Up next

- [Done] Testing for selection
- Empirical welfare analysis I: Using data on choices and claims
- Empirical welfare analysis II: What happens when you can’t use choice data
  - Don’t trust revealed preference
  - Markets don’t exist
Empirical welfare analysis I

- Welfare cost of selection
- Welfare consequences of government intervention
Two approaches to the same question:
- Einav, Finkelstein and Cullen (QJE 2010)
- Einav, Finkelstein and Schrimpf (EMA 2010)

Emphasize tradeoffs of approaches: more and less structural
- See also Chetty (AR 2009) “Sufficient Statistics”

For more discussion of welfare analysis in insurance markets see:
Einav, Finkelstein and Levin (Annual Review 2010)
Welfare analysis: emphasized in PF

- One distinguishing feature of PF (vs e.g. applied public policy, labor economics etc.) is the attention to welfare (in private markets, of government policy etc)
- But making welfare statements usually requires additional assumptions
  - Do assumptions drive the result? Is result robust to alternative plausible assumptions?
  - How far can we get w the fewest possible assumptions? If we make more assumptions what is it buying us?
Empirical welfare analysis

- Efficiency cost of adverse selection
  - Once know there is private information, want to know how great efficiency cost is

- Welfare consequences of alternative public policies
  - Can public policy improve on adverse selection equilibm?
  - Fundamentally an empirical question
    - E.g. Mandates as canonical solution to adverse selection (underinsurance) problem.
    - However, once have preference heterogeneity, potential costs from allocative distortions of mandates (vs allocative distortions from adverse selection). Recall graphs (w interior crossing; empirical question which triangles are bigger)
Welfare an empirical question

- Demand curve
- AC curve
- MC curve

Price vs. Quantity graph:
- $P_{eqm}$
- $P_{eff}$
- $Q_{eqm}$
- $Q_{max}$
- $Q_{eff}$

May not be efficient to insure all
Welfare inferences from extent of pos correlation?

- Some markets with private information about risk type appear more adversely selected than others
  - i.e. larger vs smaller positive correlation
  - Are these markets where efficiency costs likely to be greater?

- Cannot even make qualitative statements about where efficiency cost of adverse selection are likely to be larger vs smaller based on magnitude of reduced form correlation between insurance coverage and risk type
  - Play with the graphs: holding AC of insured vs uninsured same, can rotate demand to get v different welfare costs.
Welfare inferences from extent of pos correlation?

However, these demand curves generate different efficient outcomes, meaning different points at which the two demand curves intersect the MC curve, denoted in the figure by points $E_1$ and $E_2$.

As a result, they produce different-sized welfare losses, given by the corresponding triangles $CDE_1$ and $CDE_2$. This example thus illustrates how deadweight loss triangles of different sizes can be generated even though the "extent of adverse selection" as measured by the difference in average costs is the same.

One way to make some progress in quantifying the welfare consequences of selection or of potential public policy is to use bounds that are based on easily observable objects. For example, suppose we would like to bound the welfare cost of selection. We use Figure 1 (adverse selection) for this discussion, but it is easy to imagine an analogous discussion for the advantageous selection shown in Figure 4.

Suppose first that we observe only the price of the insurance sold in the market. If we are willing to assume that we observe the competitive equilibrium price ($P_{eqm}$), we can obtain a (presumably not very tight) upper bound of the welfare cost of selection.
How to estimate welfare cost of selection

- Need more than the reduced form (positive correlation)
- Will now discuss two approaches:
  - Einav, Finkelstein and Cullen (QJE 2010).
    - “Sufficient statistics” approach
    - Relatively little structure, but also limited in what analyses we can do
  - Einav, Finkelstein and Schrimpf (EMA 2010)
    - More “structural”
    - More (questionable) assumptions but ability to do richer analyses (at least in principle)
EFC (2010): The big picture

- How far can we get on welfare using relatively few assumptions?
  - In particular, if we have price variation in contracts offered, and do not try to estimate underlying primitives (risk type and risk aversion).

- Basic idea:
  - Rely on standard consumer and producer theory
  - Key feature of selection markets: firms’ costs depend on which consumers purchase their products (“endogenous cost curve”)
  - price variation can trace out demand & cost curve

- Develop approach and show application to employer provided health insurance
  - Focus: strengths and limitations of approach
Theory: Setup and notation

- Only two contracts: $H$ (full coverage) and $L$ (no coverage)
  - Easy to extend to other or more contracts (harder to draw)
  - $p = p_H - p_L$ is the relative price of contract $H$

- Key assumption: take non-price characteristics of insurance contracts as given
  - As in Akerlof (1970) compared to Rothschild and Stiglitz (1976)
  - Empirically relevant – often observably different individuals offered same menu of contract, just at different prices

- Individuals defined by a vector of attributes $\zeta_i \sim G(\zeta)$, and have to choose a contract $H$ or $L$
  - $\zeta_i$ includes preferences, information set (i.e. expected claims) etc.
  - $\zeta_i$ is what we will try to estimate in EFS (EMA 2010)
    - Clearly with underlying primitives can do a lot!
  - Key here is that we will try to do (some) welfare analysis w/o estimating $\zeta$
\( \pi(\zeta_i) \) is willingness to pay for \( H \) (i.e., \( v_H(\zeta_i, \pi(\zeta_i)) = v_L(\zeta_i) \))

\( c(\zeta_i) \) is the expected insurable costs under \( H \)

- Cost to insurance company of insuring the individual (ignoring any administrative costs)
- Abstract from moral hazard for now for notational simplicity (will come back to)
Demand:

\[ D(p) = \Pr(\pi(\zeta_i) \geq p) \]

Supply:

- \( N \geq 2 \) identical risk neutral insurance providers, who set prices in a Nash Equilibrium (a-la Bertrand)
- Average cost (AC):

\[ AC(p) = E(c(\zeta) | \pi(\zeta) \geq p) \]

- Marginal cost (MC):

\[ MC(p) = E(c(\zeta) | \pi(\zeta) = p) \]

Additional (standard) assumptions \( \rightarrow \) Equilibrium exists, unique, and given by the lowest break-even price:

\[ p^* = \min \{ p : p = AC(p) \} \]
Welfare definitions

- Total surplus from allocating $H$ to individual $i$ is

$$TS(\zeta_i) = \pi(\zeta_i) - c(\zeta_i)$$

- First best allocation: individual $i$ purchases insurance if and only if

$$\pi(\zeta_i) \geq c(\zeta_i)$$

- Constrained efficient allocation: maximizes social welfare subject to the constraint that price is the only instrument available for screening.
  - Constrained efficient: individual $i$ purchases insurance if and only if

$$\pi(\zeta_i) \geq E(c(\tilde{\zeta}) | \pi(\tilde{\zeta}) = \pi(\zeta_i))$$
Welfare cost of adverse selection

If have estimated these curves, have welfare cost of selection (CDE).

Could also evaluate consequences of: subsidies, mandates, pricing on X’s...
Graphical analysis illustrates that demand and cost curves are sufficient statistics for welfare analysis of pricing of contracts.

Empirical approach: estimate demand and cost curves but remain agnostic about underlying primitives that give rise to them.

- Analogy to Chetty (JPE 2008) who shows how key behavioral elasticities are sufficient statistics for welfare analysis of optimal level of UI benefits. (Will discuss this in next set of lectures. Key point: statistics are sufficient conditional on a model...)

We remain agnostic about underlying primitives (ζ_i) that give rise to demand and cost curve.

- e.g. active vs passive selection generating cost curve?
Sufficient statistics for welfare analysis are:

- the demand curve $D(p)$
- the average cost curve $AC(p)$

Estimation:

$$D_i = \alpha + \beta p_i + \epsilon_i \text{ for everyone}$$
$$c_i = \gamma + \delta p_i + u_i \text{ for those who endogenously chose } H$$

Requires

- To estimate $D(p)$ variation in $p$ exogenous to demand & quantity
- To estimate $AC(p)$: same variation in $p$ & cost data for sample who endogenously choose $H$

Conceptually, variation in $p$ identifies all curves non-parameterically. In practice, likely that need to make functional form assumptions.
Estimation (con’t)

- From $D(p)$ and $AC(p)$ we can back out $MC(p)$:

$$MC(p) = \frac{\partial (AC(p) \cdot D(p))}{\partial D(p)} = \left( \frac{\partial D(p)}{\partial p} \right)^{-1} \frac{\partial (AC(p) \cdot D(p))}{\partial p}$$

- Conceptually, variation in $p$ identifies all curves non-parametrically. In practice, likely that need to make functional form assumptions.
- Here structure could be useful to guide functional form
- But graphs highlight which parts of curves are important to “get right”

- Key requirement: Need variation in $p$ that is exogenous with respect to demand and cost
Market for employer provided health insurance in the U.S.

Major source of private insurance (90%)

Tax subsidy for employer provided health insurance = single largest federal tax expenditures

Prior evidence of asymmetric information in this market
  
  "positive correlation"

Death spirals (Cutler Reber 1998)
Data and setting

- Individual-level data from 2004 on U.S.-based employees of a large multi-national aluminum manufacturer
  - New health insurance options introduced for 2004

- Data include:
  - The menu of health insurance options available to each employee
  - The premium associated with each option
  - Employee choices
  - Employee (and dependents’) subsequent medical expenditure
  - Rich demographics – everything price setter likely to observe
Price variation

- Want exogenous variation in \( p_i = p_i^H - p_i^L \).
- Have 40 (decentralized) business units within company each pick from 6 pricing menus proposed by HQ.
- Is choice of pricing menu correlated with employee demand or expected costs?
  - A priori pricing variation seemed more likely exogenous / driven by idiosyncratic aspects of BU president.
    - accountants, paralegals, metallurgists, and administrative assistants may face different prices because they are affiliated with "primary metals" instead of "rigid packaging".
  - Born out by data: prices are not correlated with observables of our sample of salaried workers (see Table 2).
- But this is far from ideal.
Empirical constructs

- \( p_i = p_i^H - p_i^L \) where \( p_i^j \) is employee \( i \)'s annual contribution for coverage \( j \)
- \( D_i = 1 \) if \( i \) chose \( H \); \( D_i = 0 \) if \( i \) chose \( L \)
- \( m_i \) is employee \( i \)'s vector of medical cost during 2004
- \( c(m_i; j) \) is the insurer's cost of covering \( m_i \) under coverage \( j \)
- \( c_i = c(m_i; H) - c(m_i; L) \) is the incremental insurer's costs from covering \( i \) with \( H \) vs. \( L \) (holding behavior \( m_i \) fixed)
  - Note one will be counterfactual so need to construct (both ideally) using plan rules
We estimate (using OLS):

\[ D_i = \alpha + \beta p_i + \epsilon_i \] for everyone
\[ c_i = \gamma + \delta p_i + u_i \] for those who chose \( H \)

recall \( c_i = c(m_i; H) - c(m_i; L) \)

Marginal cost derived from these without additional estimation
### Table II

The Effect of Price on Demand and Costs

<table>
<thead>
<tr>
<th>(Relative) price ($)</th>
<th>Number of employees</th>
<th>Fraction chose contract $H$</th>
<th>Average incremental cost ($) for those covered under Contract $H$</th>
<th>Average incremental cost ($) for Contract $L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>384</td>
<td>2,939</td>
<td>0.67</td>
<td>451.40</td>
<td>425.48</td>
</tr>
<tr>
<td>466</td>
<td>67</td>
<td>0.66</td>
<td>499.32</td>
<td>423.30</td>
</tr>
<tr>
<td>489</td>
<td>7</td>
<td>0.43</td>
<td>661.27</td>
<td>517.00</td>
</tr>
<tr>
<td>495</td>
<td>526</td>
<td>0.64</td>
<td>458.60</td>
<td>421.42</td>
</tr>
<tr>
<td>570</td>
<td>199</td>
<td>0.46</td>
<td>492.59</td>
<td>438.83</td>
</tr>
<tr>
<td>659</td>
<td>41</td>
<td>0.49</td>
<td>489.05</td>
<td>448.50</td>
</tr>
</tbody>
</table>

Notes. The table presents the raw data underlying our baseline estimates. All individuals face one of six different (relative) prices, each represented by a row in the table. Column (2) reports the number of employees facing each price, and column (3) reports the fraction of them who chose contract $H$. Columns (4) and (5) report (for individuals covered by contracts $H$ and $L$, respectively) the average incremental costs to the insurer of covering these individuals with contract $H$ rather than with contract $L$, taking the family's medical expenditures as given. The graphical analog to this table is presented by the circles shown in Figure V.

This pattern of average costs indicates the existence of adverse selection (see Figure I). Column (5) shows the same for the individuals who (endogenously) select contract $L$. Recall that incremental cost is defined as the difference in costs to the insurer associated with a given employee's family's medical expenditures if those expenditures were insured under contract $H$ rather than contract $L$. As shown in Figure III, this difference is a nonlinear function of expenditures.

In the spirit of the “positive correlation” test (Chiappori and Salanie 2000), a comparison of columns (5) and (4) reveals consistently higher average costs for those covered by contract $H$ than for those covered by contract $L$. This indicates that either moral hazard or adverse selection is present. Detecting whether selection is present, and if so what its welfare consequences are, requires the use of our pricing variation, to which we now turn.

In column (1) of Table III we report OLS estimates of equation (11) with no additional controls. We obtain a downward-sloping demand curve, with a (statistically significant) slope coefficient $\beta$ of $-0.00070$. This implies that a $100 increase in price reduces the probability that the employee chooses contract $H$ by a statistically significant seven percentage points, or about 11%.

In column (2) of Table III we use OLS to separately estimate the average cost curve in equation (12). We obtain a (statistically significant) slope coefficient of $-0.00092$. This implies that a $100 increase in price reduces the probability that the employee chooses contract $H$ by a statistically significant nine percentage points, or about 21%.
### Table III

#### Estimation Results

<table>
<thead>
<tr>
<th>Dependent variable (sample)</th>
<th>1 if chose High (both High and Low)</th>
<th>Incremental cost (only High)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Estimation results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative price of High (US$)</td>
<td>$-0.00070$ (0.00032) [.034]</td>
<td>$0.15524$ (0.06388) [.021]</td>
</tr>
<tr>
<td>Constant</td>
<td>$0.940$ (0.123) [.000]</td>
<td>$391.690$ (26.789) [.000]</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>$0.652$</td>
<td>$455.341$</td>
</tr>
<tr>
<td>Number of observations</td>
<td>$3,779$</td>
<td>$2,465$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$0.008$</td>
<td>$0.005$</td>
</tr>
</tbody>
</table>

Standard errors (in parentheses) clustered on state

p-values in [square brackets]
The welfare cost of adverse selection

May not be efficient to insure all

Finkelstein ()
PF Slides
Spring 2018
This figure is the empirical analog of the theoretical Figure I. The demand curve and AC curve are graphed using the point estimates of our baseline specification (see Table III). The MC curve is implied by the other two curves, as in equation (13). The circles represent the actual data points (see Table II, columns (3) and (4)) for demand (empty circles) and cost (filled circles). The size of each circle is proportional to the number of individuals associated with it. For readability we omit the one data point from Table II with only seven observations (although it is included in the estimation). We label points C, D, and E, which correspond to the theoretical analogs in Figure I, and report some important implied point estimates (of the equilibrium and efficient points, as well as the welfare cost of adverse selection).

Figure V also provides some useful information about the fit of our estimates, and where our pricing variation is relative to the key prices of interest for welfare analysis. The circles superimposed on the figure represent the actual data points (from Table II), with the size of each circle proportional to the number of individuals who faced that price. The fit of the cost curve appears quite good. The fit of the demand curve is also reasonable, although the scatter of data points led us to assess the sensitivity of the results to a concave demand curve, which is one of the exercises reported in the Online Appendix. The price range from $384 to $659 in our data brackets our estimate of the equilibrium price (point C) of $463. The lowest (and modal) price in our sample of $384 is about 45% higher than our estimate of the efficient price.
Results: welfare benchmarks

- Estimated demand and cost curves can also provide benchmarks to help provide context.
- Preferred benchmark:
  - Cost of price subsidy required to achieve efficient price – i.e. 
    \[ \lambda (P_{eq} - P_{eff}) Q_{eff} \] 
    - is about 5 times welfare gain from moving from adverse selection equilibrium to efficient price.
- Other benchmarks (much more out of sample)
  - Welfare cost of mandatory coverage by \( H \) is about 3 times equilibrium welfare cost of adverse selection
  - Welfare cost of adverse selection \( \sim 3\% \) of total surplus at stake from efficient pricing
Many possible applications

- Relatively little work estimating welfare costs of selection
- Medigap (contracts standardized)
- Medicare Part D - rich data on insurance choices and claims (pricing variation?)
- Any insurance product... auto, homeowners, life, long term care etc.!
- Landais et al. (2017) estimate welfare consequences of choice vs mandate for supplemental UI in Sweden
Many (better!) possible sources of pricing variation

- Firm experimentation (our case)
- Regulation - e.g. pricing changes discretely with income bins (Finkelstein et al. 2017)
- Tax policy
- Coarse pricing bins (e.g. price on age in bins)
- Market structure
Comment I: What about moral hazard?

- Welfare analysis takes moral hazard effects as given
- Government generally has no comparative advantage in combating moral hazard effects
  - Part of the “technology” that we take as given
- Analysis of welfare / policy under adverse selection should take moral hazard environment as given
- NB: enormous empirical literature estimating mh effects of social insurance programs
  - This speaks to optimal level of private or social insurance
Moral hazard (con’t)

- Since costs are a function of insurance coverage, useful to define
  \[ c^H \geq c^L \]
  - \( c^j \) is expected cost of insurance coverage \( H \) when behavior is as under \( j \) coverage
  - correspondingly two average cost curves \((AC^H \text{ and } AC^L)\) and two marginal cost curves \((MC^H \text{ and } MC^L)\)

- To explicitly recognize moral hazard in preceding analysis, replace \( c, AC, \text{ and } MC \) with superscript ”\( H \)”
  - Recall that cost curve estimated on sample of individuals who endogenously choose \( H \)
Moral hazard (con’td)

- Preceding welfare analysis goes through.
  - Note that the c we defined earlier is $c^H$ – i.e. the relevant cost curve is the actual costs of coverage given the moral hazard effect of coverage on expected costs)

- Intuition: Why doesn’t $c^L$ matter for analysis:
  - Firm: only behavior of insured individuals matters ($c^H$). How would behave if not insured ($c^L$) not relevant
  - Individual: gap between $c^H$ and $c^L$ does matter but incorporated into effect on WTP ($\pi$)
  - (Caveat: when $L$ is partial coverage, need to account for any “moral hazard externality”)
Comment II: What if insurance market not perfectly competitive?

- Assumed equilibrium was $P = AC$
  - But since empirical work requires out-of-equilibrium pricing variation, don’t actually observe equilibrium

- Could easily extend welfare analysis under a different specific assumption about competition
  - Mahoney and Weyl (2017) develop this formally
Interaction of market power (imperfect competition) with selection

Example: risk adjustment subsidies to plan (based on difference between average cost of enrollees and average cost in population)

This flattens AC curve (at population average)
  - Under perfect competition, lowers average costs and creates higher Q, lower P equilibrium
  - Under imperfect competition, recall firms set price too high relative to social optimum. Adverse selection reduces incentives to mark up prices (because get worse risk pool / higher costs). Risk adjustment, by offsetting adverse selection, undermines this incentive and may lead to higher P, lower Q

Example of the theory of the second best
  - aka full employment program for empirical public finance economists
Discussion: Attractions

- Model demand and costs but not their primitives \((\zeta_i)\). Don’t have to take a stand on structure / nature of private information or preferences etc.
- Extremely simple to implement
  - Relatedly: transparent. Will see direct mapping from model to data. Makes it easier to see the key empirical assumptions.
- In principle broadly applicable.
  - Data requirements are
    - Demand and cost (as required for pos correlation test)
    - Pricing variation. = key hurdle. But many potential sources
  - Results likely relatively comparable across markets (vs more structural models where model tailored to market)
  - Caveat: settings where fixed contract assmpt seems reasonable
- Bonus: direct test of selection (shape of cost curve)
  - In one package: detect selection and examine welfare cost
Discussion: Limitations

- Requires good price variation – not always easy to find (but see many possibilities...!)
- Fixed contracts assumption
  - Cannot evaluate welfare from introducing contracts not observed
    - Requires underlying structural primitives (as in EFS EMA 2010)
    - Welfare analysis limited to policies that change price of existing contracts (mandates; subsidies; restrictions on pricing)
  - Limited to “local” welfare analysis for relatively small price changes if concerned about endogenous contract respond
- Familiar tradeoff
  - Produce-space (e.g. Almost Ideal Demand System) vs Characteristic space (e.g. BLP) approaches to differentiated demand estimation. Latter can be used to evaluate welfare from new goods before introduced.
Limitations (con’t)

- Ignores income effects
  - Using Marshallian vs. Hicksian demand
    - can produce very misleading welfare estimates (Hausman 1981)
    - If CARA not right model or approximation.
  - Again suggests more useful for local (small) policy changes
Discussion: key assumptions of framework

- "Valid" pricing variation
- Ignoring income effects
- Revealed preference
  - Or at least a particular behavioral model
- Fixed contracts
  - Estimating inefficiency selection causes via mispricing
    - Not capturing welfare cost of adverse selection from distortion of contract space (Rothschild-Stiglitz 1976)
  - Policy analysis limited to changes in prices of existing contract space
    - Preferable "small" price changes that don’t expect to trigger endogenous contract response
Sufficient Statistics for Welfare Analysis (Chetty 2009)

Primitives | Sufficient Stats. | Welfare Change
---|---|---
ω₁ | β₁(t) | \( \frac{dW}{dt} (t) \)
ω₂ | β₂(t) |
... | ... |
ωₙ | ... |

ω = preferences, constraints

β = f(ω, t)  
y = β₁X₁ + β₂X₂ + ε

dW/dt used for policy analysis

ω not uniquely identified  

β identified using program evaluation
“Sufficient statistics”

- Sufficient given a model (e.g. fixed contracts etc)
- Advantages:
  - Simplicity and transparency.
  - Ideally direct mapping from theory to empirics
    - E.g. EFC: basically just a way of transforming the data (See graph)
    - Allows for informed discussion / critique of identification, in sample fit, how far out of sample we are going etc
- Shortcoming:
  - Mostly useful for local welfare analysis
  - Have estimated behavioral elasticities that are valid locally
    - More limited set of counterfactuals
  - Important to remember: sufficient given the model
Recover underlying structural primitives (preferences and risk type)
- Use insurance company data on individual insurance choices and risk experience (claims) + modeling assumptions to recover joint distribution of (unobserved) risk type and preferences

After that, it's simple
- If have a utility based model and have estimated the parameters of it (risk type and preference) welfare analysis is easy
- Can compute welfare at observed equilibrium
- Can compare to welfare in counterfactual equilibriums
  - First best (symmetric information). Gives welfare cost of adverse selection.
  - Mandatory social insurance. Gives welfare gain / loss from a particular government intervention.

So the focus is on how we recover these parameters and what assumptions we needed to make
Why would you want to do this?

- Don’t have good pricing variation
  - Substitute structure / modeling assumptions for pricing variation
- Interested in primitives
  - E.g. recover joint distribution of risk type and risk aversion (Cohen and Einav 2007 AER). May be interested in risk aversion (average, dispersion, correlates of dispersion…)
- Want to say something about welfare from contracts not observed in the data
  - Although hopefully not too far out of sample
This paper represents an attempt to uncover several structural parameters from data on insurance claims and choices.

This basic endeavor will re-appear (in similar or different guises) in a number of other papers on insurance we’ll discuss.

Important to understand where identification comes from:
- What is in the data
- What are the key assumptions

Compare when e.g. get to “behavioral” models of insurance demand in a next lecture topic...
Setting: Semi-compulsory UK annuity market

- Individuals w/ tax preferred retirement savings required to annuitize their accumulated balance at retirement
  - 6 billion pounds in new funds annuitized in 1998 (vs. voluntary mkt)
- Annuities are survival contingent streams of payment
  - Theoretically large welfare gains.
    - Consider a retiree w/ lump sum accumulated assets facing stochastic mortality. Annuity enables him to consume more each period (vs. saving to insure against long life w low consumption at end)
- Puzzle: small voluntary annuity markets
- Important in Social Security reform discussions (will explain)
Setting (con’t)

- Semi-compulsory UK market:
  - Required to annuitize tax preferred savings
  - Choice of annuity contract: 0, 5 or 10 year guarantee.
    - During guarantee period, annuity payments are unconditional on survival
    - Guarantees trade off reduced payment per period you are alive for payments regardless of survival during guarantee
    - Choice of guarantee likely driven by private information about risk type + preference for “wealth after death”

- Attractions of setting
  - Relatively simply contracts (0, 5, or 10 year guarantee)
  - Prior evidence of asymmetric information in this market (Finkelstein and Poterba JPE 2004)
  - Moral hazard likely to be less important than in other insurance markets (attractive for estimation and identification)
  - Important market; implications for Social Security reform
Interlude: What are annuities and why are they so important?

- Defined Benefit Social Security system
  - Most Social Security systems (including US and UK) collect payroll taxes on current workers and pay benefits to current retirees as an annuity
- One key element of potential social security reform proposals: individuals accumulating their own individual funds
  - Would they be required to annuitize some / all?
  - Choice in annuitization?
- One potential rationale for Social Security is to address adverse selection in voluntary annuity markets
  - Others include forced savings (paternalism) and redistribution
  - Will discuss much more in Section VIII
References

Seminal reference: Yaari (1965) shows full annuitization is optimal
Davidoff, Diamond and Brown (AER 2005) generalize result

Basic idea: Life cycle consumer retirees with lump sum of wealth; faces stochastic mortality

How to consume in retirement?
Consume too much and live a long time $\rightarrow$ end up with little consumption
Consume too little and die early $\rightarrow$ forewent a lot of consumption
Annuities provide survival contingent stream that allows for higher consumption in all living states
Simple two period example of welfare gains from annuities

- Consumer with $U(c_1, c_2)$ alive in period 1; alive in period 2 with probability $1 - q$

- Assume two securities are available:
  - Bond returns $R_B$ units of consumption in period 2, whether or not consumer is alive, per unit of consumption in period 1
  - Annuity returns $R_A$ in period 2 if alive, 0 otherwise

- Actuarially fair annuity: $R_A = \frac{R_B}{1-q}$
  - $R_A > R_B$
Welfare gains from annuities (con’t)

- Consider consumer optimization problem via its dual (minimizing expenditure st attaining at least a given level of utility)
- Denote by $A$ savings in form of annuity, and by $B$ savings in form of bond
- Assume no other period 2 income (retirement). Therefore
  - $c_2 = R_A A + R_B B$
  - $E = c_1 + A + B$
- Expenditure minimization problem:
  \[
  \min_{c_1, A, B} c_1 + A + B \quad \text{s.t.} \quad U(c_1, R_A A + R_B B) \geq U_{\text{bar}}
  \]
- Also impose: $B \geq 0$ (cannot die in debt; otherwise with $R_A > R_B$ purchasing annuities and selling bonds in equal numbers would cost nothing and yield positive consumption when alive in period 1 but leave debt if dead, leaving lenders with expected financial losses).
Welfare gains from annuities (con’t)

- Full annuitization optimal: If $B > 0$, can reduce expenditures while holding consumption vector fixed by selling $R_A / R_B$ of the bond and purchasing one unit of annuity (noting $R_A > R_B$).
  - Solution is $B = 0$ (fully annuitize).

- Intuition: allowing individuals to substitute annuities for conventional assets yields an arbitrage-like gain when the individual places no value on wealth when not alive.
  - NB this result does not require annuities to be actuarially fair. Does require no bequest motive + $R_A > R_B$ (latter does not have to be true due to transaction costs and adverse selection but empirically appears to be).
Using above type logic, show that Yaari (1965) result on optimality of full annuitization is quite general.

- Key requirements when markets are complete are that consumers have no bequest motive and rate of return on annuities above bond (but don’t have to be actuarially fair)

- Calibration results adding things unfavorable to annuities (like incomplete markets and bequests and existence of SS) still suggest a fair amount of annuitization (although not full) should be optimal

- They conclude that need psychological / behavioral considerations to explain lack of annuity purchases
We now all understand how fascinating and important annuities are and how they interact with Social Security reforms such as allowing choice on annuitization margin.

We return to our regularly scheduled program: estimating welfare cost of adverse selection and welfare consequences of mandates in annuity markets.
Goal: recover joint distribution of unobserved preferences and risk type

Observe:
- Menu of guarantee choices (payouts as function of guarantee — by age and gender). (see next slide)
- Annuitants’ choice of guarantee
- Subsequent date of death if any (= "risk type" of annuitant)

Why buy guarantee?
- Guarantee trades of lower annuity payout while alive but continued payments in event of death during guarantee
- Longer guarantee is more attractive (at a given price) to someone who:
  - Is more likely to die sooner (adverse selection) than their risk category (age/gender) is on average
  - Has higher value for “wealth after death”
Table 3: Annuity payment rates

<table>
<thead>
<tr>
<th>Guarantee Length</th>
<th>60 Females</th>
<th>65 Females</th>
<th>60 Males</th>
<th>65 Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1078</td>
<td>0.1172</td>
<td>0.1201</td>
<td>0.1330</td>
</tr>
<tr>
<td>5</td>
<td>0.1070</td>
<td>0.1155</td>
<td>0.1178</td>
<td>0.1287</td>
</tr>
<tr>
<td>10</td>
<td>0.1049</td>
<td>0.1115</td>
<td>0.1127</td>
<td>0.1198</td>
</tr>
</tbody>
</table>

These are the rates from January 1992, which we use in our baseline specification. A rate is per pound annuitized. For example, a 60 year old female who annuitized X pounds and chose a 0 year guarantee will receive a nominal payment of 0.1078X every year until she dies.
Guarantee Choice Model

- Standard annuity framework:
  - Fully rational, forward looking, risk averse retirees
  - Retirees with stock of wealth face stochastic mortality parameterized by $\alpha_i$
  - Time separable CRRA utility

$$U(\{c_t, w_t\}_{t=0}^T) = \sum_{t=0}^T \delta^t (s_t(\alpha_i)u(c_t) + \beta_i f_t(\alpha_i)b(w_t))$$

- Heterogeneity in
  - risk type, $\alpha_i$ – mortality rate
  - preferences, $\beta_i$ – weight placed on wealth at death

- Given $\alpha_i$, $\beta_i$, individual chooses annuity contract that maximizes lifetime utility (given optimal consumption path)
  - Optimal guarantee length increases with mortality ($\alpha_i$) and preference for wealth after death ($\beta_i$)
Additional Assumptions

- Gompertz survival function with shape parameter $\lambda$ and shift parameter $\alpha$
  - Individual hazard rate as function of age $(t)$ given by $\psi^i(t) = \alpha_i e^{\lambda t}$
- $\alpha$ and $\beta$ are joint lognormally distributed
- CRRA utility function for both $u(c)$ and $b(w)$ with same coefficient of relative risk aversion
  - implies that the optimal guarantee length does not depend on initial wealth (which we do not observe)
- $\gamma = 3$
- fraction of wealth annuitized $= 0.2$
Some comments on model

- We are agnostic about structural interpretation of $\beta$ (bequests? ex ante regret? etc.)
- Relatedly, note that $\beta$ is not separately identified from risk aversion ($\gamma$), discount rate ($\delta$), etc. except by functional form.
- Perform several robustness tests to make sure that our calibrated values for other parameters is not what drives the welfare estimates.
- Baseline model assumes all preference heterogeneity is over wealth after death
  - Allowing greater heterogeneity in $\beta$ is similar to allowing heterogeneity in other preference parameters
  - Also try alternative model in which allowing for heterogeneity in other parameter (e.g. $\gamma$), rather than $\beta$
Intuition for identification

- Joint distribution of risk type and preferences identified from relationship between mortality and guarantee choice in the data
- Key idea: ex-post mortality realization identifies risk type, so guarantee choice can be used to identify preference heterogeneity and correlation with risk
- Intuition most clearly seen in two steps (estimated jointly in practice):
  1. Individual’s (ex-post) mortality experience provides information on her (ex-ante) mortality rate
     - Individual who dies sooner more likely to have had a higher (ex-ante) mortality rate
     - Key assumption of no moral hazard (mortality not a function of guarantee choice).
  2. Conditional on individual’s mortality rate, individual’s guarantee choice provides information on preferences and how they correlate with observed mortality
Some key assumptions (compare to later “behavioral” papers)

- Nature of ex ante information about risk type
  - We assume perfect information about mortality type (individuals know their own $\alpha$)

- Identifying private information about mortality requires modeling assumptions
  - Although not for existence. See conditional correlation between guarantee and mortality (e.g. Finkelstein and Poterba JPE 2004)
  - Assumed mixed proportional hazard model: $\psi^i(t) = \alpha_i e^{\lambda t}$ Imagine graph of log hazard mortality rate wrt age
    - Gompertz $\Rightarrow$ absent heterogeneity log hazard is linear in age with slope $\lambda$.
    - Heterogeneity in mortality identified by concavity of relationship between log hazard and age (over time lower mortality individuals are more likely to survive).
    - Level of graph pins down estimate of $\mu_\alpha$, average slope affects estimate of $\lambda$, and concavity affects estimate of $\sigma_\alpha$ (key parameter).
Some key assumptions (con’t)

- Identify preference heterogeneity from guarantee choice and its relationship with mortality
  - Use preference heterogeneity to rationalize choices
- Could make other assumptions (and show robustness to in paper) – e.g. different information set or different functional form for baseline hazard
  - Key is need some assumptions.
Estimation

- Estimate (by ML) \( \lambda \) using mortality data
- Calculate cutoff given \( \lambda \) using guarantee choice model (essentially no data yet)
- Estimate (by ML) distribution of \( \alpha \) and \( \beta \) using cutoffs, guarantee data, and mortality data

![Schematic indifference sets](image-url)
From one of the five largest annuity providers in the U.K.

Data on guarantee choices, age, gender, and subsequent mortality experience

All annuities purchased between January 1, 1988 and December 31, 1994 that were still active as of January 1, 1998

- Mortality experience through December 31, 2005

Limit analysis to:

- Single-life annuities
- Age at purchase of 60 or 65
- Accumulated funds within the company
- Nominal annuities
Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>60 Females</th>
<th>65 Females</th>
<th>60 Males</th>
<th>65 Males</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>1800</td>
<td>651</td>
<td>1444</td>
<td>5469</td>
<td>9364</td>
</tr>
<tr>
<td>Fraction choosing 0-year guarantee</td>
<td>14.0</td>
<td>16.0</td>
<td>15.3</td>
<td>7.0</td>
<td>10.2</td>
</tr>
<tr>
<td>Fraction choosing 5-year guarantee</td>
<td>83.9</td>
<td>82.0</td>
<td>78.7</td>
<td>90.0</td>
<td>86.5</td>
</tr>
<tr>
<td>Fraction choosing 10-year guarantee</td>
<td>2.1</td>
<td>2.0</td>
<td>6.0</td>
<td>3.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Fraction who die within observed mortality period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire sample</td>
<td>8.4</td>
<td>12.3</td>
<td>17.0</td>
<td>25.6</td>
<td>20.0</td>
</tr>
<tr>
<td>Among those choosing 0-year guarantee</td>
<td>6.7</td>
<td>7.7</td>
<td>17.7</td>
<td>22.8</td>
<td>15.7</td>
</tr>
<tr>
<td>Among those choosing 5-year guarantee</td>
<td>8.7</td>
<td>13.3</td>
<td>17.0</td>
<td>25.9</td>
<td>20.6</td>
</tr>
<tr>
<td>Among those choosing 10-year guarantee</td>
<td>8.1</td>
<td>7.7</td>
<td>16.1</td>
<td>22.9</td>
<td>18.5</td>
</tr>
</tbody>
</table>

- 5 year guarantee is by far the most common
- Individuals choosing 5 year guarantee have higher mortality than 0 guarantee; no clear pattern for 10 year guarantee (presumably due to smaller sample size)
Annuity Pricing

- Linear prices: price is quoted as an annual annuity payout rate for each pound annuitized.
- Rates at a given point in time only depend on (observed) guarantee, age, and gender.
- Ignore temporal variation and just use payment, interest, and inflation rates from January 1992:

<table>
<thead>
<tr>
<th>Guarantee Length</th>
<th>60 Females</th>
<th>65 Females</th>
<th>60 Males</th>
<th>65 Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1078</td>
<td>0.1172</td>
<td>0.1201</td>
<td>0.1330</td>
</tr>
<tr>
<td>5</td>
<td>0.1070</td>
<td>0.1155</td>
<td>0.1178</td>
<td>0.1287</td>
</tr>
<tr>
<td>10</td>
<td>0.1049</td>
<td>0.1115</td>
<td>0.1127</td>
<td>0.1198</td>
</tr>
</tbody>
</table>
### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_\alpha$</td>
<td>60 Females</td>
<td>$-5.76$</td>
</tr>
<tr>
<td></td>
<td>65 Females</td>
<td>$-5.68$</td>
</tr>
<tr>
<td></td>
<td>60 Males</td>
<td>$-4.74$</td>
</tr>
<tr>
<td></td>
<td>65 Males</td>
<td>$-5.01$</td>
</tr>
<tr>
<td>$\sigma_\alpha$</td>
<td></td>
<td>0.054</td>
</tr>
<tr>
<td>$\lambda$</td>
<td></td>
<td>0.110</td>
</tr>
<tr>
<td>$\mu_\beta$</td>
<td>60 Females</td>
<td>9.77</td>
</tr>
<tr>
<td></td>
<td>65 Females</td>
<td>9.65</td>
</tr>
<tr>
<td></td>
<td>60 Males</td>
<td>9.42</td>
</tr>
<tr>
<td></td>
<td>65 Males</td>
<td>9.87</td>
</tr>
<tr>
<td>$\sigma_\beta$</td>
<td></td>
<td>0.099</td>
</tr>
<tr>
<td>$\rho$</td>
<td></td>
<td>0.881</td>
</tr>
<tr>
<td>No. of obs.</td>
<td></td>
<td>9364</td>
</tr>
</tbody>
</table>
FIGURE 2.—Estimated distributions. The estimated indifference sets for each age–gender cell, with a scatter plots from the estimated joint distribution of $(\log \alpha/\text{ori}, \log \beta)$ superimposed; each point is a random draw from the estimated distribution in the baseline specification. The estimated indifference sets for 65-year-old males are given by the pair of dark dashed lines, for 60-year-old males by the pair of lighter dashed lines, for 65-year-old females by the pair of dotted lines, and for 60-year-old females by the pair of solid lines. The estimated indifference sets for 65-year-old males are the same as those shown in Figure 1 (but a close up and in log scale).

The results indicate that both mortality and preference heterogeneity are important determinants of guarantee choice. This is similar to recent findings in other insurance markets that preference heterogeneity can be as or more important than private information about risk in explaining insurance purchases (Finkelstein and McGarry (2006), Cohen and Einav (2007), Fang, Keane, and Silverman (2008)). As discussed, we refrain from placing a structural interpretation on the $\beta$ parameter, merely noting that a higher $\beta$ reflects a larger preference for wealth after death relative to consumption while alive. Nonetheless, our finding of heterogeneity in $\beta$ is consistent with other estimates of heterogeneity in the population in preferences for leaving a bequest (Laitner and Juster (1996), Kopczuk and Lupton (2007)).
Within sample fit:
- Fit guarantee choice proportions nearly perfectly
- Match unconditional probability of dying during the sample period very well
- Do not reproduce non-monotone relationship between guarantee choice and mortality

Out of sample fit:
- Life expectancies slightly higher than a proxy for market average (but also true within sample)
Parameter estimates allow us to calculate welfare in observed equilibrium and compare to two counterfactuals:

- Pick two counterfactuals:
  - Symmetric information (first best)
  - mandatory social insurance program (no choice over guarantee)

Choice of counterfactuals (important art)

- Limited to policies where equilibrium is easy to solve for (vs. e.g. subsidies where have to solve for fixed point...)
- Don’t want to go too far out of sample
Measuring Welfare

- Quantify welfare in terms of wealth-equivalents ($weq$):
  - The $weq$ is wealth a person would need to have without the annuity to reach same utility as achieves with initial wealth and annuity contract chosen
  - Recall we use 100 for initial wealth, and 20% annuitized
  - Higher $weq \Rightarrow$ higher welfare, $weq < 100 \Rightarrow$ prefer not to annuitize

- Compare average $weq$ under observed equilibrium and each counterfactual
  - Convert difference to annual pounds using amount annuitized in 1998 (£6 billion)
### Welfare Estimates

<table>
<thead>
<tr>
<th></th>
<th>60 Females</th>
<th>65 Females</th>
<th>60 Males</th>
<th>65 Males</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed equilibrium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wealth equivalent</td>
<td>100.24</td>
<td>100.40</td>
<td>99.92</td>
<td>100.17</td>
<td>100.16</td>
</tr>
<tr>
<td>Maximum money at stake (MMS)</td>
<td>0.56</td>
<td>1.02</td>
<td>1.32</td>
<td>2.20</td>
<td>1.67</td>
</tr>
<tr>
<td><strong>Symmetric information counterfactual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wealth equivalent</td>
<td>100.38</td>
<td>100.64</td>
<td>100.19</td>
<td>100.74</td>
<td>100.58</td>
</tr>
<tr>
<td>Absolute welfare difference (M pounds)</td>
<td>43.7</td>
<td>72.0</td>
<td>82.1</td>
<td>169.8</td>
<td>126.5</td>
</tr>
<tr>
<td>Relative welfare difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as a fraction of MMS)</td>
<td>0.26</td>
<td>0.23</td>
<td>0.21</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Mandate 0-year guarantee counterfactual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wealth equivalent</td>
<td>100.14</td>
<td>100.22</td>
<td>99.67</td>
<td>99.69</td>
<td>99.81</td>
</tr>
<tr>
<td>Absolute welfare difference (M pounds)</td>
<td>−30.1</td>
<td>−53.2</td>
<td>−73.7</td>
<td>−146.1</td>
<td>−107.3</td>
</tr>
<tr>
<td>Relative welfare difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as a fraction of MMS)</td>
<td>−0.18</td>
<td>−0.17</td>
<td>−0.19</td>
<td>−0.22</td>
<td>−0.21</td>
</tr>
<tr>
<td><strong>Mandate 5-year guarantee counterfactual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wealth equivalent</td>
<td>100.25</td>
<td>100.42</td>
<td>99.92</td>
<td>100.18</td>
<td>100.17</td>
</tr>
<tr>
<td>Absolute welfare difference (M pounds)</td>
<td>2.8</td>
<td>6.0</td>
<td>1.7</td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Relative welfare difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as a fraction of MMS)</td>
<td>0.02</td>
<td>0.02</td>
<td>0.004</td>
<td>0.002</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Mandate 10-year guarantee counterfactual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wealth equivalent</td>
<td>100.38</td>
<td>100.64</td>
<td>100.19</td>
<td>100.74</td>
<td>100.58</td>
</tr>
<tr>
<td>Absolute welfare difference (M pounds)</td>
<td>43.7</td>
<td>72.1</td>
<td>82.3</td>
<td>170.0</td>
<td>126.7</td>
</tr>
<tr>
<td>Relative welfare difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as a fraction of MMS)</td>
<td>0.26</td>
<td>0.23</td>
<td>0.21</td>
<td>0.26</td>
<td>0.25</td>
</tr>
</tbody>
</table>

---

15Our average wealth equivalent is noticeably lower than what has been calculated in the previous literature (Mitchell, Poterba, Warshawsky, and Brown (1999), Davidoff, Brown, and Diaz (2001)).
Summary of Results

Symmetric Information (first best):
- Average welfare loss due to asymmetric information = £127 million annually (2% of premiums)
- Welfare loss is due to distortion in choices: under symmetric information, all individuals choose 10 year guarantee

Government Mandates:
- Mandate can increase welfare by £127 million or decrease by £107 million depending on which contract is mandated
- Not ex-ante obvious that 10 year guarantee would be optimal mandate (rarely chosen in equilibrium)
With an estimated model of utility the sky is the limit

- Welfare cost of asym information relative to symmetric
  - What is optimal (first best) allocation?

- Welfare consequences of policies that change equilibrium allocations. Including offering policies not observed in data.
  - e.g. welfare benefits of offering 20 year guarantees (not currently allowed)
  - Welfare consequences of the compulsory annuitization requirements
  - Do we want to go that far out of sample?
Discussion: limitations

- Key challenge: estimating distribution of risk type ($\alpha$) and preferences ($\beta$)
  - Requires estimating ex ante information about risk type.
  - To get from risk realization to information requires assumptions.
    - two people w/ same death date choosing different guarantee - because of different preferences or because of different information about risk type but lower mortality person had a bad epsilon
- Without assumptions can rationalize data w very different underlying primitives
  - Fundamentally risk preferences and private information about risk type separately identified by functional form
    - Model of risk realization: Assumption that individuals have perfect information about their mortality type and that mortality risk takes the form of a gompertz mixed proportional hazard model
    - Model of choices: Guarantee choice model w all its assumptions
  - Can explore sensitivity to alternative models (including "behavioral" ones) but can't get away from modeling
Empirical welfare analysis: road map

Thus far: two approaches to empirical welfare analysis
- More vs. less structure

Up next: Exploring a key feature of both approaches: both rely on observing demand and taking a revealed preference approach
- What if we want to abandon revealed preference / “go behavioral”? 
- What if market doesn’t exist / has completely unraveled. How do we recover preferences / estimate demand?
Motivation: Small estimated welfare costs of adverse selection

- **EFC (2010):** Welfare cost from inefficient pricing of low deductible health insurance plan in Alcoa: ~3 percent of surplus at stake from efficient pricing.
- **EFS (2010):** Welfare cost of adverse selection along guarantee margin in semi-compulsory UK annuity market ~2 percent of annuitized wealth.
- Several other studies using different methodologies, but all asking about welfare cost of pricing distortion induced by adverse selection in health insurance.
  - All tend to find modest welfare costs of under-insurance from pricing distortions due to adverse selection.
Interpretation?

- Adverse selection not a big deal
  - At least given current policy environment.
  - That doesn’t mean couldn’t design policies that on the margin would create huge adverse selection
- And/or something is missing from the approach (= Next two topics)
  - Can we use observed demand to infer value of insurance?
  - Lampost problem: studying relatively small margins of contract choice in markets that exist.
    - But see Finkelstein, Hendren and Shepard (2017) on extensive margin insurance choice (will study in Section VID)
    - What about welfare costs from complete unraveling of market (ultimate distortion of contract space)?