Beliefs identification

December 11, 2025

Adverse selection can be mitigated by some factors, for example Handel (2013) shows that there is inertia in the choice of health insurance plans, which weakens the link between private information and choices. Crawford et al. (2018), in the market for credit, shows that market power weakens adverse selection by increasing the value of the marginal borrower with respect to the average borrower.

1 Model with distorted survival beliefs and guaranteed annuity

Consider an individual i has initial wealth (savings) $W > 0$ at $t = 1$.

The true survival probability between any two consecutive periods is $x_i \in (0, 1)$. Individual beliefs about survival are distorted: the perceived one-period survival probability is

$$\hat{x}_i \equiv \theta x_i,$$

with $\theta > 0$ and $\hat{x}_i \in (0, 1)$.

Individuals also have a heterogenous bequest motive β_i, with CDF F_β on $[0, \infty)$.

The individual has three options.

The first option is no annuity (N), self-insurance with savings. The second option is an immediate annuity (A) that pays $F > 0$ in each period the individual is alive, with no bequest. The third option is a guaranteed annuity (G) that pays $F_g > 0$ in each period the individual is alive; in addition, if the individual dies before T the annuity continues paying to the beneficiaries.

The corresponding values of the options are: $V^N(x, \theta, \beta_i, W)$ for no annuity, $V^A(x, \theta; F)$ for immediate annuity, and $V^G(x, \theta, \beta_i; F_g)$ for guaranteed annuity. For a microfoundation of these value functions see appendix 7.1.

Given $(x, \theta, \beta_i, W, F, F_g)$, the individual chooses the option with highest expected utility:

$$\text{Option chosen} = \arg\max \left\{ V^N(x, \theta, \beta_i, W),\ V^A(x, \theta; F),\ V^G(x, \theta, \beta_i; F_g) \right\}.$$

Non-identification The observed choice is

$$D_i \in \{A, G, N\} \quad \text{with} \quad D_i = \arg\max\{V^A, V^G, V^N\}.$$

We observe the joint distribution of (x_i, D_i); primitives of interest are θ and F_β.

In this case we can prove a non-identification result, see section 7.2.

Identification with price shifters

Previously we showed that data on (x_i, D_i) is not sufficient to identify (θ, F_β)

Adding exogenous price variation. Now allow the annuity payments to depend on an observable "price state" r (e.g. an interest rate or term-structure shifter):

$$F = F(r), \qquad F_g = F_g(r),$$

with r observed by the econometrician and exogenous:

$$r \perp (\beta, x, \theta),$$

and with sufficient support on some compact interval $\mathcal{R}$.

Now the value functions will also be indexed by r, for each (x, r, θ) we obtain two cutoffs

$$0 \leq \beta_1(x, r; \theta) \leq \beta_2(x, r; \theta) \leq \infty,$$

with

$$D = \begin{cases} A, & \beta \leq \beta_1(x, r; \theta), \\ G, & \beta_1(x, r; \theta) < \beta \leq \beta_2(x, r; \theta), \\ N, & \beta > \beta_2(x, r; \theta), \end{cases}$$

and choice probabilities

$$p_A(x, r) = F_\beta\big(\beta_1(x, r; \theta)\big), \tag{1}$$

$$p_A(x, r) + p_G(x, r) = F_\beta\big(\beta_2(x, r; \theta)\big). \tag{2}$$

Why price variation helps: breaking the simple relabeling. Without r, the previous non-identification argument rests on the fact that

$$p_A(x) = F_\beta(\beta_1(x; \theta))$$

only pins down F_β on the *image* of $x \mapsto \beta_1(x; \theta)$, and for any alternative θ_1 we can reparametrize F_β along that image. With r, we now observe $p_A(x, r)$ and $p_A(x, r) + p_G(x, r)$ for a *two-dimensional* regressor $z = (x, r)$, and the relevant composition is

$$p_A(z) = F_\beta(\beta_1(z; \theta)), \qquad p_A(z) + p_G(z) = F_\beta(\beta_2(z; \theta)).$$

A sufficient condition to break the relabeling is:

(C1) Rich price support and invertibility. For each $\theta > 0$, the mappings $z \mapsto \beta_j(z; \theta)$, $j = 1, 2$, are continuous and strictly monotone in r for all x, with images that overlap across different values of $(x, r) \in \mathcal{X} \times \mathcal{R}$. Formally, for any two points z_1, z_2 there exist z_3, z_4 such that

$$\beta_1(z_1; \theta) = \beta_2(z_3; \theta), \qquad \beta_2(z_2; \theta) = \beta_1(z_4; \theta),$$

and $z \mapsto \beta_j(z; \theta)$ are invertible on their images.

Intuitively, $(C1)$ says that as interest rates change, the thresholds $\beta_1(x, r; \theta)$ and $\beta_2(x, r; \theta)$ sweep over the support of β in a way that generates *overlapping evaluations* of F_β at common β values coming from different (x, r) pairs.

Proof of Identification under (C1)

Let (θ_0, F_β^0) be the true parameters generating the observed choice probabilities $p_A(x, r)$ and $p_G(x, r)$. Suppose there exists an alternative pair (θ_1, F_β^1) that is observationally equivalent. That is, for all $(x, r) \in \mathcal{X} \times \mathcal{R}$, the following equalities hold:

$$F_\beta^1(\beta_1(x, r; \theta_1)) = p_A(x, r) = F_\beta^0(\beta_1(x, r; \theta_0)) \tag{3}$$

$$F_\beta^1(\beta_2(x, r; \theta_1)) = p_A(x, r) + p_G(x, r) = F_\beta^0(\beta_2(x, r; \theta_0)) \tag{4}$$

We define the index $j \in \{1, 2\}$ to refer to the relevant margin, where β_1 governs the choice between A and G, and β_2 governs the choice between G and N. We can summarize (3) and (4) as:

$$F_\beta^1(\beta_j(x, r; \theta_1)) = F_\beta^0(\beta_j(x, r; \theta_0)) \quad \text{for } j \in \{1, 2\}. \tag{5}$$

Step 1: Constructing the Mapping ϕ_j

Fix an arbitrary survival probability $x \in \mathcal{X}$ and let b be a bequest level in the image of $\beta_j(x, \cdot; \theta_1)$. By Assumption (C1), the function $r \mapsto \beta_j(x, r; \theta_1)$ is continuous and strictly monotone. Therefore, for a fixed x, there exists a unique inverse price function, denoted as $r_j^*(b, x; \theta_1)$, such that:

$$\beta_j(x, r_j^*(b, x; \theta_1); \theta_1) = b \tag{6}$$

Substitute the price $r = r_j^*(b, x; \theta_1)$ into the observational equivalence condition (5):

$$F_\beta^1(b) = F_\beta^0(\beta_j(x, r_j^*(b, x; \theta_1); \theta_0)) \tag{7}$$

We define the composite mapping $\phi_j(b, x; \theta_0, \theta_1)$ as the cutoff value implied by the true parameter θ_0 at the state (x, r_j^*):

$$\phi_j(b, x; \theta_0, \theta_1) \equiv \beta_j(x, r_j^*(b, x; \theta_1); \theta_0) \tag{8}$$

Using this definition, equation (7) becomes:

$$F_\beta^1(b) = F_\beta^0(\phi_j(b, x; \theta_0, \theta_1)) \tag{9}$$

Step 2: Contradiction via Variation in x

The left-hand side of (9) depends only on b. Consequently, the right-hand side must be invariant to x. This implies that for any two survival probabilities $x, x' \in \mathcal{X}$:

$$F_\beta^0(\phi_j(b, x; \theta_0, \theta_1)) = F_\beta^0(\phi_j(b, x'; \theta_0, \theta_1)) \tag{10}$$

Since F_β^0 is a strictly increasing CDF (on the relevant support), (10) requires:

$$\phi_j(b, x; \theta_0, \theta_1) = \phi_j(b, x'; \theta_0, \theta_1) \tag{11}$$

Substituting the definition from (8), condition (11) requires that the true cutoff $\beta_j(\cdot; \theta_0)$ and the alternative cutoff $\beta_j(\cdot; \theta_1)$ move in perfect synchronization across x and r. Given the non-linear interaction between beliefs θx and prices r in the value functions (specifically the ratio of annuity flows to bequest utility), this condition generically fails unless $\theta_1 = \theta_0$. Thus, θ is point identified.

To prove this, note that equation 11 requires :

$$\frac{\partial \phi_j(b, x; \theta_0, \theta_1)}{\partial x} = 0 \quad \text{for all } x. \tag{12}$$

Differentiating the definition of ϕ_j in (8) with respect to x:

$$\frac{\partial \phi_j}{\partial x} = \frac{\partial \beta_j(\theta_0)}{\partial x} + \frac{\partial \beta_j(\theta_0)}{\partial r} \cdot \frac{\partial r_j^*}{\partial x} \tag{13}$$

From the definition of the inverse price r_j^* in (6), we apply the Implicit Function Theorem to find $\partial r_j^* / \partial x$:

$$\frac{\partial \beta_j(\theta_1)}{\partial x} + \frac{\partial \beta_j(\theta_1)}{\partial r}\frac{\partial r_j^*}{\partial x} = 0 \implies \frac{\partial r_j^*}{\partial x} = -\frac{\partial \beta_j(\theta_1)/\partial x}{\partial \beta_j(\theta_1)/\partial r} \tag{14}$$

Substituting (14) into (13), the condition $\partial \phi_j / \partial x = 0$ is equivalent to:

$$\frac{\partial \beta_j(\theta_0)/\partial x}{\partial \beta_j(\theta_0)/\partial r} = \frac{\partial \beta_j(\theta_1)/\partial x}{\partial \beta_j(\theta_1)/\partial r} \tag{15}$$

To see that (15) does not generally hold when $\theta_1 \neq \theta_0$, it is convenient to consider the case with three periods and $j = 1$ (see section 7.1), in which case:

$$\beta_1(x, r; \theta) = f(\theta x)\, g(r), \qquad f(u) \equiv \frac{1 + u + u^2}{1 - u}, \quad g(r) \equiv \frac{F(r)^\alpha - F_g(r)^\alpha}{F_g(r)}.$$

In which case replacing in equation 15 we have:

$$\frac{f'(\theta_0 x) x g(r)}{f(\theta_0 x) g'(r)} = \frac{f'(\theta_1 x) x g(r)}{f(\theta_1 x) g'(r)} \implies \frac{f'(\theta_0 x)}{f(\theta_0 x)} = \frac{f'(\theta_1 x)}{f(\theta_1 x)} \tag{16}$$

using $\frac{f'(u)}{f(u)} = \frac{-u^2+2u+2}{1-u^3}$ we have:

$$\frac{-(\theta_0 x)^2 + 2(\theta_0 x) + 2}{1 - (\theta_0 x)^3} = \frac{-(\theta_1 x)^2 + 2(\theta_1 x) + 2}{1 - (\theta_1 x)^3} \tag{17}$$

which is not generally true. Therefore, equation (15) holds if and only if $\theta_1 = \theta_0$. Thus, θ_0, F_β^0 are identified. $\square$

The core identification problem is distinguishing whether individuals are buying annuities because they have low bequest motives (low β) or because they are optimistic about their survival (high belief $\hat{x} = \theta x$). In a static setting with fixed prices, these two forces are indistinguishable: a high-bequest pessimist might make the same choice as a low-bequest optimist. We break this observational equivalence by exploiting exogenous variation in annuity prices (driven by interest rates, r). The key insight is that beliefs (θx) and bequest motives (β) interact differently with price changes.Specifically, the "exchange rate" between annuity income and bequest value depends non-linearly on beliefs. When interest rates rise, annuity payouts increase. A rational agent ($\theta = 1$) and a biased agent ($\theta \neq 1$) will re-evaluate the trade-off between the immediate annuity and the guaranteed annuity differently because their perceived "effective price" of the guarantee depends on their survival scaling θ. By observing how the demand for each product shifts across different price levels for the same underlying survival type x, we can trace out a "marginal rate of substitution" curve. The shape of this curve is unique to the specific belief parameter θ.

2 literature review

- O'Dea and Sturrock (2023) studies the same but is not really the interaction between private info and beliefs, rather that people are pessimistic, but do not study the interaction of beliefs with private information.

3 Parametric estimation

We estimate the model parameters $\psi = (\theta, \mu_\beta, \sigma_\beta)$ using the method of Maximum Simulated Likelihood (MSL). The estimation procedure recovers the belief distortion parameter θ and the distribution of bequest motives F_β (parameterized by mean μ_β and standard deviation σ_β) by matching the model-predicted choice probabilities to the observed decisions of individuals.

3.1 Likelihood Function

Let $D_i \in \{N, A, G\}$ denote the observed choice of individual i, where N represents No Annuity, A represents Immediate Annuity, and G represents Guaranteed Annuity. The observables for each individual are the survival probability x_i, the price shifter r_i, initial wealth W, and the annuity payouts $(F(r_i), F_g(r_i))$.

The individual's choice depends on their unobserved bequest motive β_i. We assume β_i is drawn from a normal distribution with mean μ_β and standard deviation σ_β:

$$\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2). \tag{18}$$

Conditional on a specific value of β_i and the parameters θ, the choice probability is deterministic in this structural model. Let $d_i(\beta_i, \theta)$ be the optimal choice function derived from comparing the value functions:

$$d_i(\beta_i, \theta) = \arg\max_{k \in \{N,A,G\}} \left\{ V^N(x_i, \theta, \beta_i, W), V^A(x_i, \theta; F_i), V^G(x_i, \theta, \beta_i; F_{g,i}) \right\}. \tag{19}$$

The unconditional probability of observing choice k for individual i is the integral of the conditional choice indicator over the distribution of unobserved heterogeneity β_i:

$$P(D_i = k | x_i, r_i; \psi) = \int \mathbb{I}(d_i(\beta, \theta) = k)\, f(\beta; \mu_\beta, \sigma_\beta)\, d\beta, \tag{20}$$

where $\mathbb{I}(\cdot)$ is the indicator function and $f(\cdot)$ is the PDF of the normal distribution.

The log-likelihood function for the sample of N individuals is:

$$\mathcal{L}(\psi) = \sum_{i=1}^{N} \ln P(D_i | x_i, r_i; \psi). \tag{21}$$

3.2 Maximum Simulated Likelihood (MSL)

Since the integral in equation (3) does not have a closed-form solution, we approximate it using Monte Carlo simulation. For each individual i, we draw S independent values from a standard normal distribution, denoted as $z_{i,s} \sim \mathcal{N}(0,1)$ for $s = 1, \dots, S$. These draws are fixed throughout the optimization to ensure the objective function is smooth (chattering control).

For a given guess of parameters $(\mu_\beta, \sigma_\beta)$, the simulated bequest motives are:

$$\beta_{i,s} = \max(0, \mu_\beta + \sigma_\beta z_{i,s}). \tag{22}$$

Note that the code enforces $\beta \geq 0$ for the economic logic of the model.

The simulated probability of the observed choice D_i is the frequency of that choice across the S simulations:

$$\hat{P}_i(\psi) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{I}(d_i(\beta_{i,s}, \theta) = D_i). \tag{23}$$

To avoid numerical issues with $\ln(0)$, a small smoothing constant $\epsilon = 10^{-6}$ is added to the probability. The simulated log-likelihood is:

$$\hat{\mathcal{L}}_{MSL}(\psi) = \sum_{i=1}^{N} \ln\left(\hat{P}_i(\psi) + \epsilon\right). \tag{24}$$

The estimator $\hat{\psi}$ is the vector that maximizes $\hat{\mathcal{L}}_{MSL}(\psi)$:

$$\hat{\psi} = \arg\max_{\psi} \sum_{i=1}^{N} \ln\left(\frac{1}{S} \sum_{s=1}^{S} \mathbb{I}(d_i(\beta_{i,s}, \theta) = D_i)\right). \tag{25}$$

3.3 Optimization

The maximization is performed using a constrained optimization algorithm (Sequential Quadratic Programming, SQP). The algorithm searches for θ, μ_β, and $\ln(\sigma_\beta)$ to satisfy the constraints $\theta > 0$ and $\sigma_\beta > 0$.

4 Estimation version

We employ a Maximum Likelihood Estimation (MLE) approach to recover the belief parameter θ and the distribution of bequest motives F_β non-parametrically from the dataset $(x_i, D_i, F_i, F_{gi})_{i=1}^N$.

4.1 Parameterization

To estimate F_β non-parametrically, we approximate the continuous cumulative distribution function (CDF) using a discrete grid. We discretize the support of β into M ordered points (or bins). Let $\mathcal{P}_n$ denote the set of parameters at iteration n:

$$\mathcal{P}_n = (\theta_n, \mathbf{c}_n)$$

where θ_n is the belief distortion parameter and $\mathbf{c}_n = \{c_{1,n}, \dots, c_{M,n}\}$ represents the discretized CDF of β. Specifically, we define a fixed grid of probabilities (quantiles) $q_m = m/M$ for $m = 1, \dots, M$, and we estimate the corresponding quantile values $c_{m,n} = F_\beta^{-1}(q_m)$. Thus, $c_{m,n}$ represents the value of the bequest motive such that $Pr(\beta \leq c_{m,n}) = m/M$. To ensure a valid CDF, we enforce the monotonicity constraint $c_{1,n} \leq c_{2,n} \leq \cdots \leq c_{M,n}$.

4.2 Choice Probabilities

Given the parameter guess $\mathcal{P}_n$, we first compute the model-implied thresholds for each individual i. These thresholds, denoted by $\beta_j(x_i, \theta_n, F_i, F_{gi})$, determine the indifference points between options:

- $\beta_1(x_i, \theta_n, \dots)$: Indifference between Immediate Annuity (A) and Guaranteed Annuity (G).
- $\beta_2(x_i, \theta_n, \dots)$: Indifference between Guaranteed Annuity (G) and No Annuity (N).

Given the thresholds, we can compute the choice probabilities by summing the probability mass that falls into the relevant regions defined by the cutoffs. Under our grid approximation, the probability mass assigned to the interval $(c_{m-1,n}, c_{m,n}]$ is simply $1/M$ (assuming uniform spacing of quantiles). The choice probabilities are:

$$p_A(x_i, F_i, F_{gi}; \mathcal{P}_n) = \sum_{m=1}^{M} \left(\frac{1}{M}\right) \cdot \mathbb{1}\Big(c_{m,n} \leq \beta_1(x_i, \theta_n, F_i, F_{gi})\Big) \tag{26}$$

$$p_G(x_i, F_i, F_{gi}; \mathcal{P}_n) = \sum_{m=1}^{M} \left(\frac{1}{M}\right) \cdot \mathbb{1}\Big(\beta_1(x_i, \theta_n, F_i, F_{gi}) < c_{m,n} \leq \beta_2(x_i, \theta_n, F_i, F_{gi})\Big) \tag{27}$$

$$p_N(x_i, F_i, F_{gi}; \mathcal{P}_n) = 1 - p_A(\cdot) - p_G(\cdot) \tag{28}$$

4.3 Likelihood Function

Once we have the individual choice probabilities, we can construct the individual likelihood contributions. The likelihood contribution of individual i is the probability that the model assigns to the choice actually observed in the data (D_i):

$$\mathcal{L}_i(\theta, \mathbf{c}) = p_A(x_i, F_i, F_{gi}; \mathcal{P}_n)^{\mathbb{1}(D_i=A)} \cdot p_G(x_i, F_i, F_{gi}; \mathcal{P}_n)^{\mathbb{1}(D_i=G)} \cdot p_N(x_i, F_i, F_{gi}; \mathcal{P}_n)^{\mathbb{1}(D_i=N)}. \quad (29)$$

The aggregate log-likelihood function to be maximized is the sum of the individual log-likelihoods:

$$\ln \mathcal{L}(\theta, \mathbf{c}) = \sum_{i=1}^{N} \ln \mathcal{L}_i(\theta, \mathbf{c}). \quad (30)$$

The estimation problem is to find:

$$(\hat{\theta}, \hat{\mathbf{c}}) = \arg\max_{\theta, \mathbf{c}} \sum_{i=1}^{N} \ln \mathcal{L}_i(\theta, \mathbf{c})$$

subject to the constraints $\theta > 0$ and $0 \leq c_1 \leq c_2 \leq \cdots \leq c_M$.

5 Parametric Estimation Algorithm

In this section, we detail the parametric estimation procedure. Unlike the non-parametric approach where we estimate the CDF F_β at discrete grid points, here we assume F_β follows a known parametric distribution (e.g., a truncated normal distribution) governed by a parameter vector γ. Our goal is to estimate the structural parameters $\psi = (\theta, \gamma)$ from the dataset $(x_i, D_i, F_i, F_{gi})_{i=1}^N$.

For concreteness, let us assume $\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$ censored at zero (since $\beta_i \geq 0$). Thus, $\gamma = (\mu_\beta, \sigma_\beta)$.

5.1 Algorithm

The estimation proceeds via Maximum Simulated Likelihood (MSL).

Step 1: Initialization Choose initial values for the parameters $\psi_0 = (\theta_0, \mu_{\beta,0}, \sigma_{\beta,0})$.

Step 2: Simulation Draws (Fixed) To approximate the integral over the unobserved bequest motives, we draw a set of fixed random shocks for each individual.

- For each individual $i = 1, \ldots, N$, draw S independent values $z_{i,s}$ from a standard normal distribution $\mathcal{N}(0, 1)$.
- These draws $\{z_{i,s}\}_{i,s}$ are held constant throughout the optimization process to ensuring the objective function is smooth with respect to the parameters.

Step 3: Likelihood Evaluation For any candidate parameter vector $\psi = (\theta, \mu_\beta, \sigma_\beta)$ during the optimization, we compute the simulated log-likelihood as follows:

1. **Construct Simulated Heterogeneity:** For each individual i and simulation s, construct the bequest motive:

$$\beta_{i,s}(\psi) = \max(0, \mu_\beta + \sigma_\beta z_{i,s})$$

2. **Solve for Optimal Decisions:** For each (i, s), compute the value functions given the current belief parameter θ:

$$\begin{aligned} v_{i,s}^N &= V^N(x_i, \theta, \beta_{i,s}(\psi), W) \\ v_{i,s}^A &= V^A(x_i, \theta; F_i) \\ v_{i,s}^G &= V^G(x_i, \theta, \beta_{i,s}(\psi); F_{gi}) \end{aligned}$$

   Determine the optimal choice $d_{i,s}^* = \arg\max_{k \in \{N,A,G\}} \{v_{i,s}^k\}$.

3. **Compute Choice Probabilities:** Calculate the simulated probability of the observed choice D_i for individual i:

$$\hat{P}_i(D_i | x_i, \ldots; \psi) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(d_{i,s}^* = D_i)$$

Note: In practice, we smooth this indicator function (e.g., using a logistic kernel) or simply add a small $\epsilon > 0$ to avoid taking the log of zero.

4. **Aggregate Log-Likelihood:** Sum the individual contributions:

$$\ln \hat{\mathcal{L}}(\psi) = \sum_{i=1}^{N} \ln \hat{P}_i(D_i | x_i, \ldots; \psi)$$

Step 4: Maximization Find the estimator $\hat{\psi}$ by maximizing the simulated log-likelihood:

$$\hat{\psi} = \arg\max_{\psi} \ln \hat{\mathcal{L}}(\psi)$$

subject to the constraints $\theta > 0$ and $\sigma_\beta > 0$. Standard errors are computed using the inverse Hessian of the likelihood function at the optimum.

6 Other thoughts

- We observe choices, which are a function of beliefs and preferences. Jingi asked me why would we care about identifying beliefs separately from preferences. Is there any policy relevance to this? I think there is none but is just very interesting to separate it. But I should think about welfare implications, for example that there are a lot of people who are actually buying insurance because they do not take into account their private information.

References

Crawford, G. S., Pavanini, N., & Schivardi, F. (2018). Asymmetric Information and Imperfect Competition in Lending Markets. *American Economic Review*, *108*(7), 1659–1701. https://doi.org/10.1257/aer.20150487

Handel, B. R. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts [Publisher: American Economic Association]. *American Economic Review*, *103*(7), 2643–2682. https://doi.org/10.1257/aer.103.7.2643

O'Dea, C., & Sturrock, D. (2023). Survival Pessimism and the Demand for Annuities. *The Review of Economics and Statistics*, *105*(2), 442–457. https://doi.org/10.1162/rest_a_01048

7 Appendix

7.1 Microfoundations for value functions

Time is $t = 1, 2, 3$. Each individual i has initial wealth (savings) $W > 0$ at $t = 1$.

The true survival probability between any two consecutive periods is $x_i \in (0, 1)$. Individual beliefs about survival are distorted: the perceived one-period survival probability is

$$\hat{x}_i \equiv \theta x_i,$$

with $\theta > 0$ and $\hat{x}_i \in (0, 1)$. Period utility from consumption is

$$u(c) = c^{\alpha}, \qquad \alpha \in (0, 1).$$

The bequest weight β_i is heterogeneous with CDF F_{β} on $[0, \infty)$. Bequest utility is linear:

$$v(B) = \beta_i B, \qquad \beta_i \geq 0.$$

We assume no time discounting.

The individual has three options at $t = 1$. The first option is no annuity (N), self-insurance with savings. The second option is an immediate annuity (A) that pays $F > 0$ in each period the individual is alive, with no bequest. The third option is a guaranteed annuity (G) that pays $F_g > 0$ in each period the individual is alive; in addition, if the individual dies between periods 1 and 2, the contract pays a guaranteed amount F_g at $t = 2$ as a pure bequest (no consumption).

1. No annuity: consumption–savings problem and value. If the individual does not buy an annuity, she chooses a consumption–savings plan:

$$c_1, c_2, c_3 \geq 0$$

subject to $c_1 + c_2 + c_3 = W$ and $c_1, c_2, c_3 \geq 0$.

Given beliefs $\hat{x} = \theta x$, the expected utility from a given plan (c_1, c_2, c_3) is

$$U^N(c_1, c_2, c_3; x, \theta, \beta_i) = u(c_1) + (1 - \hat{x})\, \beta_i (W - c_1) + \hat{x} \Big[u(c_2) + (1 - \hat{x})\, \beta_i (W - c_1 - c_2) + \hat{x}\, u(c_3) \Big].$$

The value of self-insurance is

$$V^N(x, \theta, \beta_i, W) \equiv \max_{\substack{c_1, c_2, c_3 \geq 0 \\ c_1 + c_2 + c_3 = W}} U^N(c_1, c_2, c_3; x, \theta, \beta_i).$$

Sometimes we also use s_t for savings at the beginning of period t, for instance $s_1 = W - c_1$.

2. Immediate annuity (A). If the individual buys the immediate annuity, she uses all wealth W to purchase a contract that pays F in each period she is alive. There is no bequest from the annuity.

Expected utility under the immediate annuity is

$$V^A(x, \theta) = \left(1 + \hat{x} + \hat{x}^2\right) F^{\alpha}.$$

3. Guaranteed annuity (G). If the individual buys the guaranteed annuity, she uses all wealth W to purchase a contract that:

- pays F_g each period she is alive (as with an immediate annuity),
- if she dies between $t = 1$ and $t = 2$, pays a guaranteed amount F_g at $t = 2$ as a bequest (no consumption).

Hence, the expected utility under the guaranteed annuity is

$$\begin{aligned} V^G(x, \theta, \beta_i) &= u(F_g) + \hat{x}\, u(F_g) + \hat{x}^2\, u(F_g) + (1 - \hat{x})\, v(F_g) \\ &= \left(1 + \hat{x} + \hat{x}^2\right) F_g^\alpha + (1 - \hat{x})\, \beta_i F_g. \end{aligned}$$

4. Choice among N, A, and G. Given (x, θ, β_i, W), the individual chooses the option with highest expected utility:

$$\text{Option chosen} = \arg\max \left\{V^N(x, \theta, \beta_i, W),\, V^A(x, \theta),\, V^G(x, \theta, \beta_i)\right\}.$$

7.2 Non-identification result

Assumption 1 (Monotone partition in β_i)**.** *Assume that there exist cutoffs* $0 \le \beta_1(x;\theta) \le \beta_2(x;\theta)$ *such that,*

$$D_i = \begin{cases} A, & \text{if } \beta_i \le \beta_1(x_i;\theta), \\ G, & \text{if } \beta_1(x_i;\theta) < \beta_i \le \beta_2(x_i;\theta), \\ N, & \text{if } \beta_i > \beta_2(x_i;\theta). \end{cases}$$

Assumption 1 is intuitive, it says that individuals with a higher bequest motive will buy less insurance since they do not mind leavin savings to their beneficiaries. A sufficient condition is that $s_1(\beta_i = 0; x) > F_g$. [1]

Define the conditional choice probabilities

$$p_k(x) \equiv \Pr(D_i = k \mid x_i = x), \qquad k \in \{A, G, N\}.$$

By Assumption 1, the model implies:

$$p_A(x) = \Pr(\beta_i \le \beta_1(x;\theta)) = F_\beta\big(\beta_1(x;\theta)\big), \tag{31}$$

$$p_A(x) + p_G(x) = \Pr(\beta_i \le \beta_2(x;\theta)) = F_\beta\big(\beta_2(x;\theta)\big), \tag{32}$$

$$p_N(x) = 1 - \big[p_A(x) + p_G(x)\big] = 1 - F_\beta\big(\beta_2(x;\theta)\big). \tag{33}$$

[1] Given that $\frac{\partial V^G}{\partial \beta_i} = (1-x)\beta_i F_g$ and that $\frac{\partial V^N}{\partial \beta_i} > \beta_i(1-x)[s_1(\beta_i) + x s_2(\beta_i)]$ if $s_1(\beta_i = 0; x) > F_g$ then, given that savings are increasing on β_i, we have that $s_1(\beta_i; x) > F_g$. Then $\frac{\partial V^G}{\partial \beta_i}(1-x) = \beta_i F_g < \beta_i(1-x)[s_1(\beta_i) + x s_2(\beta_i)] < \frac{\partial V^N}{\partial \beta_i}$

Non-identification of (θ, F_β) (nonparametric F_β). Assume:

(A1) For each $\theta > 0$, the functions $x \mapsto \beta_1(x;\theta)$ and $x \mapsto \beta_2(x;\theta)$ are strictly monotone and continuous on the support of x, hence invertible with inverses $b \mapsto x_1^\theta(b)$ and $b \mapsto x_2^\theta(b)$.

(A2) The images of $\beta_1(\cdot;\theta)$ and $\beta_2(\cdot;\theta)$ are disjoint (or can be made disjoint by restricting attention to suitable subsets of x).

In section 7.3 we provide intuitive conditions under which this assumptions are satisfied.

Suppose the data are generated by some true pair (θ_0, F_β^0), yielding observed choice probabilities

$$p_A(x) = F_\beta^0\big(\beta_1(x;\theta_0)\big), \qquad p_A(x) + p_G(x) = F_\beta^0\big(\beta_2(x;\theta_0)\big),$$

for all x.

Proposition 2 (Non-identification)**.** *Under (A1)–(A2), for any alternative $\theta_1 > 0$ there exists a CDF F_β^1 such that the pair (θ_1, F_β^1) generates the same conditional choice probabilities $\{p_A(x), p_G(x), p_N(x)\}$ for all x. Hence (θ, F_β) is not point-identified.*

Proof. Fix $\theta_1 > 0$ and consider the cutoff functions $\beta_1(x;\theta_1)$ and $\beta_2(x;\theta_1)$. By (A1), define inverses $x_1^{\theta_1}(b)$ and $x_2^{\theta_1}(b)$ on the images of $\beta_1(\cdot;\theta_1)$ and $\beta_2(\cdot;\theta_1)$, respectively, such that

$$\beta_1\big(x_1^{\theta_1}(b);\theta_1\big) = b, \qquad \beta_2\big(x_2^{\theta_1}(b);\theta_1\big) = b.$$

Define F_β^1 on these images by

$$\begin{aligned} F_\beta^1(b) &\equiv p_A\big(x_1^{\theta_1}(b)\big), && \text{for } b \in \operatorname{Im}\big(\beta_1(\cdot;\theta_1)\big), \\ F_\beta^1(b) &\equiv p_A\big(x_2^{\theta_1}(b)\big) + p_G\big(x_2^{\theta_1}(b)\big), && \text{for } b \in \operatorname{Im}\big(\beta_2(\cdot;\theta_1)\big). \end{aligned}$$

Disjointness of the images (A2) guarantees that F_β^1 is well-defined. Monotonicity of p_A and $p_A + p_G$ in x and of $x_1^{\theta_1}$, $x_2^{\theta_1}$ in b implies F_β^1 is nondecreasing in b on the union of these images, and it can be extended arbitrarily (but monotonically) to a full CDF on $[0, \infty)$.

Now, for any x,

$$F_\beta^1\big(\beta_1(x;\theta_1)\big) = p_A(x), \qquad F_\beta^1\big(\beta_2(x;\theta_1)\big) = p_A(x) + p_G(x).$$

Hence the model-implied probabilities under (θ_1, F_β^1) satisfy

$$p_A(x;\theta_1, F_\beta^1) = F_\beta^1\big(\beta_1(x;\theta_1)\big) = p_A(x),$$

and

$$p_A(x;\theta_1, F_\beta^1) + p_G(x;\theta_1, F_\beta^1) = F_\beta^1\big(\beta_2(x;\theta_1)\big) = p_A(x) + p_G(x),$$

so $p_G(x;\theta_1, F_\beta^1) = p_G(x)$ and $p_N(x;\theta_1, F_\beta^1) = 1 - p_A(x) - p_G(x) = p_N(x)$ for all x.

Therefore the observed conditional choice probabilities do not uniquely determine (θ, F_β): for each $\theta_1 > 0$ we can construct a CDF F_β^1 that exactly reproduces the same $\{p_A(x), p_G(x), p_N(x)\}_x$. $\square$

7.3 Microfoundations for (A1) and (A2)

In this subsection we show how the cutoff properties used in Step 3 follow from the underlying value functions, under mild assumptions on primitives, instead of being imposed directly on the reduced-form objects $\beta_1(\cdot;\theta)$, $\beta_2(\cdot;\theta)$.

Explicit expression and monotonicity of $\beta_1(x;\theta)$. Recall that the A–G cutoff $\beta_1(x;\theta)$ is defined implicitly by

$$V^A(x,\theta) \;=\; V^G(x,\theta,\beta_1(x;\theta)).$$

Replacing for the value functions, and rearranging we have that the cutoff solves:

$$\beta_1(x;\theta) = \frac{1+\hat{x}+\hat{x}^2}{1-\hat{x}} \cdot \frac{F^\alpha - F_g^\alpha}{F_g}, \tag{34}$$

Under our maintained primitives $\hat{x} \in (0,1)$, $F > 0$, $F_g > 0$, and $u(c) = c^\alpha$ with $\alpha \in (0,1)$, (34) is well-defined and continuous in x for any $\theta > 0$. In addition, if we assume that the guaranteed annuity pays a (weakly) lower flow than the plain annuity,

$$F > F_g > 0, \tag{35}$$

then $F^\alpha - F_g^\alpha > 0$ and the factor $(F^\alpha - F_g^\alpha)/F_g$ is strictly positive.

To study monotonicity, define

$$f(\hat{x}) \;\equiv\; \frac{1+\hat{x}+\hat{x}^2}{1-\hat{x}}, \qquad \hat{x} \in (0,1).$$

Then

$$f'(\hat{x}) \;=\; \frac{(1+2\hat{x})(1-\hat{x}) + (1+\hat{x}+\hat{x}^2)}{(1-\hat{x})^2} \;=\; \frac{2+2\hat{x}-\hat{x}^2}{(1-\hat{x})^2}.$$

For $\hat{x} \in (0,1)$, the numerator satisfies $2+2\hat{x}-\hat{x}^2 > 2 > 0$ and the denominator is positive, so $f'(\hat{x}) > 0$ on $(0,1)$. Since $\hat{x} = \theta x$ with $\theta > 0$, the composite $x \mapsto f(\theta x)$ is strictly increasing and continuous on the support of x. Combining this with (34) and (35), we obtain:

Lemma 3 (Microfoundations for the A–G cutoff)**.** *Under $\hat{x} = \theta x \in (0,1)$, $F > F_g > 0$, and $u(c) = c^\alpha$ with $\alpha \in (0,1)$, the A–G cutoff $\beta_1(x;\theta)$ is given by (34) and is continuous and strictly increasing in x for any fixed $\theta > 0$.*

This shows that the monotonicity and continuity of $\beta_1(\cdot;\theta)$ in Assumption (A1) are direct implications of the primitives.

A sufficient condition for the monotonicity of $\beta_2(x;\theta)$ is that $s_1(\beta_i = 0; x) > F_g$. Given that $\frac{\partial V^G}{\partial \beta_i} = (1-x)\beta_i F_g$ and that $\frac{\partial V^N}{\partial \beta_i} > \beta_i(1-x)[s_1(\beta_i) + x s_2(\beta_i)]$ if $s_1(\beta_i = 0; x) > F_g$ then, given that savings are increasing on β_i, we have that $s_1(\beta_i; x) > F_g$. Then $\frac{\partial V^G}{\partial \beta_i}(1-x) = \beta_i F_g < \beta_i(1-x)[s_1(\beta_i) + x s_2(\beta_i)] < \frac{\partial V^N}{\partial \beta_i}$

A simple sufficient condition for (A2). Recall that (A2) requires the images of the cutoff functions $\beta_1(\cdot;\theta)$ and $\beta_2(\cdot;\theta)$ to be disjoint (or made disjoint by restricting the range of x). A convenient and transparent sufficient condition is a *uniform separation* of the two cutoffs in β.

Let $\mathcal{X}$ denote the support of x and assume $\mathcal{X}$ is compact.

Assumption 4 (Uniform separation of cutoffs)**.** *For a given $\theta > 0$, suppose there exists a constant $\delta > 0$ such that*

$$\beta_2(x;\theta) - \beta_1(x;\theta) \;\geq\; \delta \qquad \textit{for all } x \in \mathcal{X}.$$

Lemma 5 (Assumption 4 implies (A2))**.** *Fix $\theta > 0$ and suppose Assumption 4 holds. Then the images* $\mathrm{Im}(\beta_1(\cdot;\theta))$ *and* $\mathrm{Im}(\beta_2(\cdot;\theta))$ *are disjoint. In particular, condition* (A2) *holds.*

Proof. Because $\mathcal{X}$ is compact and $\beta_j(\cdot;\theta)$ is continuous for $j = 1, 2$, each image $\mathrm{Im}(\beta_j(\cdot;\theta))$ is a compact interval in $\mathbb{R}$. Define

$$\overline{\beta}_1 \equiv \sup_{x\in\mathcal{X}} \beta_1(x;\theta), \qquad \underline{\beta}_2 \equiv \inf_{x\in\mathcal{X}} \beta_2(x;\theta).$$

By Assumption 4,

$$\beta_2(x;\theta) \;\geq\; \beta_1(x;\theta) + \delta \quad \text{for all } x.$$

Taking the infimum over x on the left-hand side and the supremum over x on the right-hand side yields

$$\underline{\beta}_2 \;\geq\; \overline{\beta}_1 + \delta.$$

In particular, $\overline{\beta}_1 < \underline{\beta}_2$. Hence

$$\mathrm{Im}(\beta_1(\cdot;\theta)) \;\subseteq\; (-\infty, \overline{\beta}_1], \qquad \mathrm{Im}(\beta_2(\cdot;\theta)) \;\subseteq\; [\underline{\beta}_2, \infty),$$

and the two images are disjoint because the right endpoint of the first is strictly smaller than the left endpoint of the second. This is exactly the content of (A2). □

7.4 Microfoundations for C1

[WORK IN PROGRESS]

Support of the price state and plausibility of (C1)

We now formalize what we mean by "sufficient support" of the price state r, and we provide simple primitive conditions under which condition (C1) is satisfied. Throughout, recall that r affects payoffs only through the annuity payments $(F(r), F_g(r))$, is observed by the econometrician, and is exogenous: $r \perp (x, \beta, \theta)$.:contentReference[oaicite:0]index=0

Formalizing "sufficient support on a compact interval $\mathcal{R}$". Let $\mathcal{R} \subset \mathbb{R}$ denote the support of r.

Assumption 6 (Rich support of the price state)**.** *There exist real numbers $\underline{r} < \overline{r}$ such that:*

1. *$\mathcal{R} = [\underline{r}, \overline{r}]$ is a non-degenerate compact interval.*
2. *The distribution of r has* full support *on $\mathcal{R}$; that is, for any subinterval $[a, b] \subseteq [\underline{r}, \overline{r}]$ with $a < b$,*
$$\Pr\big(r \in [a, b] \mid x, \beta\big) > 0.$$
Equivalently, the conditional density $f_{r|x,\beta}(r \mid x, \beta)$ is strictly positive on $(\underline{r}, \overline{r})$.

Assumption 6 formalizes "sufficient support" as saying that the price state r can take any value in a whole interval $[\underline{r}, \overline{r}]$ with positive probability.

Primitive conditions for (C1). We want conditions stated in terms of primitives $(F(r), F_g(r))$ and the value functions (V^A, V^G, V^N) under which (C1) is satisfied.

Assumption 7 (Price monotonicity of annuity flows)**.** *The annuity flows respond smoothly and non-degenerately to r:*

1. *$F(r)$ and $F_g(r)$ are continuously differentiable on $\mathcal{R}$.*
2. *$F(r)$ and $F_g(r)$ are strictly monotone in r (either both increasing or both decreasing), and their ratio varies with r:*
$$\frac{d}{dr}\left(\frac{F(r)^\alpha - F_g(r)^\alpha}{F_g(r)}\right) \neq 0 \quad \textit{for all } r \in (\underline{r}, \overline{r}).$$

Assumption 7 says that interest-rate movements generate genuine shifts in the relative attractiveness of the plain and guaranteed annuities: as r changes, the "implicit price" of the guarantee (in units of $F_g(r)$) moves in a strictly monotone way.

Under our microfoundations, the A–G cutoff is explicitly

$$\beta_1(x, r; \theta) = \frac{1 + \hat{x} + \hat{x}^2}{1 - \hat{x}} \cdot \frac{F(r)^\alpha - F_g(r)^\alpha}{F_g(r)}, \qquad \hat{x} = \theta x,$$

so Assumption 7 immediately implies that for each fixed (x, θ), $r \mapsto \beta_1(x, r; \theta)$ is continuous and strictly monotone on $\mathcal{R}$, and, by Assumption 6, its image is an interval of β-values with non-empty interior.

For the G–N cutoff, recall that $\beta_2(x, r; \theta)$ is defined by

$$V^N(x, \theta, \beta_2(x, r; \theta), W) = V^G(x, r, \theta, \beta_2(x, r; \theta)),$$

where
$$V^G(x, r, \theta, \beta) = (1 + \hat{x} + \hat{x}^2) F_g(r)^\alpha + (1 - \hat{x})\beta\, F_g(r), \quad \hat{x} = \theta x.$$

[Why does β_2 satisfy the assumptions]