of the efficient Q and thus less welfare loss. The vertical, red dashed line shows the specification choice that we use in our primary models. Since it is towards the right of the graph, our choices can be interpreted as conservative.

The lower portion of the figure shows how each choice impacts the strength of our estimates. Each tick in the bottom panel of Figure 7 represents the specification for that regression and that choice. Options that appear more often on the left of the figure tend to increase the welfare loss while those on the right tend to decrease it. Only three choices have a meaningful effect: the discount rate, the minimum allowed observed years, and keeping small firms. For the discount rate, our preferred choice of 0.03 is slightly less conservative than either 0.01 or 0, however the most noticeable jump is from 0.03 to 0.05. The most important choice is limiting how far into the future we have to observe firms for them to be included. Our base choice of 5 years appears extremely conservative, likely since most lapses happen in the first few years of the contract. Finally, our choice to keep small firms appears slightly less conservative than dropping them would be.

We perform the same exercise for all of the other regressions. In the interest of brevity, we delegate the curves themselves to Online Appendix E and summarize their results here in Table 7. The primary take-way is that nearly all of our results are extremely robust to specification. This includes our null result on mortality. While it is possible to create a specification where the mortality channel is statistically significant, the 95th percentile is still less than a third of the lapse channel

Figure 7: Specification Curve: Q_{eq} as a Percentage of Q_{eff}



Note: Each point is one of 384 estimates of $Q_{\text{eq}}/Q_{\text{eff}}$ obtained by jointly varying seven modeling choices: (i) discount rate (0, 0.01, 0.03, 0.05); (ii) minimum observable horizon (3–6 years); (iii) inclusion of small firms; (iv) truncation at 10 future years; (v) duration weighting; (vi) Winsorization; and (vii) horizon fixed effects. Whiskers denote 90 % confidence intervals. The red dashed line marks the baseline specification and *s denote the specific choices.

in our preferred specification. The one result that does not appear robust is the surrender channel. While 77% of the specifications are statistically significant, a number of those include cases where the coefficient is positive instead of negative. However, this is relatively immaterial for our overall

results given that surrender is much less likely than lapsing and has less influence on policy costs. In summary, our finding of adverse selection in life insurance through the lapsation channel but not through the mortality channel appears extremely robust to research design choices.

Table 7: Specification Curve Summary Statistics

Variable	Pref. Coef.	% Significant	5th Pctl.	25th Pctl.	75th Pctl.	95th Pctl.
Efficient Quantity (%)	85.799	100.0%	77.264	81.289	85.557	88.271
Future Cost	0.125	100.0%	0.105	0.127	0.161	0.188
Demand	-0.906	100.0%	-1.234	-1.115	-0.946	-0.830
Lapse Channel	-0.145	100.0%	-0.249	-0.229	-0.157	-0.126
Mort. Channel	0.003	29.2%	-0.022	-0.009	0.016	0.034
Surr. Channel	-0.034	77.1%	-0.065	-0.050	-0.017	0.037

Note: Each row summarizes the distribution of price coefficients across all specifications for the corresponding dependent variable. 'Pref. Coef.' refers to the coefficient obtained under a baseline specification (3% discount rate, no weighting, full horizon, FE for years observed, 1% Winsorization, a minimum of 5 observed years, and keeping small firms). '% Significant' is the share of specifications for which the 90% confidence interval excludes zero. The 5th, 25th, 75th, and 95th percentiles describe the empirical distribution of the point estimates across all specifications.

4.4.3 Robustness to Functional Form

Our result relies on the log-linear specification of our estimation as it is presented in Equations (26) and (27). Kang and Vasserman (2025) raise the important issue that in settings where demand functions are estimated based upon few points, welfare implications are extrapolated from average treatment effects through functional form assumptions. Our setting is an especially pronounced example. For every insurance company, in every market, in every year; we infer welfare effects through counterfactual statements extrapolated from two slope measurements and a firm-market-year specific data point. Our counterfactuals are thus identified at not few, but one data point and rely on the functional form assumption of the log-linear specification.

Kang and Vasserman (2025) suggest a procedure for quantifying the robustness of the welfare implications of a limited estimation. In their procedure, they allow for arbitrary functional forms under the restriction that between two observed points, the local slope does not deviate from the average slope by more that a proportional factor k. Unfortunately, their approach does not directly apply to Akerlovian selection markets. We describe both their approach and the problems

in applying it to selection markets in more detail in Appendix D. The most crucial issue is that Akerlovian selection markets require two functional form assumptions – both the demand and cost curve – to determine the welfare consequences, and these functions interact in non-trivial ways (as can be seen Equation (A46)).

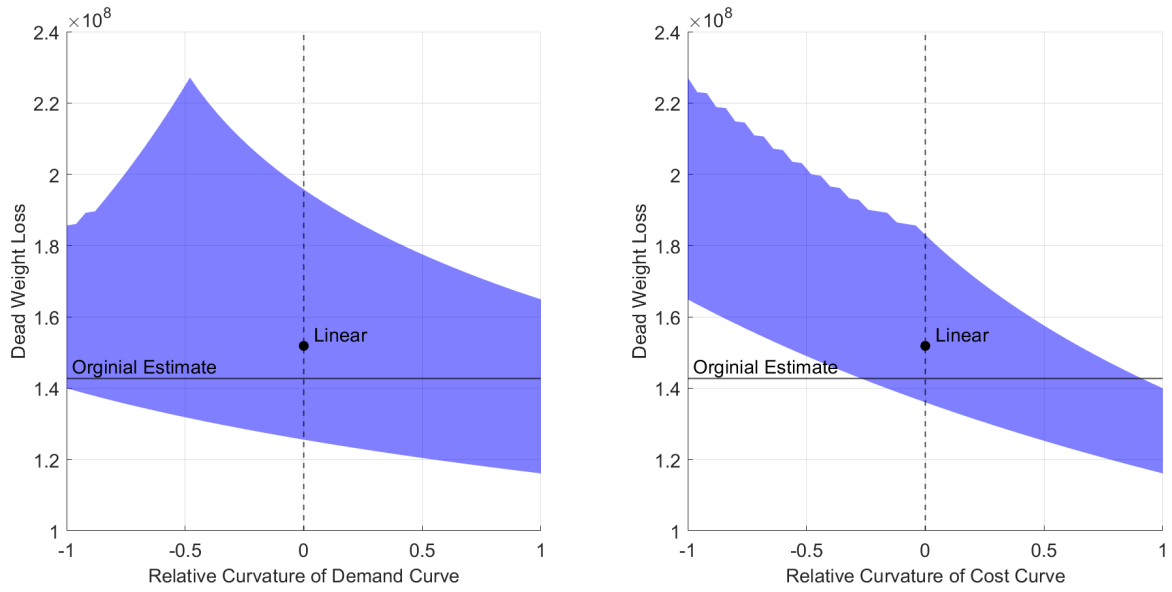
We thus propose an alternative robustness measure for functional form assumptions when estimating the dead-weight loss in an EFC framework. In the same spirit as Kang and Vasserman (2025), we stick to the idea that different functional forms lead to different welfare outcomes. However, to solve the issue of the complicated interactions between demand and cost curves, we restrict the functional forms of each curve to second-order polynomials and solve for the welfare bounds numerically. While still allowing us to consider both concave and convex functions for both curves, this assumption considerably simplifies the problem's numerical complexity. The exact procedure is described in Appendix D. Conceptually, we determine the most concave and convex feasible demand and average cost functions given our empirical estimates. We then use a grid between these two functions to calculate the resulting dead-weight loss for various combinations of functional forms. This lets us determine how much the curvature of the demand and cost functions influences our welfare estimates.

The results of this analysis are summarized graphically in Figure 8. Estimates of the average annual dead-weight loss for new business range between \$120 and \$220 million. This range is smaller than the confidence interval due to potential estimation error as reported in Table 5, but shifted upward due to the conservativeness of the log-linear specification. Even compared to the linear specification that is, for example, used in Einav et al. (2010), our specification underreports the dead-weight loss by about 5%. The figure thus provides confidence in the fact that the functional form assumption made in our estimation is not necessarily decisive for our conclusions about welfare consequences of adverse selection through lapsation. It should be noted that the boundaries on the ranges in both panels are sometimes ragged. This is because some combinations of curvatures for the demand and cost curve do not lead to an efficient market outcome of the form $P(Q_{\text{eff}}) = MC(Q_{\text{eff}})$ for positive prices. In that case, we ignore the combination of functional forms and use the combination with the highest dead-weight loss and positive prices instead. This

Figure 8: Range of Dead-Weight-Loss due to Variations in Functional Assumptions

(a) Range as a function of curvature in the demand curve

(b) Range as a function of curvature in the average cost curve



Note: The figure presents the range of possible annual aggregate dead-weight losses due to adverse selection when varying the functional form assumptions on the demand and the average cost curve. For each curve, we assume a second-order polynomial function and vary the curvature of the function between the minimum and the maximum as is described in Appendix D. Each panel varies the curvature of one of the functions on the x-axis and displays the range due to varying the other function on the y-axis. The linear specification where curvature is zero for both curves is indicated in each panel. We also show the dead-weight loss due to our empirical specification described in Table 5.

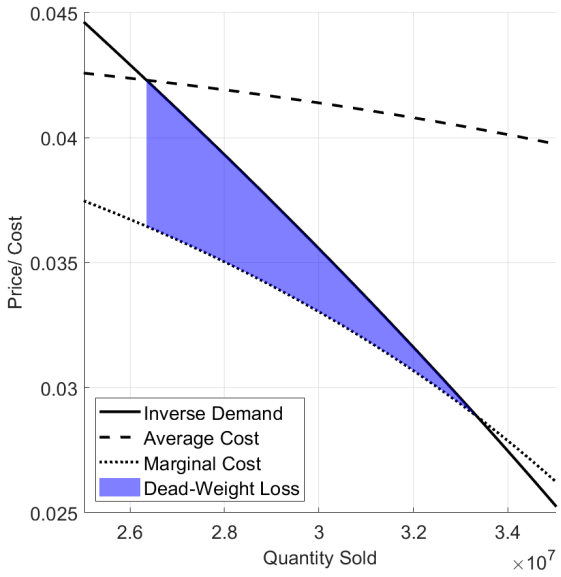
case is particularly likely to appear for very concave demand functions.

Under the caveat of an existing efficient market equilibrium at positive prices, Figure 8 shows that the most extreme values for the dead-weight loss are achieved by the most extreme curvatures assumed for the demand and cost functions. Specifically, assuming the most concave possible demand and average cost function leads to the maximum possible dead-weight loss, while using the most convex functions leads to the minimum possible value. To show the consequences of these functional forms, Figure 9 displays the two resulting market equilibria for the firm Aetna Life in New York State in 1998, the same observation we used as an example for the original estimate in Figure 6.

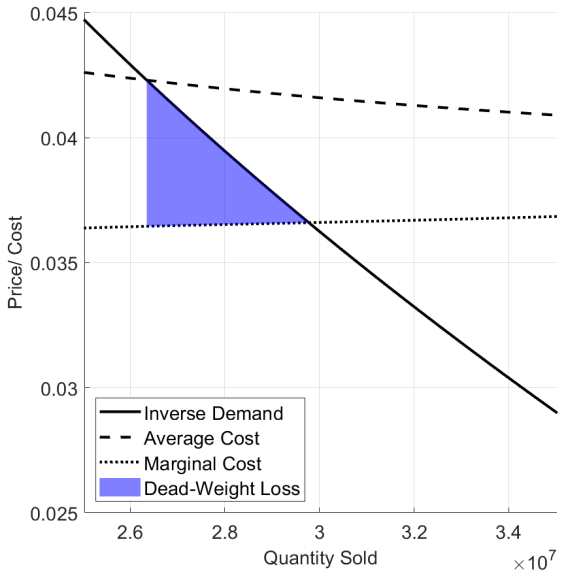
Considering both panels in combination, we can see that changing the functional form does

Figure 9: Possible extremes of dead-weight loss due to variation in functional form for a sample firm

(a) Maximum dead-weight loss



(b) Minimum dead-weight loss



Note: The figure presents the largest (panel(a)) and smallest (panel (b)) dead-weight loss due to variations in the curvature of the demand and average cost function for the firm Aetna Life in the state of New York in 1998. It thus shows the possible extremes of the sample firm shown in panel (a) of Figure 6. For each curve, we assume a second-order polynomial function with the parameter range chosen as is described in Appendix D. The largest possible dead-weight loss results from the most concave demand and average cost curves, while the smallest results from the most convex curves.

not shift the level of the marginal cost curve, but rather tilts it around the point defined by the observed equilibrium quantity. In other words, our point estimate exactly defines the welfare loss of the marginal consumer and our functional form assumption and estimated slopes allow us to impute the welfare loss of the inframarginal consumers. In the appendix, we show that this holds generally in our empirical situation and is not dependent on the polynomial approach. In panel (a) we can see that using concave demand and cost curves shifts the efficient market solution outward, leading not only to larger dead-weight losses, but also to a significantly stronger relative contraction of the market due to adverse selection. In contrast, panel (b) shows that using convex demand and average cost curves can lead to marginal cost curves which are even increasing in the quantity sold. This does not remove adverse selection (which is due to the marginal costs being below the average cost), but it does lessen the potential consequences.

5 Discussion of Equity

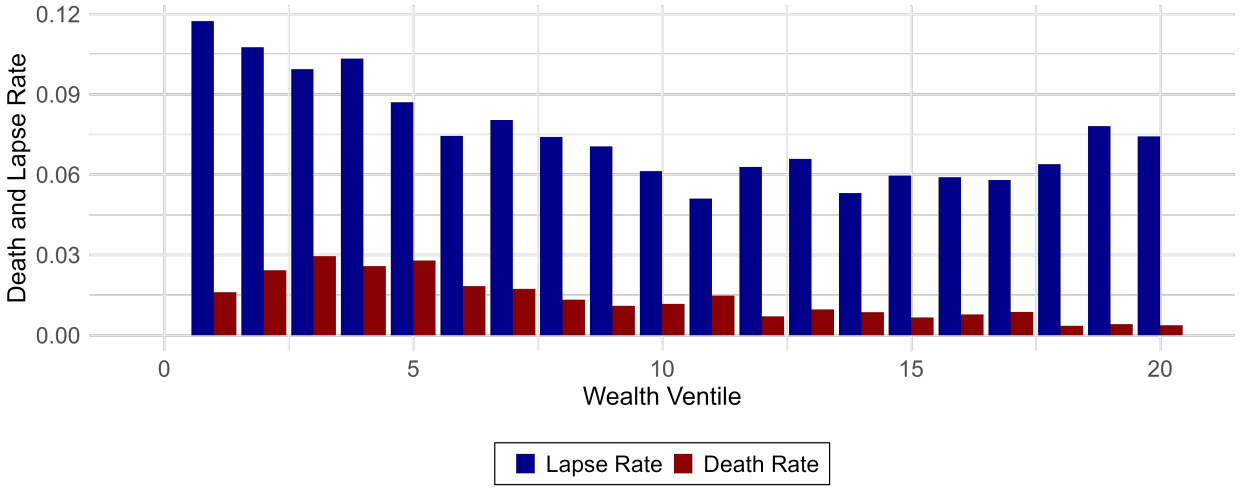
Akerlovian models of adverse selection such as the equilibrium by Einav et al. (2010) have a hard time speaking to the distributional and equity consequences of the market outcome. Our reduced form estimates show a dead-weight loss in total welfare due to lapsation-based adverse selection. This result suggests that measures to make prices contingent on lapse rate or to make prices equal costs for each contract year will increase the size of the market and the overall welfare. As with all results on the population level, however, there is heterogeneity in the individual welfare consequences such measures could have (Jansen et al., 2022; Geruso et al., 2023; Koijen et al., 2024). As pointed out by Rothschild (2024), this is especially the case in Akerlovian markets of adverse selection. In such markets, ex-post welfare gains cannot be redistributed heterogeneously across the population because, by definition, the relevant dimension of heterogeneity is unknown to the central planner.

Our empirical results show that while there is an adverse selection channel through lapse rates, there is no mortality channel. Any change in the market targeted at increasing efficiency would thus have to use the lapsation channel. One such change would be insurers pricing life insurance contracts contingent on expected individual lapse rates. This is effectively the same as increasing the amount of information used for pricing in the sense of Jansen et al. (2022). The distributional consequences thus depend on the correlation of wealth with the take-up of life insurance coverage as well as on the correlation of wealth with the tendency to lapse. Previous empirical studies, such as Gropper and Kuhnen (2025) as well as our own analysis of the Survey of Consumer Finances in Figure 2, provide a clear indication that the former correlation is positive. To analyze the latter, we plot lapse rates and death rates among insureds in the past six waves of the Health and Retirement Survey (Health and Retirement Study, 2020; RAND Corporation, 2020) by wealth ventile in Figure 10. We can see that the low wealth ventiles are significantly more likely to lapse, while there seems to be no discernible relationship for higher wealth levels.[22]

To model the equity impacts of selection through lapsation, we employ a Berry et al. (1995,

[22]See Appendix C for details on why we use the Health and Retirement Survey rather than the Survey of Consumer Finances and the variable construction.

Figure 10: Empirical correlation of wealth and lapse rates



Note: The figure reports the lapse rates of respondents aged 64 and younger in the Health and Retirement survey for years 2010 through 2020. Only respondents which reported to have life insurance coverage in the last wave of the survey are considered. Wealth is taken as the sum of all assets minus the sum of all debt and the ventiles are constructed according to the prior period's wealth. See Appendix C for details on the variable construction.

BLP) style random coefficients discrete choice demand model with micro-moments. We derive the micro-moments on the association of household wealth and lapse, death and purchase rates from the Health and Retirement Survey (HRS).[23] After estimating the base model, we run a counterfactual simulation of individually-underwritten lapse pricing and estimate the distributional impact of the resulting change in consumer welfare. This lets us understand both the equity consequences of the selection implied by our results and the consequences of the policy remedies we discussed.

5.1 Utility Specification

We index the S state-level markets by $s = 1, \ldots, S$, with J insurance firms, indexed by $j = 1, \ldots, J$, and I consumers, indexed by $i = 1, \ldots, I$. In each period t, consumers who do not currently have a policy choose one product from the set of available firms or an outside option (denoted by

[23]We merge surrender and lapse together because, though the HRS does report if consumers who have lapsed received a cash value, many of the observations are missing and the reported rates are an order of magnitude lower than the surrender rates in the NAIC data.

0).[24] Instead of modeling a dynamic utility function, we define an indirect, instantaneous utility function for each the individual-firm-market combination in period t. The utility that consumer i derives from purchasing from firm j in state s in period t is specified as:

$$u_{ijst} = -\beta_{p,i} \times P_{jt} + \beta_{w,i} + \xi_{jt} + \eta_{st} + \phi_{js} + \varepsilon_{ijst} \tag{31}$$

where P_{jt} is the price offered by firm j at time t; $\beta_{p,i}$ is the individual-specific price sensitivity; ξ_{jt} captures unobserved (to the econometrician) firm-year characteristics that influence demand; η_{st} and ϕ_{js} are state-year and firm-state fixed effects, respectively; and ε_{ijst} is an i.i.d. Type I extreme value error term. $\beta_{w,i}$ represents the impact of consumer i's wealth on the demand for life insurance and can be interpreted as the aggregate effect of "direct" selection (i.e. the consumer's knowledge of their personal lapse rates) and the generally positive correlation of wealth and demand.

The individual-specific price and wealth coefficients are:

$$\beta_{p,i} = exp(\mu_p + \nu_i) \tag{32}$$

$$\beta_{w,i} = \sum_{z \in \mathbb{Q}} \beta_{w,z} \times \mathbb{1}(\mathcal{Q}(w_i) \in z) \tag{33}$$

where μ_p is the base price sensitivity, and ν_i represents the unobserved heterogeneity in the consumer's price elasticity and is assumed to be distributed with mean zero and standard deviation σ_p. We model the impact of wealth as a series of intercept terms, where $\mathbb{Q}$ is a collection of equally sized sets that partition the unit interval, $\mathcal{Q}(w_i)$ is the consumer's wealth quantile, and $\mathbb{1}()$ is an indicator function.

5.2 Estimation

We estimate the model in four stages. First, we follow the "FRAC" Taylor approximation method developed by Salanié and Wolak (2019) to obtain initial parameter estimates for μ_p and σ_p. Second,

[24] We define the total potential market size as the dollars of personal income minus the face value of individual insurance in force.

we use these estimates as starting values for the traditional BLP GMM estimation. Third, we use the GMM results to develop the "optimal" instruments following the procedure of Conlon and Gortmaker (2020). Finally, we adapt the model to match the wealth micro-moments. The four steps are described in more detail in Appendix F.

5.3 Counterfactual

Our counterfactual considers lapse-based pricing, where premiums are adjusted based on the lapse risk of each wealth group. This approach captures how insurers could price discriminate based on observable correlates of lapsation risk, addressing the core equity tradeoff in selection. The counterfactual pricing formula adjusts premiums based on the ratio of group-specific to market-average lapse rates.

To derive the pricing formula, consider baseline quantities of the pooled price P_b and lapse rate L_b. For individual of type m, let the lapse rate be L_m. Define the cost-price elasticity as $\delta = \frac{\partial \ln AC}{\partial \ln P}$ and the lapse-price elasticity as $\lambda = \frac{\partial \ln L}{\partial \ln P}$. Both parameters are known from the reduced form estimations reported in Section 4. Because cost responds to price only through lapses, we can write

$$\frac{\partial \ln AC}{\partial \ln L} = \frac{\delta}{\lambda}. \tag{34}$$

Denote the baseline price and lapse rate for an insurer-year observation as P_{bjt} and L_{bjt}, respectively. They represent demand-weighted average quantities across all wealth quantiles. As for the reduced-form market model, we assume a multiplicative profit factor ζ_{jt} such that $P_{bjt} = \zeta_{jt} AC_{bjt}$. From this assumption on profits and due to constant elasticities, we can write

$$AC(P_{mjt}) = \frac{1}{\zeta_{jt}} P_{bjt} \left(\frac{L(P_{mjt})}{L_{bjt}} \right)^{\delta/\lambda}, \tag{35}$$

$$L(P_{mjt}) = L_{mjt} \left(\frac{P_{mjt}}{P_{bjt}} \right)^{\lambda}. \tag{36}$$

We further assume that the relative profits are equal for all wealth quantiles (and thus all customers) such that $P_{mjt} = \zeta_{jt} AC_{mjt}$. This renders

$$\frac{P_{mjt}}{P_{bjt}} = \left(\frac{L_{mjt}}{L_{bjt}}\right)^{y}, \qquad y = \frac{\delta}{\lambda(1-\delta)}, \qquad \delta \neq 1. \tag{37}$$

Individuals' lapse probabilities in our sample only differ by the wealth ventile that they are in such that $L_{ijt} = \frac{L_{\mathcal{Z}(w_i)}}{L_b} L_{bjt}$. Here, $L_{\mathcal{Z}(w_i)}$ is the ventile specific lapse probability for individual i, $\mathcal{Z}(w_i) = z_i$ is the associated ventile function, and L_b is the average lapse probability in the entire sample. Our counterfactual thus renders an insurer-year specific price for each wealth ventile ω given by

$$P_{jt}(z) = P_{jt}^{\text{obs}} \times \left(\frac{L_z}{L_b}\right)^{y} \tag{38}$$

where P_{jt}^{obs} is the observed equilibrium price for firm j in market t and $y = \frac{\delta}{\lambda(1-\delta)}$ is given by the reduced-form estimates. The exponent y captures the equilibrium price adjustment needed to maintain relative profits when pricing is based on lapse risk, reflecting the effects of both changes in demand and cost induced by prices.

The lapse-based pricing applies a uniform multiplier to all firms' prices within each wealth quantile, preserving relative price differences while adjusting absolute price levels based on lapse risk. Given the estimated mean utility parameters ϑ_{jt}, the base price sensitivity μ_p, and wealth ventile parameters $\beta_{w,z}$, we simulate consumers at different wealth quantiles and integrate over heterogeneous price sensitivities. Under observed prices, the choice probabilities are:

$$\widehat{\varphi}_{jt}^{\text{orig}} = \int \frac{\exp\Big(\vartheta_{jt} + (\mu_p + \nu_i)\, P_{jt} + \beta_{w,z_i}\Big)}{1 + \sum_h \exp\Big(\vartheta_{ht} + (\mu_p + \nu_i)\, P_{ht} + \beta_{w,z_i}\Big)} dF_\nu, \tag{39}$$

where $\nu_i \sim N(0, \sigma_p^2)$ captures unobserved heterogeneity in price sensitivity. Under counterfactual

lapse-based prices:

$$\widehat{\varphi}_{jt}^{\text{cf}} = \int \frac{\exp\Big(\vartheta_{jt} + (\mu_p + \nu_i)\, P_{jt}(z_i) + \beta_{w,z_i}\Big)}{1 + \sum_h \exp\Big(\vartheta_{ht} + (\mu_p + \nu_i)\, P_{ht}(z_i) + \beta_{w,z_i}\Big)} dF_\nu. \tag{40}$$

Consumer surplus is computed by integrating the log-sum formula over the distribution of price sensitivities:

$$CS(z) = \frac{1}{20} \int \frac{1}{|\mu_p + \nu_i|} \log \left(1 + \sum_j \exp\Big(\vartheta_{jt} + (\mu_p + \nu_i)\, P_{jt}(z) + \beta_{w,z}\Big) \right) dF_\nu. \tag{41}$$

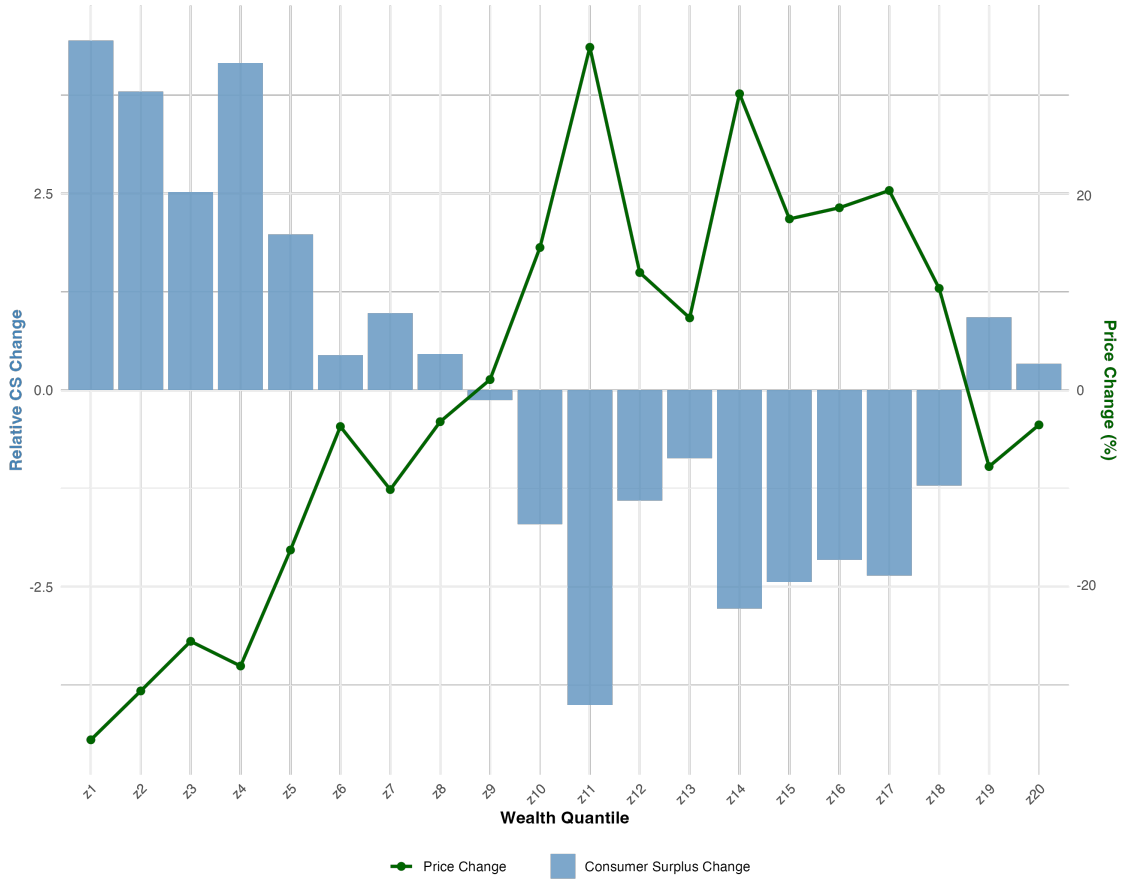
We implement this integration numerically using Gauss-Hermite quadrature. Finally, we aggregate predicted consumer choices across wealth quantiles—weighted by market sizes—and summarize changes in consumer surplus, prices paid, and market participation rates. The key equity implications emerge from how lapse-based pricing affects different wealth groups: lower-wealth consumers benefit from price decreases due to their higher lapse rates, while higher-wealth consumers face price increases.

Figure 11 illustrates the relative distributional impact of implementing individual lapse-based pricing in the life insurance market. The findings reveal a largely progressive redistribution of consumer surplus across wealth groups, with significant gains for lower-wealth households and corresponding losses for middle- and upper-middle-wealth households.

The most prominent result is the substantial increase in consumer surplus for households in the lowest wealth quantiles. The bottom five wealth ventiles (z1–z5) capture the majority of positive consumer surplus changes, with z1 experiencing the largest relative gain. These gains correspond to price reductions of approximately 2–4% for the lowest wealth groups, as shown by the green line. This progressive redistribution arises directly from correcting the cross-subsidization inherent in uniform pricing: lower-wealth households, typically characterized by higher lapse rates, currently pay premiums that exceed their actuarially fair cost.

The distributional pattern reverses for middle-wealth quantiles. Starting around z9 and continuing through z18, households experience negative changes in consumer surplus, as these groups

Figure 11: Wealth-Based Distributional Impacts of Risk-Based Pricing



Note: This figure presents the distributional impacts of lapse-based pricing across wealth quantiles. The blue bars (left axis) show each wealth quantile's share of the total consumer surplus change, with positive values indicating gains and negative values indicating losses. The green line (right axis) displays the percentage change in prices faced by each wealth group. Wealth quantiles range from z1 (lowest wealth) to z20 (highest wealth). The analysis incorporates both direct price effects and behavioral responses, including changes in market participation. Under lapse-based pricing, premiums are adjusted according to each group's predicted lapse rate relative to the market average, with the adjustment factor $(L_z/L_b)^{\delta/(\lambda(1-\delta))}$ where λ and δ are the estimated lapse-price and cost-price elasticities and z denotes the ventile.

effectively lose the implicit subsidy they receive under uniform pricing. The price changes for these groups range from small increases for z9 to more substantial increases of 1–4% for z10–z18. These households, with lower lapse rates than the population average, would pay premiums closer to their true expected costs under lapse-based pricing.

These empirical results support our theoretical framework regarding adverse selection driven by policy lapsation. Under the existing flat-premium structure, a regressive subsidy exists where

lower-wealth households, characterized by higher lapse probabilities, implicitly subsidize wealthier households, who tend to retain policies for longer periods. Introducing individual lapse-based pricing would correct this distortion, making life insurance more affordable and accessible for financially vulnerable households. Because it removes the implicit subsidization of low-lapse-rate consumers by high-lapse-rate consumers, this counterfactual captures the welfare consequences of all three proposed policy solution. Whether we consider lapse-based pricing, the introduction of policy reserves, or the removal of flat premium schedules, all three solutions have the effect that they remove this cross-subsidization, leading to the welfare effects observed in Figure 11.

6 Conclusions

We examine lapsing behavior as a channel for selection in life insurance markets. Using insurance regulatory data, we find strong evidence of adverse selection. Further, we find that this selection is driven solely by lapse behavior, with mortality having no influence. We find that a 1% increase in premiums leads to a 0.15% decrease in lapse rates, while having no effect on mortality. This decrease in lapse rates leads to a 0.13% increase in future costs. Adverse selection reduces market size by 14% and creates an annual net present value for the dead-weight losses of $1.3 billion for each new generation of policyholders.

Our findings challenge the conventional wisdom that life insurance markets exhibit little (or potentially advantageous) selection. By demonstrating that lapse behavior is a primary, and previously ignored, selection channel, we help to explain conflicting results in the literature. We also show that sophisticated consumer knowledge about their own risk type is not necessary for this adverse selection to occur and detail that, unlike other selection markets, the distributional effects are also negative.

Our analysis does have limitations. The reduced-form approach likely understates welfare costs by assuming price changes affect all contracts rather than just new ones. Additionally, our lack of individual contract data limits our ability to model to other contract features that would more accurately allow us to estimate demand.

Our theoretical contribution extends beyond life insurance markets. By showing how selection can operate through multiple channels with opposing effects, we provide a framework for analyzing complex selection patterns in other insurance markets where contracts span multiple periods and early termination is possible, such as long-term care insurance or annuities. We suggest three potential policy solutions: incorporating individual-based lapse estimates into pricing, individual policy reserves, or shifting from flat premiums to rates that reflect contemporaneous mortality risk. These changes would benefit low-income consumers by reducing cross-subsidization of wealthy policyholders who maintain long-term coverage. They would also potentially reshape the secondary life insurance market. Lapse-based pricing in particular would create new arbitrage opportunities in life settlements and stranger-originated life insurance schemes. Non-flat premiums that contemporaneously price mortality risk would have fewer negative consequences for secondary markets, but are unpopular with customers and thus the insurance industry. Our favored solution are individual policy reserves. Here, any difference between premium and actuarial mortality risk is paid into a reserve. Once mortality risk exceed premiums, the reserve is used to fund the difference. Any amount in the reserve is paid out to the customer upon lapse or surrender. This way, there is no cost effect and thus no selection through the lapse channel while premiums are kept level and no new arbitrage opportunities are created on secondary markets.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? The Quarterly Journal of Economics 138(1), 1–35.

Ai, M. and P. Steinorth (2024, March). Revisiting selection in life insurance. Unpublished manuscript. German Socio-Economic Panel (SOEP).

Aizawa, N. and Y. S. Kim (2018). Advertising and risk selection in health insurance markets. American Economic Review 108(3), 828–867.

Ameriks, J., A. Caplin, S. Laufer, and S. Van Nieuwerburgh (2011). The joy of giving or assisted living? using strategic surveys to separate public care aversion from bequest motives. The Journal of Finance 66(2), 519–561.

Azar, J., M. C. Schmalz, and I. Tecu (2018). Anticompetitive effects of common ownership. Journal of Finance 73(4), 1513–1565.

Azevedo, E. M. and D. Gottlieb (2017). Perfect competition in markets with adverse selection. Econometrica 85(1), 67–105.

Bernheim, B. D. (1991). How strong are bequest motives? evidence based on estimates of the demand for life insurance and annuities. Journal of Political Economy 99(5), 899–927.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, 841–890.

Bissonnette, L., M. D. Hurd, and P.-C. Michaud (2017). Individual survival curves comparing subjective and observed mortality risks. Health Economics 26(12), e285–e303.

Brown, J. R. and A. Goolsbee (2002). Does the internet make markets more competitive? evidence from the life insurance industry. Journal of Political Economy 110(3), 481–507.

Carson, J. M., C. M. Ellis, R. E. Hoyt, and K. Ostaszewski (2020). Sunk costs and screening: two-part tariffs in life insurance. Journal of Risk and Insurance 87(3), 689–718.

Carson, J. M. and M. D. Forster (2000). Suitability and life insurance policy replacement: An analytical tool. Journal of Insurance Regulation 18(4), 427–447.

Cawley, J. and T. Philipson (1999). An empirical examination of information barriers to trade in insurance. The American Economic Review 89(4), 827–846.

Chen, R., K. A. Wong, and H. C. Lee (2001). Age, period, and cohort effects on life insurance purchases in the us. Journal of Risk and Insurance, 303–327.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016). The association between income and life expectancy in the United States, 2001-2014. Jama 315(16), 1750–1766.

Collier, B. L. and M. A. Ragin (2020). The influence of sellers on contract choice: Evidence from flood insurance. Journal of Risk and Insurance 87(2), 523–557.

Conlon, C. and J. Gortmaker (2020). Best practices for differentiated products demand estimation with pyblp. The RAND Journal of Economics 51(4), 1108–1161.

Cutler, D. M., A. Finkelstein, and K. McGarry (2008). Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. The American Economic Review 98(2), 157–162.

Daily, G., I. Hendel, and A. Lizzeri (2008). Does the secondary life insurance market threaten dynamic insurance? American Economic Review 98(2), 151–156.

DeFusco, A. A., H. Tang, and C. Yannelis (2022). Measuring the welfare cost of asymmetric information in consumer credit markets. Journal of Financial Economics 146(3), 821–840.

Einav, L. and A. Finkelstein (2011). Selection in insurance markets: Theory and empirics in pictures. The Journal of Economic Perspectives 25(1), 115–38.

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating welfare in insurance markets using variation

in prices. The Quarterly Journal of Economics 125(3), 877.

Eling, M. and M. Kochanski (2013). Research on lapse in life insurance: What has been done and what needs to be done? The Journal of Risk Finance 14(4), 392–413.

Fang, H., M. P. Keane, and D. Silverman (2008). Sources of advantageous selection: Evidence from the medigap insurance market. Journal of Political Economy 116(2), 303–350.

Fang, H. and E. Kung (2010). How does life settlement affect the primary life insurance market? NBER Working Paper Series.

Fang, H. and E. Kung (2021). Why do life insurance policyholders lapse? the roles of income, health, and bequest motive shocks. Journal of Risk and Insurance 88(4), 937–970.

Fei, W., C. Fluet, and H. Schlesinger (2015). Uncertain bequest needs and long-term insurance contracts. Journal of Risk and Insurance 82(1), 125–148.

Fier, S. G. and A. P. Liebenberg (2013). Life insurance lapse behavior. North American Actuarial Journal 17(2), 153–167.

Gao, J., S. Ge, L. Schmidt, and C. J. T. Trillo (2023). How do health insurance costs affect firm labor composition and technology investment? Working Paper.

Ge, S. (2022). How do financial constraints affect product pricing? evidence from weather and life insurance premiums. The Journal of Finance 77(1), 449–503.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2014). Competition and ideological diversity: Historical evidence from us newspapers. American Economic Review 104(10), 3073–3114.

Geruso, M., T. Layton, and A. Leive (2023). The incidence of adverse selection: Theory and evidence from health insurance choices. Technical report, National Bureau of Economic Research.

Geruso, M., T. Layton, and D. Prinz (2019). Screening in contract design: Evidence from the aca health insurance exchanges. American Economic Journal: Economic Policy 11(2), 64–107.

Giesbert, L., S. Steiner, and M. Bendig (2011). Participation in micro life insurance and the use of other financial services in ghana. Journal of Risk and Insurance 78(1), 7–35.

Gottlieb, D. and K. Smetters (2021). Lapse-based insurance. American Economic Review 111(8), 2377–2416.

Gropper, M. J. and C. M. Kuhnen (2025). Wealth and insurance choices: Evidence from us households. The Journal of Finance 80(2), 1127–1170.

Hackmann, M. B., J. T. Kolstad, and A. E. Kowalski (2015). Adverse selection and an individual mandate: When theory meets practice. American Economic Review 105(3), 1030–1066.

Harris, T. and A. Yelowitz (2014). Is there adverse selection in the life insurance market? evidence from a representative sample of purchasers. Economics Letters 124(3), 520–522.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In The Economics of New Goods. University of Chicago Press.

He, D. (2009). The life insurance market: Asymmetric information revisited. Journal of Public Economics 93(9-10), 1090–1097.

He, D. (2011). Is there dynamic adverse selection in the life insurance market? Economics Letters 112(1), 113–115.

Health and Retirement Study (2020). Health and retirement study (hrs) public use datasets, 2008–2020 waves. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (U01AG009740) and the Social Security Administration.

Hedengren, D. and T. Stratmann (2016). Is there adverse selection in life insurance markets? Economic inquiry 54(1), 450–463.

Holzhausen, A., M. Grimm, K. Stoffel, P. Pelayo-Romero, and P. Hoffmann (2025, May). Allianz Global Insurance Report 2025: Rising demand for protection. Research report, Allianz Research, Munich, Germany. Accessed: January 7, 2026.

Hurd, M. D. and K. McGarry (2002). The predictive validity of subjective probabilities of survival. The Economic Journal 112(482), 966–985.

Jansen, M., F. Nagel, C. Yannelis, and A. L. Zhang (2022). Data and welfare in credit markets. Technical Report 2022-88, University of Chicago, Becker Friedman Institute for Economics.

Jayachandran, S. and A. Lleras-Muney (2009, 02). Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines*. The Quarterly Journal of Economics 124(1), 349–397.

Kang, Z. Y. and S. Vasserman (2025). Robust bounds for welfare analysis. American Economic Review. forthcoming.

Kinge, J. M., J. H. Modalsli, S. Øverland, H. K. Gjessing, M. C. Tollånes, A. K. Knudsen, V. Skirbekk, B. H. Strand, S. E. Håberg, and S. E. Vollset (2019). Association of household income with life expectancy and cause-specific mortality in Norway, 2005-2015. Jama 321(19), 1916–1925.

Koijen, R. S., H. K. Lee, and S. Van Nieuwerburgh (2024). Aggregate lapsation risk. Journal of Financial Economics 155, 103819.

Koijen, R. S. and M. Yogo (2015). The cost of financial frictions for life insurers. American Economic Review 105(1), 445–475.

Koijen, R. S. and M. Yogo (2016). Shadow insurance. Econometrica 84(3), 1265–1287.

Kuo, W., C. Tsai, and W.-K. Chen (2003). An empirical study on the lapse rate: The cointegration approach. Journal of Risk and Insurance 70(3), 489–508.

Lai, G. C., H. Nakamura, S. Yamamoto, and T. Yoneyama (2021). Adverse retention: Strategic renewal of guaranteed renewable term life insurance policies. Journal of Risk and Insurance 88(4), 1001–1022.

Landais, C., A. Nekoei, P. Nilsson, D. Seim, and J. Spinnewijn (2021). Risk-based selection in unemployment insurance: Evidence and implications. American Economic Review 111(4), 1315–1355.

Liebenberg, A. P., J. M. Carson, and R. E. Dumm (2012). A dynamic analysis of the demand for life insurance. Journal of Risk and Insurance 79(3), 619–644.

Linton, M. (1932). Panics and cash values. Transactions of the Actuarial Society of America 33, 265–394.

Lockwood, L. M. (2018). Incidental bequests and the choice to self-insure late-life risks. American Economic Review 108(9), 2513–2550.

Martin, G. J. and A. Yurukoglu (2017). Bias in cable news: Persuasion and polarization. American Economic Review 107(9), 2565–2599.

McCarthy, D. and O. S. Mitchell (2010). International adverse selection in life insurance and annuities. Springer.

Miller, N. H. and M. C. Weinberg (2017). Understanding the price effects of the millercoors joint venture. Econometrica 85(6), 1763–1791.

Mulholland, B. S. and M. S. Finke (2014). Does cognitive ability impact life insurance policy lapsation? Available at SSRN 2512076.

National Association of Insurance Commissioners (2017). Annual statement data (life/fraternal): State page and exhibit of life insurance. Administrative statutory filings; firm- and state-level; years 1993–2017; restricted-access by license.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica 69(2), 307–342.

Nolte, S. and J. C. Schneider (2017). Don't lapse into temptation: A behavioral explanation for policy surrender. Journal of Banking & Finance 79, 12–27.

Outreville, J. F. (1990). Whole-life insurance lapse rates and the emergency fund hypothesis. Insurance: Mathematics and Economics 9(4), 249–255.

Pesando, J. E. (1974). The interest sensitivity of the flow of funds through life insurance companies: an econometric analysis. Journal of Finance 29(4), 1105–1121.

RAND Corporation (2020). Rand hrs longitudinal file 2020. Harmonized longitudinal HRS variables (waves

1992–2020). Produced by the RAND Center for the Study of Aging with funding from the National Institute on Aging and the Social Security Administration. This paper uses waves 2008–2020; life-insurance lapse variables were merged from the raw HRS Core files.

Rothschild, C. (2024). Welfare analysis in insurance markets. The Geneva Risk and Insurance Review, 1–23.

Salanié, B. and F. A. Wolak (2019). Fast," robust", and approximately correct: estimating mixed demand systems. Technical report, National Bureau of Economic Research.

Schott, F. H. (1971). Disintermediation through policy loans at life insurance companies. The Journal of Finance 26(3), 719–729.

Shapiro, B. T. (2020). Advertising in health insurance markets. Marketing Science 39(3), 587–611.

Shefrin, H. (2002). Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing. Oxford University Press on Demand.

Simonsohn, U., J. P. Simmons, and L. D. Nelson (2020). Specification curve analysis. Nature Human Behaviour 4(11), 1208–1214.

Truett, D. B. and L. J. Truett (1990). The demand for life insurance in mexico and the united states: A comparative study. Journal of Risk and Insurance, 321–328.

Wenning, D. (2024). Equal prices, unequal access: The effects of national pricing in the us life insurance industry. Working Paper.

Wuppermann, A. C. (2017). Private information in life insurance, annuity, and health insurance markets. The Scandinavian Journal of Economics 119(4), 855–881.

Appendix A Proofs of Propositions

A.1 Proof of Proposition 1

Proof. Let Ψ denote the support of $g(\psi)$. Because $\bar{\Pi}(\psi) : \Psi \mapsto \mathbb{R}$ is monotone, $\bar{\Pi}^{-1}(\Pi) : \mathbb{R} \mapsto \Psi$ is well-defined. We can thus write

$$MC(\Pi) = E[C(\psi)|\bar{\Pi}(\psi) = \Pi] = E[C(\psi)|\psi = \bar{\Pi}^{-1}(\Pi)] = C(\bar{\Pi}^{-1}(\Pi)). \tag{A1}$$

The last step follows, because monotonicity in Π ensures that there is a single value of ψ such that $\bar{\Pi}(\psi) = \Pi$. From the above it follows that $\frac{\partial MC}{\partial \Pi} = \frac{\partial C}{\partial \psi}\frac{\partial \bar{\Pi}^{-1}}{\partial \Pi}$ and thus $\frac{\partial MC}{\partial \Pi} > 0$ if and only if $sgn\left(\frac{\partial C}{\partial \psi}\right) = sgn\left(\frac{\partial \bar{\Pi}^{-1}}{\partial \Pi}\right)$. The proposition follows, because monotonicity implies that $sgn\left(\frac{\partial \bar{\Pi}^{-1}}{\partial \Pi}\right) = sgn\left(\frac{\partial \bar{\Pi}}{\partial \psi}\right)$. □

A.2 Comparative Statics when Forgetting Causes Lapses

We can see the influence of the relevant characteristics on $\bar{P}_i$ from the implicit function theorem. Specifically, we see that

$$\frac{\partial \bar{P}_i}{\partial \Upsilon_i} = \frac{u_1'(\Upsilon_i - \bar{P}_i) - u_1'(\Upsilon_i) + (1-\rho_i^l)\left[u_2'(\Upsilon_i - \bar{P}_i) - u_2'(\Upsilon_i) + (1-\rho_i^d)[u_3'(\Upsilon_i - \bar{P}_i) - u_3'(\Upsilon_i)]\right]}{u_1'(\Upsilon_i) + (1-\rho_i^l)\left[u_2'(\Upsilon_i - \bar{P}_i) + (1-\rho_i^d)u_3'(\Upsilon_i - \bar{P}_i) + \rho_i^d v_3'(\kappa + B - \bar{P}_i)\right]}, \tag{A2}$$

$$\frac{\partial \bar{P}_i}{\partial \rho_i^d} = \frac{(1-\rho_i^l)\left[u_3(\Upsilon_i) - u_3(\Upsilon_i - \bar{P}_i) + v_3(\kappa + B - \bar{P}_i) - v_3(\kappa)\right]}{u_1'(\Upsilon_i) + (1-\rho_i^l)\left[u_2'(\Upsilon_i - \bar{P}_i) + (1-\rho_i^d)u_3'(\Upsilon_i - \bar{P}_i) + \rho_i^d v_3'(\kappa + B - \bar{P}_i)\right]}, \tag{A3}$$

$$\frac{\partial \bar{P}_i}{\partial \rho_i^l} = \frac{u_2(\Upsilon_i) - u_2(\Upsilon_i - \bar{P}_i) + (1-\rho_i^d)\left[u_3(\Upsilon_i) - u_3(\Upsilon_i - \bar{P}_i)\right] + \rho_i^d\left[v_3(\kappa) - v_3(\kappa + B - \bar{P}_i)\right]}{u_1'(\Upsilon_i) + (1-\rho_i^l)\left[u_2'(\Upsilon_i - \bar{P}_i) + (1-\rho_i^d)u_3'(\Upsilon_i - \bar{P}_i) + \rho_i^d v_3'(\kappa + B - \bar{P}_i)\right]}. \tag{A4}$$

$\frac{\partial \bar{P}_i}{\partial \Upsilon_i} > 0$ follows from concavity of u. $\frac{\partial \bar{P}_i}{\partial \rho_i^d} > 0$ follows from $B > \bar{P}_i$ which has to hold by first order stochastic dominance. $\frac{\partial \bar{P}_i}{\partial \rho_i^l} < 0$ follows from rearranging $EU^{LI}(\Upsilon_i, \rho_i^l, \rho_i^d, \bar{P}_i) - EU^0(\Upsilon_i, \rho_i^d) = 0$, realizing that $u_1(\Upsilon_i - \bar{P}_i) - u_1(\Upsilon_i) < 0$ and dividing both sides by $-(1-\rho_i^l)$.

The effects on costs are staight forward as can be seen by $\frac{\partial C_i}{\partial \rho_i^d} = (1-\rho_i^l)B > 0$ and $\frac{\partial C_i}{\partial \rho_i^l} = -\rho_i^d B < 0$.

A.3 Proof of Proposition 2

Proof. Because Υ, ρ^d, and ρ^l are only driven by type, we can set $EU^{LI}(\Upsilon_i, \rho_i^l, \rho_i^d, P) = EU^{LI}(\Upsilon(\psi_i), \rho^l(\psi_i), \rho^d(\psi_i), P) = EU^{LI}(\psi_i, P)$ and similarly $EU^0(\Upsilon_i, \rho_i^d) = EU^0(\psi_i)$ and then determine the influence of type on the maximum willingness to pay for insurance as (suppressing i for legibility)

$$\frac{\partial \bar{\Pi}}{\partial \psi} = \left(\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} + \frac{\partial \bar{P}}{\partial \rho^l} \frac{\partial \rho^l}{\partial \psi} + \frac{\partial \bar{P}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi} \right) (1 - \rho^l) - \frac{\partial \rho^l}{\partial \psi} \bar{P}. \tag{A5}$$

The influence of type on costs is given by

$$\frac{\partial C}{\partial \psi} = (1 - \rho^l) B \frac{\partial \rho^d}{\partial \psi} - \rho^d B \frac{\partial \rho^l}{\partial \psi}. \tag{A6}$$

Item 2: $\frac{\partial \rho^l}{\partial \psi} = 0$ and $\frac{\partial \rho^d}{\partial \psi} < 0$ imply $\frac{\partial C}{\partial \psi} < 0$ and $\frac{\partial \bar{\Pi}}{\partial \psi} = \left(\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} + \frac{\partial \bar{P}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi} \right) (1 - \rho^l)$. Thus $sgn\left(\frac{\partial C}{\partial \psi}\right) = sgn\left(\frac{\partial \bar{\Pi}}{\partial \psi}\right)$ if $\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} < \frac{\partial \bar{P}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi}$ and $sgn\left(\frac{\partial C}{\partial \psi}\right) \neq sgn\left(\frac{\partial \bar{\Pi}}{\partial \psi}\right)$ if $\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} > \frac{\partial \bar{P}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi}$. The statement follows from Proposition 1.

Item 1: Follows from Item 2.

Item 3: $\frac{\partial \rho^d}{\partial \psi} = 0$ and $\frac{\partial \rho^l}{\partial \psi} < 0$ imply $\frac{\partial C}{\partial \psi} > 0$. Further, they imply $\frac{\partial \bar{\Pi}}{\partial \psi} = \left(\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} + \frac{\partial \bar{P}}{\partial \rho^l} \frac{\partial \rho^l}{\partial \psi} \right) (1 - \rho^l) - \frac{\partial \rho^l}{\partial \psi} \bar{P} > 0$ which follows from the assumption on type and the comparative statics regarding $\bar{P}$. The statement follows from Proposition 1. □

A.4 Proof of Proposition 3

Proof. Because consumers do not consider the influence of their own ρ^d or ρ^l when determining $\bar{P}^n$ (with n indexing naïve consumers), $\frac{\partial \bar{P}_i^n}{\partial \rho_i^l} = \frac{\partial \bar{P}_i^n}{\partial \rho_i^d} = 0$. The derivative of the maximum willingness to pay towards type is now

$$\frac{\partial \bar{\Pi}_i^n}{\partial \psi_i} = (1 - \rho_b^l) \frac{\partial \bar{P}_i^n}{\partial \Upsilon_i} \frac{\partial \Upsilon_i}{\partial \psi_i} > 0. \tag{A7}$$

The proposition follows from Proposition 1. □

A.5 Proof of Proposition 4

Proof. For the sake of legibility, we drop the subscript i from the parameters. To show that both adverse and advantageous selection can occurr in the market, we need to determine the signs of $\frac{\partial \bar{\Pi}}{\partial \psi}$ and $\frac{\partial C}{\partial \psi}$.

We begin with $\hat{\Gamma}$ and consider the influence of Υ and ρ^d, the two heterogeneity dimensions remaining after endogenizing lapses. We also consider P which is necessary later.

$$\frac{\partial \hat{\Gamma}}{\partial \Upsilon} = \frac{u_2'(\Upsilon-\hat{\Gamma}) - u_2'(\Upsilon-\hat{\Gamma}-P) + (1-\rho^d)[u_3'(\Upsilon-\hat{\Gamma}) - u_3'(\Upsilon-\hat{\Gamma}-P)]}{u_2'(\Upsilon-\hat{\Gamma}) - u_2'(\Upsilon-\hat{\Gamma}-P) + (1-\rho^d)[u_3'(\Upsilon-\hat{\Gamma}-P) - u_3'(\Upsilon-\hat{\Gamma})]} = 1 > 0 \tag{A8}$$

$$\frac{\partial \hat{\Gamma}}{\partial \rho^d} = \frac{v_3(\kappa) - v_3(\kappa+B-P) + u_3(\Upsilon-\hat{\Gamma}-P) - u_3(\Upsilon-\hat{\Gamma})}{u_2'(\Upsilon-\hat{\Gamma}) - u_2'(\Upsilon-\hat{\Gamma}-P) + (1-\rho^d)[u_3'(\Upsilon-\hat{\Gamma}) - u_3'(\Upsilon-\hat{\Gamma}-P)]} > 0 \tag{A9}$$

$$\frac{\partial \hat{\Gamma}}{\partial P} = \frac{u_2'(\Upsilon-\hat{\Gamma}-P) + \rho^d v_3'(\kappa+B-P) + (1-\rho^d)u_3'(\Upsilon-\hat{\Gamma}-P)}{u_2'(\Upsilon-\hat{\Gamma}) - u_2'(\Upsilon-\hat{\Gamma}-P) + (1-\rho^d)[u_3'(\Upsilon-\hat{\Gamma}) - u_3'(\Upsilon-\hat{\Gamma}-P)]} < 0 \tag{A10}$$

Further, it is convenient to define

$$g_L(\Upsilon,\rho^d) = E[u_2(\Upsilon-\Gamma) + \rho^d v_3(\kappa) + (1-\rho^d)u_3(\Upsilon-\Gamma) | \Gamma \geq \hat{\Gamma}] \tag{A11}$$

$$g_N(\Upsilon,\rho^d) = E[u_2(\Upsilon-\Gamma-P) + \rho^d v_3(\kappa+B-P) + (1-\rho^d)u_3(\Upsilon-\Gamma-P) | \Gamma < \hat{\Gamma}]. \tag{A12}$$

From this, consider that

$$g_L(\Upsilon,\rho^d) = \frac{\int_{\hat{\Gamma}}^{\bar{\Gamma}} [u_2(\Upsilon-\Gamma) + \rho^d v_3(\kappa) + (1-\rho^d)u_3(\Upsilon-\Gamma)]h(\Gamma)d\Gamma}{\int_{\hat{\Gamma}}^{\bar{\Gamma}} h(\Gamma)d\Gamma}. \tag{A13}$$

Thus, using Leibniz' rule, we can see that

$$\begin{aligned}\frac{\partial g_L(\Upsilon,\rho^d)}{\partial \Upsilon} =& \frac{\int_{\hat{\Gamma}}^{\bar{\Gamma}} [u_2'(\Upsilon-\Gamma) + (1-\rho^d)u_3'(\Upsilon-\Gamma)]h(\Gamma)d\Gamma - h(\hat{\Gamma})[u_2(\Upsilon-\hat{\Gamma}) + \rho^d v_3(\kappa) + (1-\rho^d)u_3(\Upsilon-\hat{\Gamma})]}{\int_{\hat{\Gamma}}^{\bar{\Gamma}} h(\Gamma)d\Gamma} \\ &+ \frac{h(\hat{\Gamma})\int_{\hat{\Gamma}}^{\bar{\Gamma}} [u_2(\Upsilon-\Gamma) + \rho^d v_3(\kappa) + (1-\rho^d)u_3(\Upsilon-\Gamma)]h(\Gamma)d\Gamma}{\left(\int_{\hat{\Gamma}}^{\bar{\Gamma}} h(\Gamma)d\Gamma\right)^2} \qquad \text{(A14)} \\ =& \frac{h(\hat{\Gamma})}{1-H(\hat{\Gamma})} g_L - \frac{h(\hat{\Gamma})}{1-H(\hat{\Gamma})}[u_2(\Upsilon-\hat{\Gamma}) + \rho^d v_3(\kappa) + (1-\rho^d)u_3(\Upsilon-\hat{\Gamma})] \\ &+ E[u_2'(\Upsilon-\Gamma) + (1-\rho^d)u_3'(\Upsilon-\Gamma) | \Gamma \geq \hat{\Gamma}] \qquad \text{(A15)}\end{aligned}$$

Similarly, we can use the same steps to obtain

$$\begin{aligned}\frac{\partial g_N(\Upsilon,\rho^d)}{\partial \Upsilon} =& \frac{h(\hat{\Gamma})}{H(\hat{\Gamma})}[u_2(\Upsilon-\hat{\Gamma}-P) + \rho^d v_3(\kappa+B-P) + (1-\rho^d)u_3(\Upsilon-\hat{\Gamma}-P)] - \frac{h(\hat{\Gamma})}{H(\hat{\Gamma})} g_N \\ &+ E[u_2'(\Upsilon-\Gamma-P) + (1-\rho^d)u_3'(\Upsilon-\Gamma-P) | \Gamma < \hat{\Gamma}] \qquad \text{(A16)}\end{aligned}$$

To observe the relevant comparative statics on the maximum willingness to pay, we start with the

influence on the maximal premium consumers are willing to pay. It is defined by $\Delta EU = EU^{LI} - EU^{0} = 0$. We begin by considering the influence of income. By the implicit function theorem $\frac{\partial \bar{P}}{\partial \Upsilon} = -\frac{\frac{\partial \Delta EU}{\partial \Upsilon}}{\frac{\partial \Delta EU}{\partial \bar{P}}}$. We start with the numerator.

Consider the influence of income on the expected utility with life insurance

$$\begin{aligned}
\frac{\partial EU^{LI}}{\partial \Upsilon} =& u_1'(\Upsilon - P) + h(\hat{\Gamma})(g_N - g_L) + (1 - H(\hat{\Gamma}))\frac{\partial g_L}{\partial \Upsilon} + H(\hat{\Gamma})\frac{\partial g_N}{\partial \Upsilon} && \text{(A17)}\\
=& u_1'(\Upsilon - P) + h(\hat{\Gamma})(g_N - g_L) + h(\hat{\Gamma})(g_L - g_N)\\
&+ h(\hat{\Gamma})[u_2(\Upsilon - \hat{\Gamma} - P) + \rho^d v_3(\kappa + B - P) + (1 - \rho^d)u_3(\Upsilon - \hat{\Gamma} - P)]\\
&- h(\hat{\Gamma})[u_2(\Upsilon - \hat{\Gamma}) + \rho^d v_3(\kappa) + (1 - \rho^d)u_3(\Upsilon - \hat{\Gamma})]\\
&+ (1 - H(\hat{\Gamma}))E[u_2'(\Upsilon - \Gamma) + (1 - \rho^d)u_3'(\Upsilon - \Gamma)|\Gamma \geq \hat{\Gamma}]\\
&+ H(\hat{\Gamma})E[u_2'(\Upsilon - \Gamma - P) + (1 - \rho^d)u_3'(\Upsilon - \Gamma - P)|\Gamma < \hat{\Gamma}] && \text{(A18)}\\
=& u_1'(\Upsilon - P) + (1 - H(\hat{\Gamma}))E[u_2'(\Upsilon - \Gamma) + (1 - \rho^d)u_3'(\Upsilon - \Gamma)|\Gamma \geq \hat{\Gamma}]\\
&+ H(\hat{\Gamma})E[u_2'(\Upsilon - \Gamma - P) + (1 - \rho^d)u_3'(\Upsilon - \Gamma - P)|\Gamma < \hat{\Gamma}] && \text{(A19)}
\end{aligned}$$

Here, the last equality follows from (7). The influence of income on the expected utility without life insurance is given by

$$\frac{\partial EU^{0}}{\partial \Upsilon} = u_1'(\Upsilon) + E[u_2'(\Upsilon - \Gamma) + (1 - \rho^d)u_3'(\Upsilon - \Gamma)]. \quad \text{(A20)}$$

Because utility is increasing and concave, we know that $u_1'(\Upsilon - P) - u_1'(\Upsilon) > 0$. Further, it follows from the same argument that $(1-H(\hat{\Gamma}))E[u_2'(\Upsilon-\Gamma)+(1-\rho^d)u_3'(\Upsilon-\Gamma)|\Gamma \geq \hat{\Gamma}]+H(\hat{\Gamma})E[u_2'(\Upsilon-\Gamma-P)+(1-\rho^d)u_3'(\Upsilon-\Gamma-P)|\Gamma < \hat{\Gamma}] - E[u_2'(\Upsilon - \Gamma) + (1 - \rho^d)u_3'(\Upsilon - \Gamma)] = H(\hat{\Gamma})E[u_2'(\Upsilon - \Gamma - P) + (1 - \rho^d)u_3'(\Upsilon - \Gamma - P)|\Gamma < \hat{\Gamma}] - H(\hat{\Gamma})E[u_2'(\Upsilon - \Gamma) + (1 - \rho^d)u_3'(\Upsilon - \Gamma)|\Gamma < \hat{\Gamma}] > 0$. We can thus write $\frac{\partial \Delta EU}{\partial \Upsilon} > 0$.

The denominator of $\frac{\partial \bar{P}}{\partial \Upsilon} = -\frac{\frac{\partial \Delta EU}{\partial \Upsilon}}{\frac{\partial \Delta EU}{\partial \bar{P}}}$ is given by $\frac{\partial EU^{LI}}{\partial \bar{P}}$, because $\frac{\partial EU^{0}}{\partial \bar{P}} = 0$. From (8), we can see that

$$\frac{\partial EU^{LI}}{\partial \bar{P}} = -u_1'(\Upsilon - \bar{P}) + h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \bar{P}}(g_N - g_L) + H(\hat{\Gamma})\frac{\partial g_N}{\partial \bar{P}}. \quad \text{(A21)}$$

The first two parts are negative by increasing utility, $g_N > g_L$ and (A10). The sign of the last par is deter-

mined by $\frac{\partial g_N}{\partial \bar{P}}$. By applying Leibniz' rule, we get

$$\frac{\partial g_N}{\partial \bar{P}} = -\frac{\int_0^{\hat{\Gamma}}(u_2'(I-\Gamma-\bar{P})+\rho^d v_3'(\kappa+B-\bar{P})+(1-\rho^d)u_3'(I-\Gamma-P))h(\Gamma)d\Gamma}{H(\hat{\Gamma})} + \frac{\partial \hat{\Gamma}}{\partial \bar{P}}\frac{h(\hat{\Gamma})[u_2(\Upsilon-\hat{\Gamma}-P)+\rho^d v_3(\kappa+B-P)+(1-\rho^d)u_3(\Upsilon-\hat{\Gamma}-P)]}{H(\hat{\Gamma})} - \frac{h(\hat{\Gamma})}{H(\hat{\Gamma})}g_N. \tag{A22}$$

This expression is clearly negative due to the properties of the utility functions and (A10). This lets us conclude that $\frac{\partial \bar{P}}{\partial \Upsilon} > 0$.

We now consider the influence of the probability of death on the maximum premium the consumer is willing to pay. By the same argument as before, we need to consider $\frac{\partial \Delta EU}{\partial \rho^d}$. By the same arguments as before

$$\frac{\partial g_L}{\partial \rho^d} = -\frac{h(\hat{\Gamma})}{1-H(\hat{\Gamma})}\frac{\partial \hat{\Gamma}}{\partial \rho^d}[u_2(I-\hat{\Gamma})+\rho^d v_3(\kappa)+(1-\rho^d)u_3(I-\hat{\Gamma})] + \frac{h(\hat{\Gamma})}{1-H(\hat{\Gamma})}\frac{\partial \hat{\Gamma}}{\partial \rho^d}g_L + E[v_3(\kappa)-u_3(I-\Gamma)|\Gamma \geq \hat{\Gamma}] \tag{A23}$$

$$\frac{\partial g_N}{\partial \rho^d} = +\frac{h(\hat{\Gamma})}{1-H(\hat{\Gamma})}\frac{\partial \hat{\Gamma}-P}{\partial \rho^d}[u_2(I-\hat{\Gamma}-P)+\rho^d v_3(\kappa+B-P)+(1-\rho^d)u_3(I-\hat{\Gamma}-P)] - \frac{h(\hat{\Gamma})}{1-H(\hat{\Gamma})}\frac{\partial \hat{\Gamma}}{\partial \rho^d}g_N + E[v_3(\kappa+B-P)-u_3(I-\Gamma-P)|\Gamma < \hat{\Gamma}]. \tag{A24}$$

Thus, we can write

$$\frac{\partial \Delta EU}{\partial \rho^d} = -h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}(g_L-g_N)+(1-H(\hat{\Gamma}))\frac{\partial g_L}{\partial \rho^d}+H(\hat{\Gamma})\frac{\partial g_N}{\partial \rho^d}-E[v_3(\kappa)-u_3(I-\Gamma)] \tag{A25}$$

$$\begin{aligned} &= -h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}(g_L-g_N)+h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}(g_L-g_N) \\ &\quad + h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}[u_2(I-\hat{\Gamma}-P)+\rho^d v_3(\kappa+B-P)+(1-\rho^d)u_3(I-\hat{\Gamma}-P) \\ &\quad - u_2(I-\hat{\Gamma})-\rho^d v_3(\kappa)-(1-\rho^d)u_3(I-\hat{\Gamma})]+(1-H(\hat{\Gamma}))E[v_3(\kappa)-u_3(I-\Gamma)|\Gamma \geq \hat{\Gamma}] \\ &\quad + H(\hat{\Gamma})E[v_3(\kappa+B-P)-u_3(I-\Gamma-P)|\Gamma < \hat{\Gamma}]-E[v_3(\kappa)-u_3(I-\Gamma)] \end{aligned} \tag{A26}$$

$$= H(\hat{\Gamma})E[v_3(\kappa+B-P)-v_3(\kappa)+u_3(I-\Gamma)-u_3(I-\Gamma-P)|\Gamma < \hat{\Gamma}] > 0 \tag{A27}$$

where the inequality follows directly from the properties of the utility functions and first-order stochastic dominance ($B > P$). This lets us conclude that $\frac{\partial \bar{P}}{\partial \rho^d} > 0$.

The maximum willingness to pay for insurance is defined as a function of type by $\bar{\Pi}(\psi_i) = \bar{P}(\psi_i)(3 - 2(1 - H(\hat{\Gamma}(\psi_i)))) = \bar{P}(\psi_i)(1 + 2H(\hat{\Gamma}(\psi_i)))$. Both $\bar{P}$ and $\hat{\Gamma}$ are influenced by consumer type ψ_i through the

income channel and the probability of death channel. Thus

$$\frac{\partial \bar{\Pi}}{\partial \psi} = \left(\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} + \frac{\partial \bar{P}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi} \right) (1 + 2H(\hat{\Gamma})) + 2\bar{P}h(\hat{\Gamma}) \left(\frac{\partial \hat{\Gamma}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} + \frac{\partial \hat{\Gamma}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi} \right) \tag{A28}$$

For the cost, $C(\psi_i) = \rho^d(\psi_i) H(\hat{\Gamma}(\psi_i)) B$. The influence of type on cost is thus given by

$$\frac{\partial C}{\partial \psi} = \frac{\partial \rho^d}{\partial \psi} H(\hat{\Gamma}) B + \rho^d B h(\hat{\Gamma}) \left(\frac{\partial \hat{\Gamma}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} + \frac{\partial \hat{\Gamma}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi} \right). \tag{A29}$$

Item 1: There are various different circumstances under which both adverse and advantageous selection appear. Below, we make a particular set of assumptions such that either adverse or advantageous selection results. These are not necessary assumptions. They are merely sufficient to show the statement in the proposition.

Adverse selection: Consider the situation in which $\frac{\partial C}{\partial \psi} > 0$. This implies

$$\rho^d h(\hat{\Gamma}) \frac{\partial \hat{\Gamma}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} > -\frac{\partial \rho^d}{\partial \psi} \left(H(\hat{\Gamma}) + \rho^d h(\hat{\Gamma}) \frac{\partial \hat{\Gamma}}{\partial \rho^d} \right) \tag{A30}$$

which further implies $\frac{\partial \hat{\Gamma}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} > -\frac{\partial \hat{\Gamma}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi}$. From this, we know that

$$\begin{aligned} \frac{\partial \Upsilon}{\partial \psi} \Big[u_2'(I - \hat{\Gamma} - P) - u_2'(I - \hat{\Gamma}) + (1 - \rho^d)[u_3'(I - \hat{\Gamma} - P) - u_3'(I - \hat{\Gamma})] \Big] \\ > -\frac{\partial \rho^d}{\partial \psi} \Big[v_3(\kappa + B - P) - v_3(\kappa) + u_3(I - \hat{\Gamma}) - u_3(I - \hat{\Gamma} - P) \Big]. \end{aligned} \tag{A31}$$

By concavity of the utility function, we can see that $u_3(I - \hat{\Gamma}) - u_3(I - \hat{\Gamma} - P) > E[u_3(I - \Gamma) - u_3(I - \Gamma - P) | \Gamma < \hat{\Gamma}]$. Thus,

$$\frac{\partial \Upsilon}{\partial \psi} \Big[u_2'(I - \hat{\Gamma} - P) - u_2'(I - \hat{\Gamma}) + (1 - \rho^d)[u_3'(I - \hat{\Gamma} - P) - u_3'(I - \hat{\Gamma})] \Big] > \frac{\partial \rho^d}{\partial \psi} \frac{\partial \Delta EU}{\partial \rho^d} \tag{A32}$$

When now assuming that $u_2''' = u_3''' = 0$, then $u_2'(I - \hat{\Gamma} - P) - u_2'(I - \hat{\Gamma}) + (1 - \rho^d)[u_3'(I - \hat{\Gamma} - P) - u_3'(I - \hat{\Gamma})] = E[u_2'(\Upsilon - \Gamma - P) - u_2'(\Upsilon - \Gamma) + (1 - \rho^d)[u_3'(\Upsilon - \Gamma - P) - u_3'(\Upsilon - \Gamma)] | \Gamma < \hat{\Gamma}]$. Because $\frac{\partial \Delta EU}{\partial \Upsilon} > \frac{\partial \Upsilon}{\partial \psi} E[u_2'(\Upsilon - \Gamma - P) - u_2'(\Upsilon - \Gamma) + (1 - \rho^d)[u_3'(\Upsilon - \Gamma - P) - u_3'(\Upsilon - \Gamma)] | \Gamma < \hat{\Gamma}]$, it follows that $\frac{\partial \Delta EU}{\partial \Upsilon} > -\frac{\partial \Delta EU}{\partial \rho^d}$ and thus $\frac{\partial \bar{P}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} > -\frac{\partial \bar{P}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi}$. Combined with $\frac{\partial \hat{\Gamma}}{\partial \Upsilon} \frac{\partial \Upsilon}{\partial \psi} > -\frac{\partial \hat{\Gamma}}{\partial \rho^d} \frac{\partial \rho^d}{\partial \psi}$, it follows that $\frac{\partial \bar{\Pi}}{\partial \psi} > 0$. Adverse selection results by Proposition 1.

Advantageous selection: We now consider $\frac{\partial C}{\partial \psi} < 0$, however, we further assume

$$-\frac{\partial \rho^d}{\partial \psi}\left(H(\hat{\Gamma}) + \rho^d h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}\right) > \rho^d h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \Upsilon}\frac{\partial \Upsilon}{\partial \psi} > -\rho^d h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}\frac{\partial \rho^d}{\partial \psi}. \tag{A33}$$

We can then follow the same steps as above to derive that under $u_2''' = u_3''' = 0$ it has to hold that $\frac{\partial \bar{\Pi}}{\partial \psi} > 0$. Advantageous selection results from Proposition 1.

Item 2: We begin by setting $\frac{\partial \Upsilon}{\partial \psi} = 0$. Then

$$\frac{\partial \bar{\Pi}}{\partial \psi} = \frac{\partial \bar{P}}{\partial \rho^d}\frac{\partial \rho^d}{\partial \psi}(1 + 2H(\hat{\Gamma})) + 2\bar{P}h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}\frac{\partial \rho^d}{\partial \psi} < 0 \text{ and} \tag{A34}$$

$$\frac{\partial C}{\partial \psi} = \frac{\partial \rho^d}{\partial \psi}H(\hat{\Gamma})B + \rho^d B h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \rho^d}\frac{\partial \rho^d}{\partial \psi} < 0. \tag{A35}$$

Alternatively, set $\frac{\partial \rho^d}{\partial \psi} = 0$. Then

$$\frac{\partial \bar{\Pi}}{\partial \psi} = \frac{\partial \bar{P}}{\partial \Upsilon}\frac{\partial \Upsilon}{\partial \psi}(1 + 2H(\hat{\Gamma})) + 2\bar{P}h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \Upsilon}\frac{\partial \Upsilon}{\partial \psi} > 0 \text{ and} \tag{A36}$$

$$\frac{\partial C}{\partial \psi} = \rho^d B h(\hat{\Gamma})\frac{\partial \hat{\Gamma}}{\partial \Upsilon}\frac{\partial \Upsilon}{\partial \psi} > 0. \tag{A37}$$

The statement follows from Proposition 1. □

A.6 Proof of Proposition 5

Proof. Subjective maximum willingness to pay is given by $\bar{\Pi}(\psi_i) = (2 + \rho_b^d)\bar{P}(\psi_i)$, resulting in

$$\frac{\partial \bar{\Pi}_i}{\partial \psi_i} = (2 + \rho_b^d)\frac{\partial \bar{P}_i}{\partial \Upsilon_i}\frac{\partial \Upsilon_i}{\partial \psi_i} > 0. \tag{A38}$$

The proposition follows from (A29) and Proposition 1. □

Appendix B Description of Back-of-the-Envelope Estimation of NPV of Deadweight Loss.

The net present value (NPV) of deadweight losses is determined over a 30-year time horizon. This calculation accounts for policy exits—such as lapses, deaths, and surrenders—and applies a 3% discount rate. The first step is to compute the average exit rate, Λ_{exit}, which is the sum of the individual exit rates:

$$\Lambda_{exit} = \Lambda_{lapse} + \Lambda_{death} + \Lambda_{surrender} \tag{A39}$$

Next, the survival probability, Ω_t, which represents the probability that a policy remains active up to period t, is given by:

$$\Omega_t = (1 - \Lambda_{exit})^{t-1}. \tag{A40}$$

The discount factor, ϱ_t, which adjusts future values to their present value, is calculated as:

$$\varrho_t = \left(\frac{1}{1.03}\right)^{t-1}. \tag{A41}$$

The combined NPV factor—representing the expected present value of \$1 received annually over 30 years—is obtained by summing the product of the survival probability and the discount factor:

$$\text{Combined Factor} = \sum_{t=1}^{30} \Omega_t \cdot \varrho_t. \tag{A42}$$

Finally, the NPV of deadweight losses is calculated by multiplying the annual deadweight loss (DWL), as determined from the structural model, by the combined factor:

$$\text{NPV_DWL} = \text{DWL} \cdot \text{Combined Factor}. \tag{A43}$$

Appendix C Description of Consumer Finance Data

C.1 Survey of Consumer Finances Data

In panel (a) of Figure 2, we consider the share of households with life insurance by income relative to the federal poverty line. Income is measured in the Survey of Consumer Finances (SCF) as variable x5729, the total income of the household before taxes and other deductions. The federal poverty line, adjusted by the number of people in the household, was taken from the Office of the Assistant Secretary for Planning and Evaluation. For 2016, the adjustment by family members was variable and we used the average value of $4,140. For this and all other figures using data from the SCF, we limit teh analysis to households whose repsondent was between 18 and 64 to limit the influence of tax considerations on life insurance demand.

In panel (b) of Figure 2, we consider the share of households with life insurance by wealth quintile. For the figure, we use the last four waves (2013, 2016, 2019, and 2022) of the SCF. Wealth is defined as the sum of all assets (both liquid and illiquid) minus the sum of all debt. We list all variables used in the construction of the different types of assets and debt in Table C1.

The general trends taken from Figure 2 are robust to alternative definitions of wealth or poverty. In Figure C1 below, we report the same diagram, but use two alternative definitions. Panel (a) uses the answer to question x7362 in the SCF for income. This question asks what the total income of the household in a "normal" year would be, thus correcting for unexpected shocks to income. Panel (b) uses only liquid assets as a definition of wealth. Both panels show the same general trends as the figure in the main text. Insurance demand is increasing in both wealth and income and decreasing over time.

C.2 Health and Retirement Survey Data

To analyze how consumers' tendencies to lapse correlate with wealth in Figure 10, we use data from the Health and Retirement Survey (HRS) rather than data from the SCF. With the exception of a special wave during the financial crisis, the SCF is not a panel and does not ask consumers whether they have lapsed a life insurance policy in some given time period. This makes an analysis of life insurance lapse decisions impossible except for doing it in a very specific and probably not representative economic environment. In the HRS, respondents are surveyed in a panel structure and asked whether they have lapsed a life insurance policy, such that the data are more suited for the analysis. The biggest downside of the HRS for our purpose is that, by design, it oversamples older individuals. We deal with this limitation, by restricting our analysis to respondents who are 64 and younger. As in the analysis of the SCF, this limits the relevance of tax

Table C1: Variables used for constructing wealth in the Survey of Consumer Finances

Wealth type	Variables in the SCF
Liquid Assets	
Checking accounts, including prepaid payment cards	x3506, x3510, x3514, x3518, x3522, x3526, x3529, x7596
IRAs and Keogh accounts	x6551, x6559, x6567, x6552, x6560, x6568, x6553, x6561, x6569, x6554, x6562, x6570, x6756, x6757, x6758
Savings and money market accounts	x3730, x3736, x3742, x3748, x3754, x3760
Mutual funds	x3822, x3824, x3826, x3828, x3830, x7787
Bonds	x3902, x7635, x7637, x7636, x7638, x7639
Stocks	x3915
Certificates of deposit	x3721
Illiquid Asset	
Real Estate	x513, x526, x604, x614, x623, x716, x1306, x1325, x1339, x1705/10000*x1706, x1805/10000*x1806, x2002, x2012
Businesses	x3125, x3224, x3129, x3229,x3335, x3408, x3412, x3452, x3416, x3420, x3428, x8452
Vehicles	x8166, x8167, x8168, x8188, x2422, x8430, x2506, x2606, x2623, x8435
Other assets	x3930, x6577, x3960, x4018, x4022, x4026, x4030
Debt	
Mortgages and real estate loans	x805, x905, x1005, x1044, x1108, x1119, x1130, x1136, x1215, x1219, x1715, x1815, x2006, x2016
Vehicle loans	x2218, x2318, x2418, x7169, x2424, x8432, x2519, x2619, x2625, x8437
Consumer debt (incl. credit cards)	x413, x421, x427, x443, x7575, x2723, x2740, x2823, x2840, x2923, x2940, x7183, x8443
Student loans	x7824, x7847, x7870, x7924, x7947, x7970, x7179, x8440
Other debt	x3932, x4032

Notes: The table lists all wealth categories and the corresponding variables in the SCF used for the construction of the wealth variable in Figure 2.

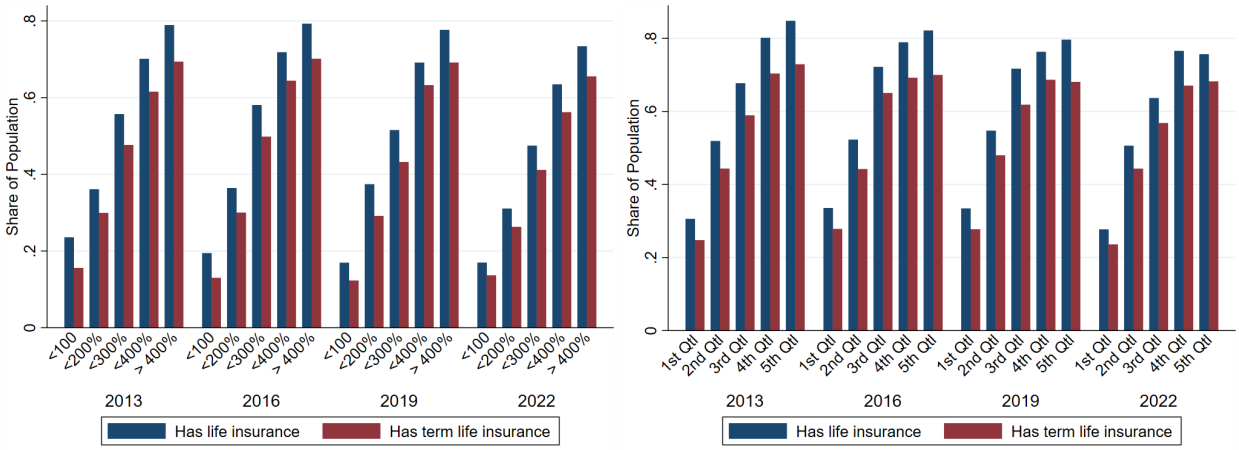
considerations in the life insurance demand and lapse decisions of the respondents.

For the construction of the sample, we utilize the last six waves of the RAND version of the HRS, which are 2010, 2012, 2014, 2016, 2018 and 2020. Because we use lagged variables, we also use data from 2008. For wealth, we use the variable "atotb", which is the primary wealth proxy provided by RAND. Ventiles are defined on this variable for all respondents with age younger or equal 64. For the display in Figure 10, we use the wealth ventile of the previous wave to sort respondents, because the lapse decision was made between the previous and the current wave. Note that data on life insurance lapses are not part

Figure C1: Life Insurance Demand by Income and Wealth - Alternative Measures

(a) Normal Income relative to the federal poverty line

(b) Quintile of liquid assets



Note: Figure repeats the analysis of panel (b) in Figure 2 using alternative definitions of wealth and poverty. The graphs show the amount of households with life insurance coverage for any type of policy (blue) or term life (red) in the US by quintile of liquid assets (panel (a)) or by income relative to the federal overty line (panel (b)). Data is taken from four waves of the Survey of Consumer Finances, see Table C1 for details.

of the RAND HRS panel and need to be merged into the panel from the raw HRS files. Because lapsing is only possible if respondents already had a life insurance policy and because life insurance ownership is influenced by wealth (see Figure 2), we restrict the sample to those individuals which reported having a life insurance policy in the previous wave.

Appendix D Robustness to Functional Form Assumptions in Akerlovian Selection Markets

In this appendix, we describe the approach we employ to assess the robustness to functional assumptions when calculating the dead-weight loss in Akerlovian selection markets. We begin by describing the problem generally and then exemplify the procedure using the original estimates of Einav et al. (2010, henceforth EFC). Note that some of the details stated below also appear in the main text, but are repeated here to allow for a stand-alone introduction of the method.

Kang and Vasserman (2025, henceforth KV) raise an important issue about the robustness of welfare inference in settings where demand functions are estimated based upon few (or two) points: when researchers can only measure a few points on the demand curve, welfare implications must be extrapolated from average treatment effects through functional form assumptions. Our setting is an especially pronounced example. For every insurance company, in every market, in every year; we infer welfare effects through counterfactual statements extrapolated from two slope measurements and a firm-market-year specific data point. Our counterfactuals are thus identified at not few, but one data point.

KV suggest a procedure for quantifying the robustness of the welfare implications of a limited estimation. In their procedure, they allow for arbitrary functional forms under the restriction that between two observed points, the local slope does not deviate from the average slope between these two points by more that a proportional factor k. The researcher then finds the highest k^* such that, within this limited set of functions, the welfare implications of their preferred functional form could flip from positive to negative (or vice-versa). Higher values of k^* then indicate more robustness, with an $k^* = 0$ implying <u>any</u> deviation of the researcher's assumed gradient between identified points could reverse the welfare implications.

Unfortunately, this approach does not apply to Akerlovian selection markets. As long as the demand curve slopes down and the cost curve slopes up, dead-weight loss (DWL) will be positive and only the magnitude will vary, regardless of k.[25] While KV do extend their functional bounds on the demand curve to bounding DWL, Akerlovian selection markets instead require two functional forms—both the demand and cost curve—to determine DWL, and these functions interact in non-trivial ways (see Equation (A46) below).

We thus propose an alternative robustness measure for functional form assumptions when estimating DWL in an EFC framework. In the same spirit as KV, we stick to the idea that different functional forms

[25] Note that we define the cost curve as a function of premiums. When plotted against quantity, both curves slope down.

lead to different welfare outcomes. However, we seek only to bound the possible outcomes, rather than find a critical value that flips the sign. To solve the issue of the complicated interactions between demand and cost curves, we restrict the functional forms of each curve to second-order polynomials and solve for the bounds numerically. While still allowing us to consider both concave and convex functions for both curves, this assumption considerably simplifies the problem's numerical complexity.

In what follows, we assume that, for a given market, we observe the current competitive equilibrium of price (P_{eq}) and quantity sold (Q_{eq}). We also assume that we observe the slopes of the demand curve ($D(P)$) and the average cost curve ($AC(P)$) at this point, and only at this point.[26] For simplicity and in congruence with the notation in EFC, we denote the slope of the demand curve as $\frac{\partial D(P)}{\partial P} = \beta$ and the slope of the average cost curve as $\frac{\partial AC(P)}{\partial P} = \delta$. Note that this deviates from the definition of β and δ in Section 4.1. We discuss how to translate to our specific setting at the end of this appendix.

To exemplify our procedure, we will use the estimated slopes and equilibrium point of EFC. Figure D2 shows the health insurance market analyzed by EFC under the linearity assumptions used in their study. We denote the (inverse) demand function, that is the function that maps quantities to prices, as $P(Q)$. While the actual estimated cost function maps prices to cost, we follow custom and show cost functions mapping quantities to cost, which we denote $AC(Q)$ and $MC(Q)$.

As long as demand is decreasing and average costs are increasing in price (and thus decreasing in quantity, given the demand assumption), Figure D2 makes clear that the relevant price range for calculating DWL is given by all positive prices below the equilibrium price. From this, the observed equilibrium and the measured slope, we have three conditions on the demand function:

(i) $\frac{\partial D(P)}{\partial P} < 0; \forall\ P \in]0, P_{\text{eq}}]$,

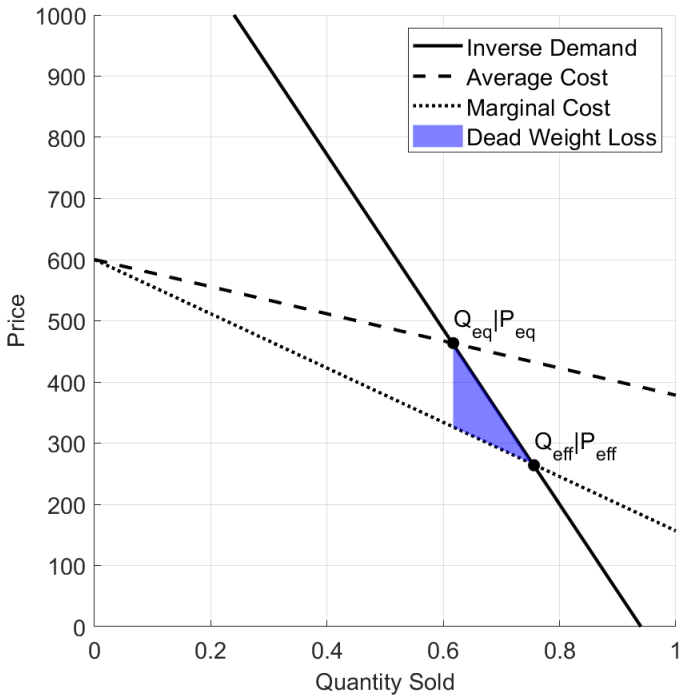
(ii) $D(P_{\text{eq}}) = Q_{\text{eq}}$ and

(iii) $\frac{\partial D(P)}{\partial P}\Big|_{P=P_{\text{eq}}} = \beta$.

We consider second-order polynomial demand functions of the form $D(P) = a_D + b_D P + c_D P^2$. The slope is given by $\frac{\partial D(P)}{\partial P} = b_D + 2c_D P$. From condition (iii), we thus know that $b_D = \beta - 2c_D P_{\text{eq}}$. Because $\beta < 0$, condition (i) will never be binding when $c_D > 0$, that is when $D(P)$ is convex. When $c_D < 0$, the most concave admissible demand curve is going to have its highest slope at $P = 0$. Condition (iii) can thus be used to find the smallest admissible value of c_D, which we will denote $\underline{c_D}$. Substituting from above, the

[26] Note that we use $D(P)$ for the demand curve rather than $Q(P)$ to avoid confusion with Q as an argument of the cost functions introduced below. This is in conflict with the use of D as the death rate in the main text but improves clarity of the appendix.

Figure D2: Market situation and dead-weight loss estimate in EFC



Note: Figure shows the inverse demand curve and average cost curve from the estimates in EFC. Marginal costs and dead-weight loss are inferred from the estimated curves. The marked points are the competitive equilibrium and the efficient market outcome.

condition implies that $\underline{c_D} = \frac{\beta}{2P_{\text{eq}}}$. From this, we can identify the values of a_D and b_D for the most concave function, using $b_D(\underline{c_D}) = \beta - 2\underline{c_D}P_{\text{eq}}$ and $a_D(\underline{c_D}) = Q_{\text{eq}} - b_D(\underline{c_D})P_{\text{eq}} - \underline{c_D}P_{\text{eq}}^2$.

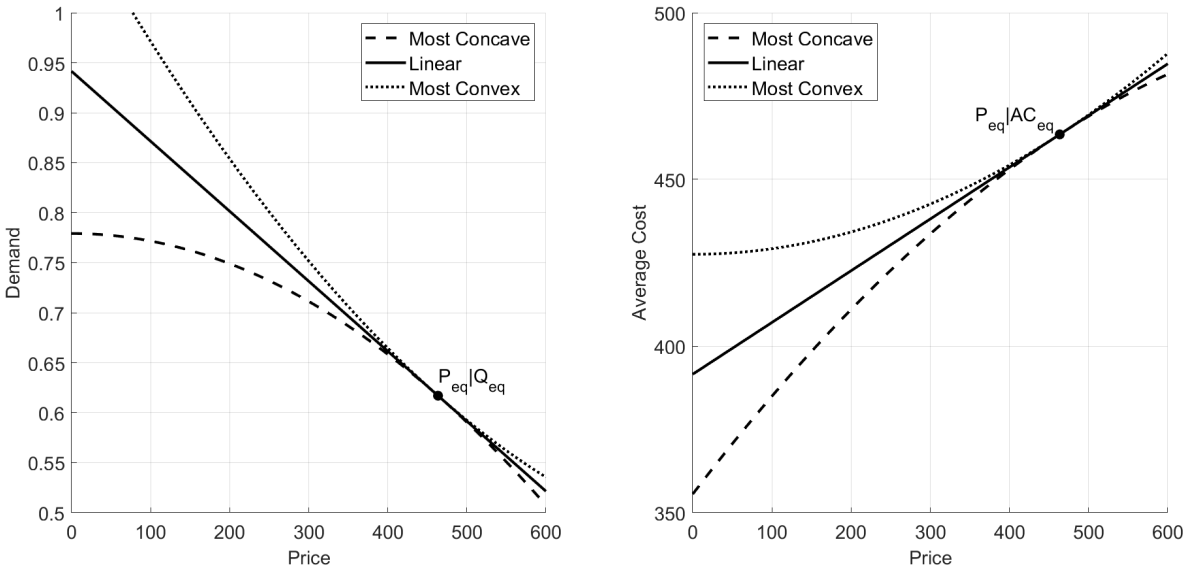
Because demand functions can be infinitely convex and still fulfill conditions (i) through (iii), we need a sensible upper bound for the range of functions to consider. Given that curvature of the polynomial is given by c_D and that the most concave function has curvature $\underline{c_D}$, we implement the upper curvature bound of $\overline{c_D} = -\underline{c_D}$. $a_D(\overline{c_D})$ and $b_D(\overline{c_D})$ are defined analogously to their most concave counterparts. These two functions then define our space of admissible demand functions and are displayed in panel (a) of Figure D3. Functions between the two extremes are constructed by a grid between $\underline{c_D}$ and $\overline{c_D}$ and the implied values for a_D and b_D are calculated for each value in the grid.

We use a similar procedure as for the demand curves to define the admissible average cost curves. All cost curves are of the form $AC(P) = a_C + b_C P + c_C P^2$. The three conditions for the functions are given by

(i) $\frac{\partial AC(P)}{\partial P} > 0; \forall\ P \in]0, P_{\text{eq}}]$,

Figure D3: Ranges of demand and cost curves used in robustness analysis

(a) Range of demand curves (b) Range of cost curves



Note: Figures show the range of possible demand curves and cost curves that are used in the robustness analysis. The procedure developed here varies the curvature of the the demand curves between the two extremes in a discrete grid.

(ii) $AC(P_{\text{eq}}) = P_{\text{eq}}$ and

(iii) $\left.\frac{\partial AC(P)}{\partial P}\right|_{P=P_{\text{eq}}} = \delta$.

Because these curves need to have a positive slope, the most convex function has to be bound by condition (iii) implying that $\overline{c_C} = \frac{\delta}{2P_{\text{eq}}}$. It then follows that $b_C(\overline{c_C}) = \delta - 2\overline{c_C}P_{\text{eq}}$ and $a_C(\overline{c_C}) = (1 - b_C(\overline{c_C}))P_{\text{eq}} - \overline{c_C}P_{\text{eq}}^2$. Having defined the average cost curve with the highest admissible curvature, we again define the opposite of the range via $\underline{c_C} = -\overline{c_C}$. Corresponding parameters and the grid of functions are defined as for the demand functions. The range of average cost curves is displayed in panel (b) of Figure D3.

EFC use the linear functional form assumption to measure the dead-weight loss as $DWL = \frac{1}{2}(Q_{\text{eff}} - Q_{\text{eq}})(P_{\text{eq}} - MC(P_{\text{eq}})$. Without the luxury of linear functional forms, we instead use

$$DWL = \int_{Q_{\text{eq}}}^{Q_{\text{eff}}} P(Q)dQ - \int_{Q_{\text{eq}}}^{Q_{\text{eff}}} MC(Q)dQ. \tag{A44}$$

Q_{eq} is known, so we require the derivation of Q_{eff}, $P(Q)$, and $MC(Q)$. $P(Q)$ is simply given by the inverse

of $D(P)$. We can use the polynomial form to derive the inverse demand curve. It is given by

$$P(Q) = \begin{cases} \frac{-b_D}{2c_D} - \frac{\sqrt{b_D^2 - 4a_D c_D + 4c_D Q}}{2c_D} & \text{for } c_D \neq 0 \\ \frac{-a_D}{b_D} + \frac{1}{b_D} Q & \text{for } c_D = 0 \end{cases} \tag{A45}$$

Because $AC(Q)$ maps quantities to cost, we can write $AC(Q) = AC(P(Q))$. Marginal cost as a function of quantities is given by

$$MC(Q) = \frac{\partial AC(Q) \cdot Q}{\partial Q} = \frac{\partial AC(P(Q))}{\partial P(Q)} \cdot \frac{\partial P(Q)}{\partial Q} \cdot Q + AC(P(Q)). \tag{A46}$$

Finally, Q_{eff} is given by the implicit definition $P(Q_{\text{eff}}) = MC(Q_{\text{eff}})$.

An important observation from Equation (A46) is that by the guidelines of our robustness analysis, namely that β, δ, P_{eq} and Q_{eq} are known, the point $MC(Q_{\text{eq}})$ will not change due to changes on the functional forms of D and AC. We know that $P(Q_{\text{eq}}) = P_{\text{eq}}$ and thus that $\frac{\partial AC(P(Q))}{\partial P(Q)}\Big|_{Q=Q_{\text{eq}}} = \frac{\partial AC(P)}{\partial P}\Big|_{P=P_{\text{eq}}} = \delta$. Further, from the chain rule and the fact that $P(Q)$ is the inverse function of $D(P)$, we know that $\frac{\partial P(Q)}{\partial Q}\Big|_{Q=Q_{\text{eq}}} = 1/\beta$. Thus

$$MC(Q_{\text{eq}}) = \frac{\delta}{\beta} Q_{\text{eq}} + AC(Q_{\text{eq}}) = \frac{\delta}{\beta} Q_{\text{eq}} + P_{\text{eq}}. \tag{A47}$$

The important implication is that a change in the magnitude of DWL due to a change in the functional form of D or AC cannot come from a shift of the marginal cost curve. Changing these function can only tilt the marginal cost curve as is shown in Figure 9 of the main text.

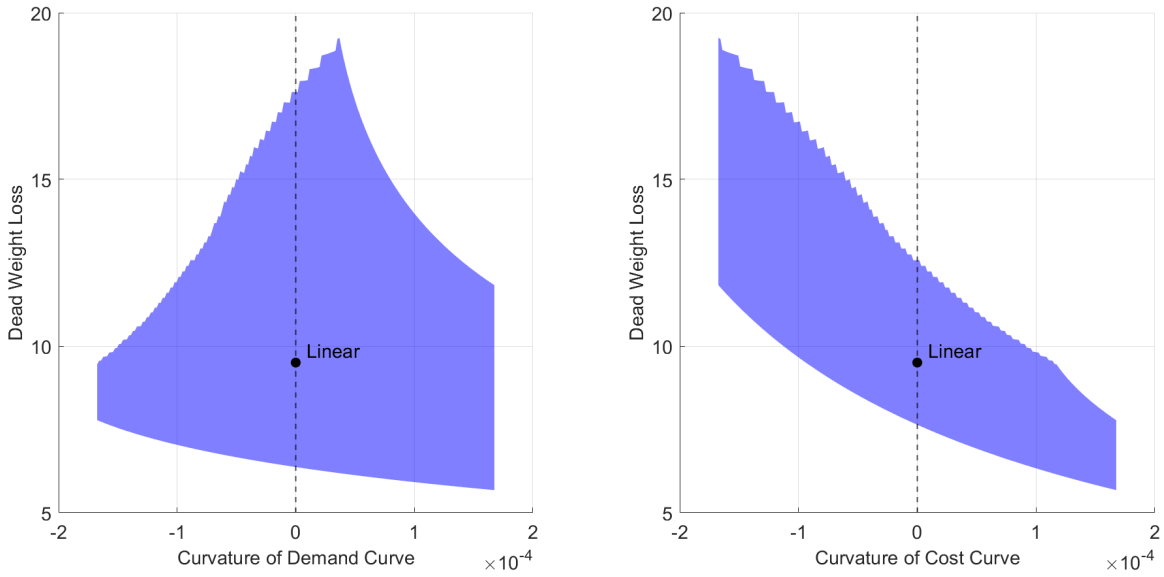
The result of the robustness analysis is shown in Figure D4. The results show that the dead-weight loss can vary widely with the functional form of the cost and demand curve. From the original value of \$9.55 per person reported in EFC, the value can increase by almost 100% or decrease by over 40% allowing for a total range between \$5.67 and \$19.24 per person. From the figure, we can see that the curvature of the cost curve seems to be more influential than that of the demand curve. In fact, the range of possible DWL values is highest when the demand curve is linear or slightly convex. For a linear demand curve, for example, changing the curvature of the cost function can lead to DWL estimates between \$6.34 and \$17.62. No range of this magnitude can be achieved by varying the curvature of the demand curve for any fixed curvature of the cost curve.

The boundary of the DWL range is somewhat ragged in certain areas of both graphs in Figure D4. This is because not all curvature combinations between demand and cost curve support an efficient solution

Figure D4: Possible range of the dead-weight loss in EFC by curvature of cost and demand function

(a) Range by curvature of the demand function

(b) Range by curvature of the cost function



Note: Analysis uses the measured slopes and market outcomes of EFC. Curvature of the function is set by allowing the function to be a second degree polynomial and varying the quadratic term within feasible ranges. The linear solution (red dot) is the dead-weight loss implied by linear cost and demand curves as is specified in EFC.

of the market in the sense that there is no intersection of $P(Q)$ and $MC(Q)$ for a positive price and a quantity smaller 1. While one could assume that the market simply clears with $Q = 1$, this assumption seems unrealistic especially when considering the low prices implied by this situation. We thus refrain from reporting a DWL for these situations and restrict the DWL range to those markets in which a solution of the form $P(Q) = MC(Q)$ exists.

Having described our procedure and exemplified it using the EFC estimates, we now consider how it can be applied to other estimates of DWL due to adverse selection. The application of the procedure to our own estimates demonstrates that the original functional forms used to estimate the slopes of the demand and cost curves are immaterial beyond their impact on the local estimates of the slope of demand and average cost in P at point $P_{\text{eq}}|Q_{\text{eq}}$. Our estimates of α_{jst}, β, γ_{jst} and δ from Equations (13) and (14) as well

as the backed out time-frame and profit-corrected value of $\tilde{\gamma}_{jst}$ imply

$$\left.\frac{\partial Q_{jst}}{\partial P}\right|_{P=P_{\text{eq}}} = e^{\alpha_{jst}}\beta P_{\text{eq},jst}^{\beta-1} \tag{A48}$$

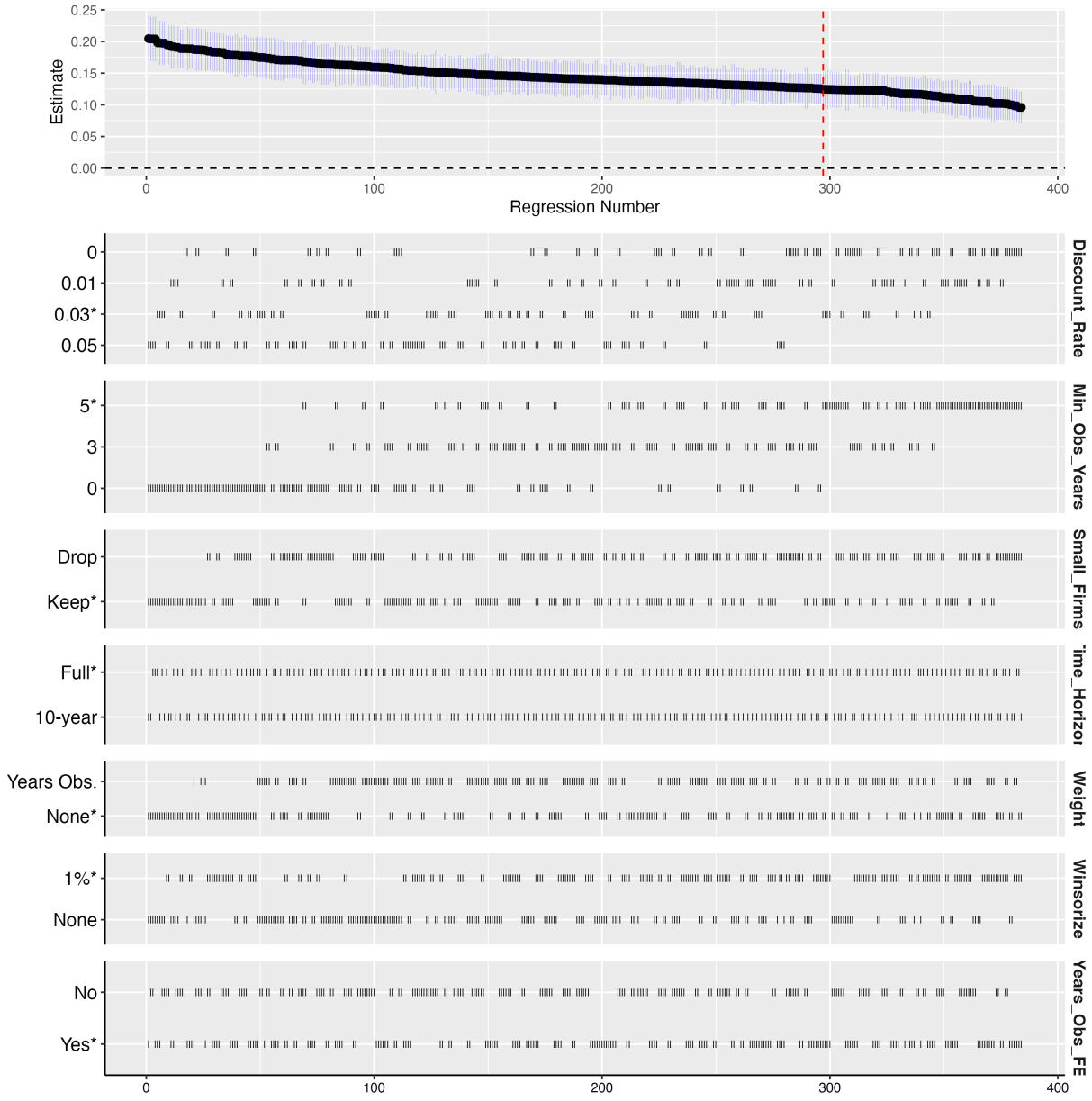
$$\left.\frac{\partial \tilde{AC}_{jst}}{\partial P}\right|_{P=P_{\text{eq}}} = e^{\tilde{\gamma}_{jst}}\delta P_{\text{eq},jst}^{\delta-1}. \tag{A49}$$

Using these slopes and the observed prices and quantities we can apply the robustness procedure to our estimates along the steps of our procedure given above. The results are shown in Figure 8 of the main text.

A bigger caveat for applications to other potential settings is whether the assumptions of an observed equilibrium point ($P_{\text{eq}}|Q_{\text{eq}}$) and slopes measured at exactly this point are actually fulfilled. While that assumption fits our particular setting very well, it can be questioned in other settings. In fact, even EFC observe the slope of the demand and cost function for a set of different prices and infer the equilibrium point from their estimate. While one can make the argument that for the majority of their data, this range is not particularly large, their estimates are certainly not made at only one point. We nevertheless use their estimates for the demonstration of the procedure due to their prominence and the simplicity of the linear estimation. In theory, more points with measured slopes allow for higher order polynomials in the cost and demand curves. We leave the exploration of this possibility for future research.

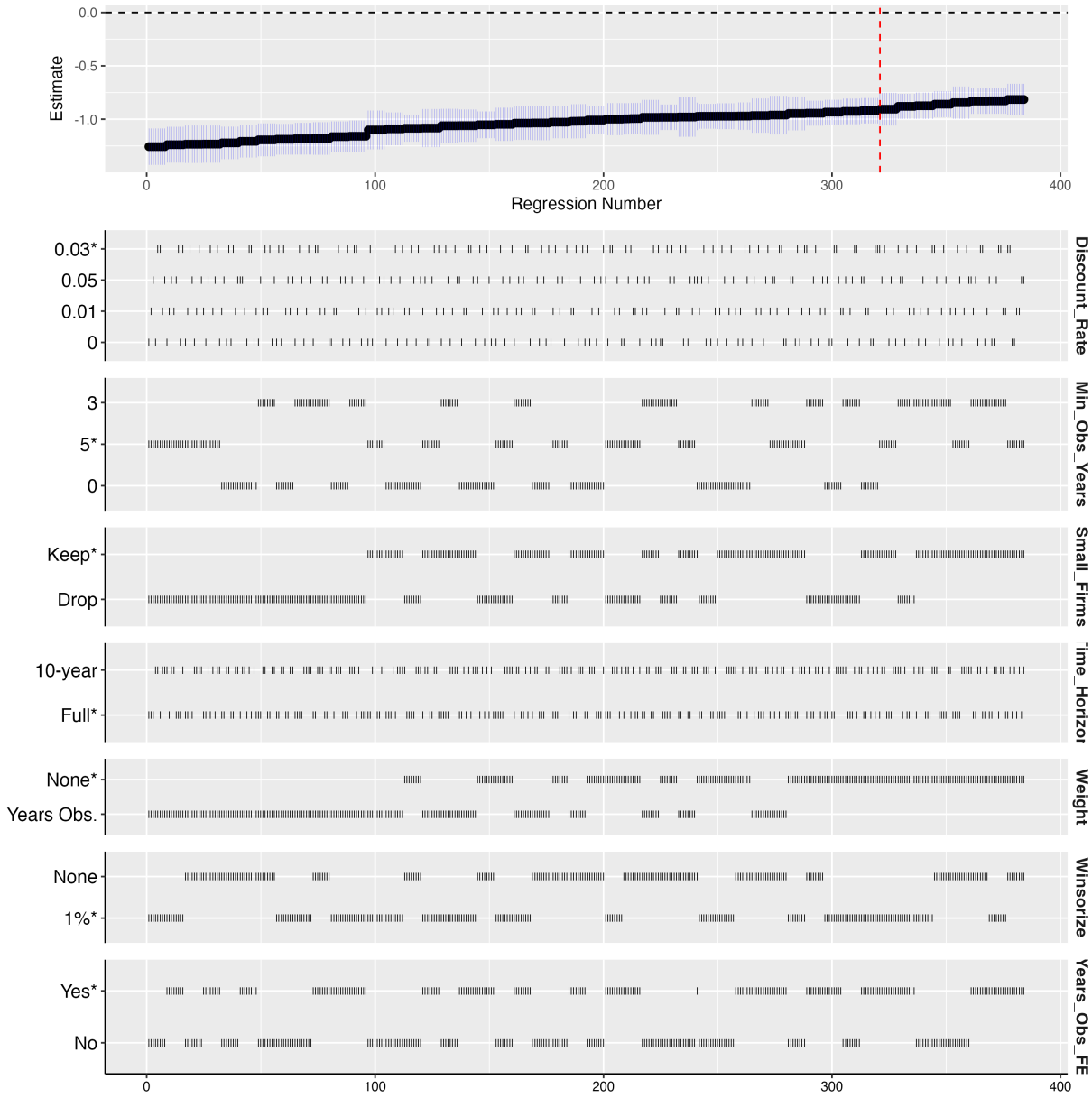
Appendix E Robustness to Research Design

Figure E5: Specification Curve: Cost Curve



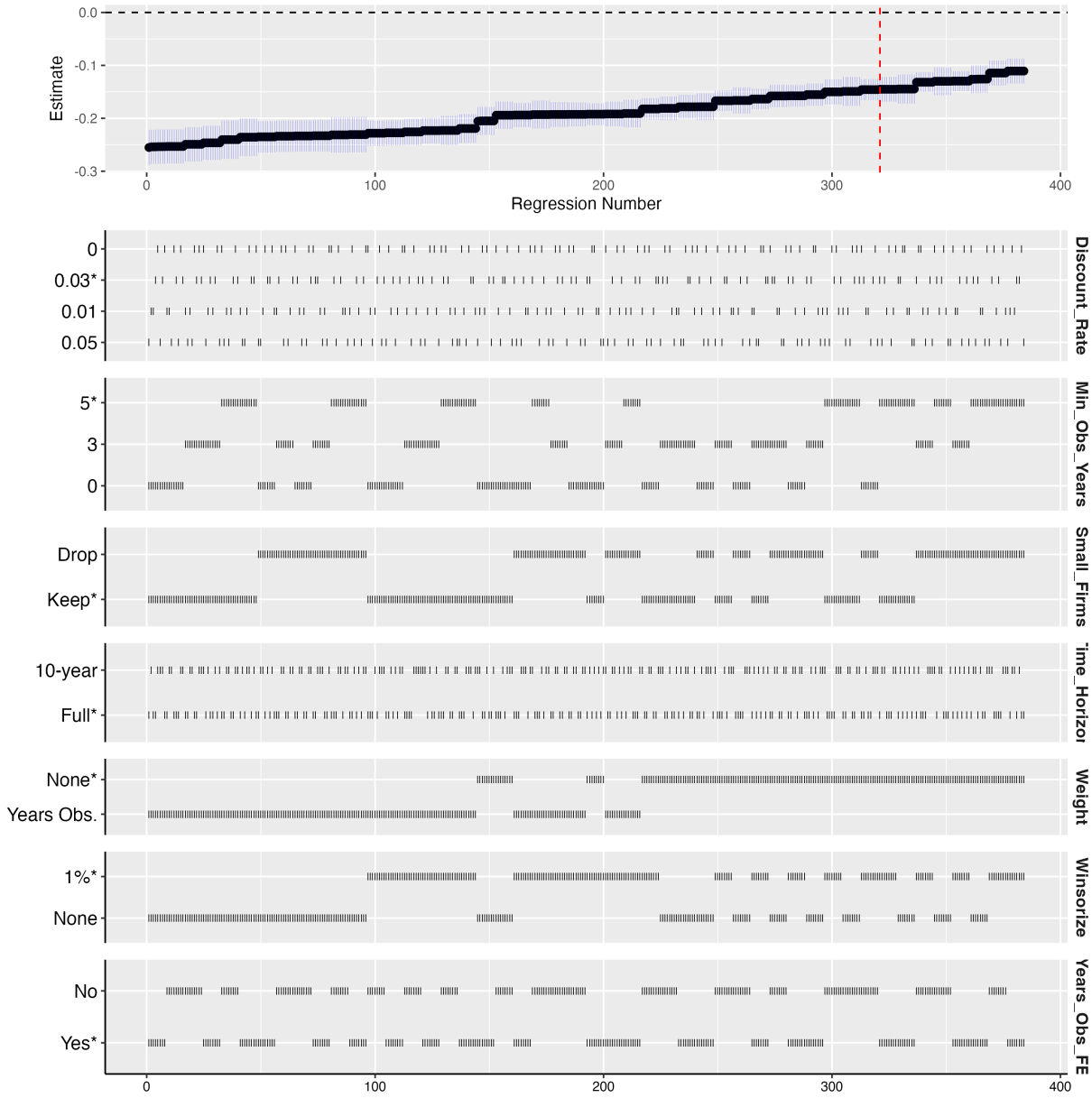
Note: This figure plots plots our estimated cost function, with 90% confidence intervals, from 384 sets of regressions featuring different combinations of modeling choices, as described in the text.

Figure E6: Specification Curve: Demand Curve



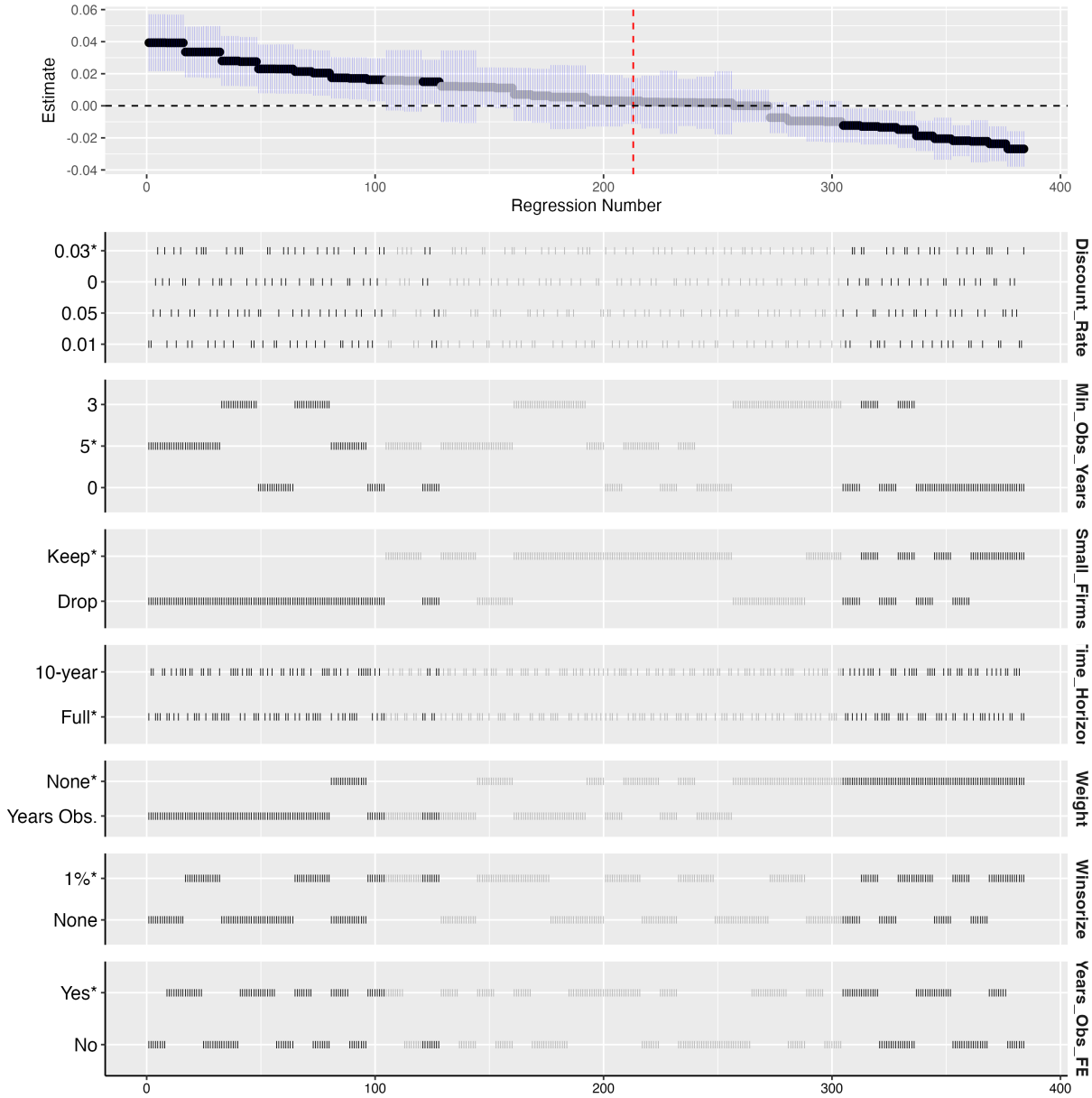
Note: This figure plots plots our estimated demand function, with 90% confidence intervals, from 384 sets of regressions featuring different combinations of modeling choices, as described in the text.

Figure E7: Specification Curve: Lapse Channel



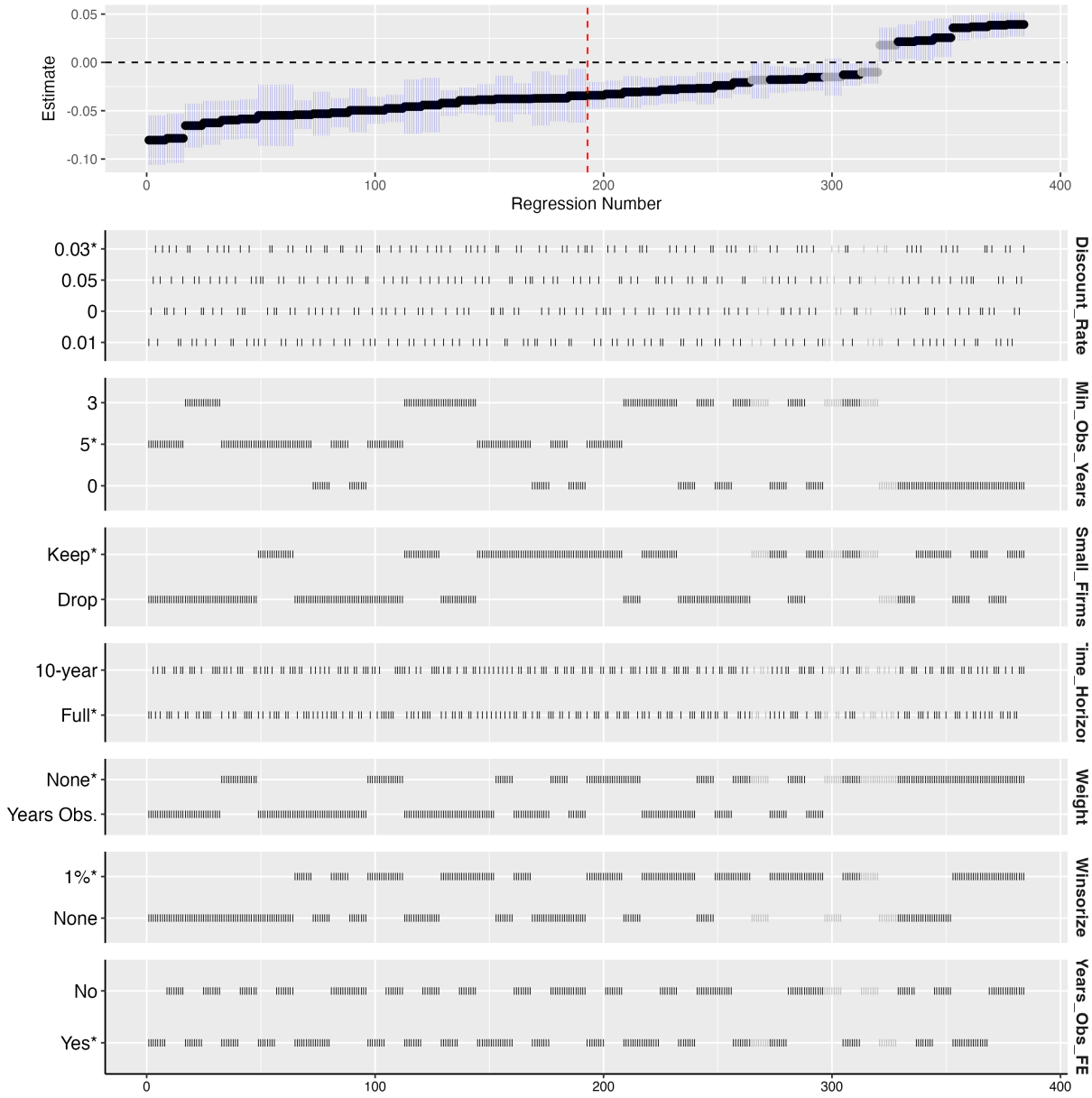
Note: This figure plots plots our estimated lapse regression, with 90% confidence intervals, from 384 sets of regressions featuring different combinations of modeling choices, as described in the text.

Figure E8: Specification Curve: Mortality Channel



Note: This figure plots plots our estimated mortality channel, with 90% confidence intervals, from 384 sets of regressions featuring different combinations of modeling choices, as described in the text.

Figure E9: Specification Curve: Surrender Channel



Note: This figure plots plots our estimated surrender regression, with 90% confidence intervals, from 384 sets of regressions featuring different combinations of modeling choices, as described in the text.

Appendix F Structural Estimation Details

Note that the following appendix requires additional notation. In the interest of legibility, we opted for standard notation using Latin and Greek lettering that is sometimes in conflict with the notation of the main manuscript. In doubt the definition given in this appendix is the relevant one.

F.1 First Stage: FRAC

To estimate the model, we follow the Taylor approximation method developed by Salanié and Wolak (2019) to estimate our discrete-choice demand model efficiently. Specifically, we linearize the random coefficient model by applying a Taylor approximation around its definition at zero variance to circumvent computational difficulties associated with numerical integration over consumer heterogeneity and multiple micro-moments. The procedure leads to reliable estimates as long as the actual variance is small.

We first define a function F, which observed product characteristics $(P_{jt}, \eta_{st}, \phi_{js})$ and parameters (μ_p, σ_p) to unobserved product characteristics ξ_{jst} conditional on observed market shares φ_{jst}. We then approximate F with a second order Taylor approximation around $\sigma_p = 0$

$$F(\mu_p, \sigma_p|\varphi_{jst}) \approx f_0(\varphi_{jst}) - f_{1,p}(\varphi_{jst})\mu_p - f_{1,c}(\varphi_{jst}) - \frac{1}{2} f_2(\varphi_{jst})\sigma_p^2. \tag{A50}$$

Explicitly, the approximating terms become (with φ_{0st} being the market share of the outside option in state s of year t):

$$f_0(\varphi_{jst}) = \ln\left(\frac{\varphi_{jst}}{\varphi_{0st}}\right) \tag{A51}$$

$$f_{1,p}(\varphi_{jst}) = P_{jt} \tag{A52}$$

$$f_{1,c}(\varphi_{jst}) = [\eta_{st}, \phi_{js}] \tag{A53}$$

$$f_2(\varphi_{jst}) = [P_{jt}^2] \tag{A54}$$

Where the first-order term captures average impact (at the firm-state-year) of each part of the utility function and the second-order terms (f_2), approximate the non-linear impacts of the randomness of individual-level wealth and price impacts. We then simply run a 2SLS model regressing f_0 on $f_{1,p}$, $f_{1,c}$ and the "artificial regressors" of f_2 to recover μ_p and σ_p. We address price endogeneity through the same instrument (group-market premiums) as in Section 4.1. We additionally include the square of the group-market premiums to

instrument for P_{jt}^2.

After the initial estimation, we refine the estimates through a Newton-Raphson correction. This involves explicitly solving for unobserved product characteristics ξ_{jt} through the Berry inversion, incorporating the estimated unobserved characteristics into our dependent variable, and iteratively re-estimating the model parameters. This correction substantially reduces approximation biases from the linearized first-stage estimation. See Salanié and Wolak (2019) for further details.

F.2 Second Stage: BLP

Following the FRAC estimation, we use the recovered parameters $(\widehat{\mu_p}, \widehat{\sigma_p})$ as starting values for the full BLP estimation. This stage implements the standard Berry et al. (1995) generalized method of moments (GMM) estimator to account for the full nonlinearity of the random coefficients model.

We solve the BLP contraction mapping to recover the mean utility levels ϑ_{jst} that rationalize the observed market shares:

$$\varphi_{jst} = \int \frac{\exp(\vartheta_{jst} + \kappa_{ijst}(\sigma))}{1 + \sum_{h \in J_s} \exp(\vartheta_{hst} + \kappa_{ihst}(\sigma))} dF_\nu(\nu_i) \tag{A55}$$

where $\kappa_{ijst}(\sigma) = -\sigma_p \nu_i P_{jt}$ captures the individual-specific deviation from mean utility due to unobserved heterogeneity in price sensitivity.

Defining $\theta = \{\mu_p, \sigma_p\}$ as the vector of coefficients, the GMM objective function minimizes:

$$Q(\theta) = F(\theta)' Z W Z' F(\theta) \tag{A56}$$

where $F(\theta) = \vartheta_{jst} - X_{jst}\gamma'$ are the structural errors, Z is the matrix of instruments (group-market premiums and their squares), and W is the GMM weighting matrix. X_{jst} is the vector of product characteristics, i.e. prices and fixed effects, and γ is the associated coefficient vector. We employ a two-step efficient GMM procedure, first using the identity matrix for W, then updating with the optimal weighting matrix $W = (Z'FF'Z)^{-1}$.

F.3 Third Stage: Optimal Instruments

The third stage improves estimation efficiency by constructing optimal instruments following Conlon and Gortmaker (2020). These instruments exploit the model's structure to maximize the correlation with en-

dogenous variables while maintaining orthogonality to the structural errors.

We compute the optimal instruments as:

$$Z_{jst}^{opt} = \mathbb{E}\left[\frac{\partial \xi_{jst}}{\partial \theta}\middle| Z_{jst}\right] \tag{A57}$$

where the expectation is taken with respect to the distribution of unobserved product characteristics, conditional on the original instruments.

In the demand-only setting, we approximate this expectation using the "approximate" method of Conlon and Gortmaker (2020), which replaces the unobserved product characteristics with their expected value of zero and computes the derivative of the demand residual with respect to the parameters. This requires evaluating the Jacobian of the mean utility inversion with respect to the random coefficients:

$$\left.\frac{\partial \xi_{jst}}{\partial \theta}\right|_{\theta=\hat{\theta}} = -\left.\frac{\partial \vartheta_{jst}(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} \tag{A58}$$

where $\hat{\theta}$ is the coefficient vector estimated from Equation (A56), and $\vartheta_{jst}(\theta)$ is the mean utility implied by observed shares under parameter values θ.

We then re-estimate the model using these optimal instruments, which improves the finite-sample properties of the GMM estimator compared to traditional instruments.

F.4 Fourth Stage: Micro-Moments

The final stage incorporates micro-moments from the Health and Retirement Survey to identify the wealth-specific intercepts $\beta_{w,z}$. These moments match the model-predicted insurance ownership rates to observed rates across wealth ventiles.

Define Γ_z as the observed probability of having life insurance for wealth ventile z in the HRS data. The model-predicted probability for ventile z is:

$$\hat{\Gamma}_z(\beta_w) = \frac{1}{|I_z|}\sum_{i \in I_z}\sum_{s,t} \omega_{st} \cdot Pr(i \text{ purchases insurance in market } st|\beta_w) \tag{A59}$$

where I_z is the set of individuals in wealth ventile z, and ω_{st} are market weights defined by total premiums.

We estimate the wealth parameters by minimizing the sum of squared deviations:

$$\min_{\beta_w} \sum_{z \in \mathbb{Q}} (\Gamma_z - \hat{\Gamma}_z(\beta_w))^2 \tag{A60}$$

Given the high dimensionality of the parameter space (20 wealth ventiles), we employ a combination of L-BFGS-B and Nelder-Mead optimization algorithms. The optimization is initialized with a linear gradient from positive values for low-wealth ventiles (reflecting higher baseline utility from insurance) to negative values for high-wealth ventiles.

This micro-moment approach allows us to capture the heterogeneous impact of wealth on insurance demand while maintaining computational tractability. The resulting wealth-specific parameters provide crucial variation for our counterfactual analysis of lapse-contingent pricing and its distributional consequences.