14.472 Public Finance II

Topic II_b & c: Adverse selection: Testing and Welfare I

Amy Finkelstein

Fall 2022

Topic II: Empirical analysis of selection

- Theory: adverse selection can impair efficient operation of insurance markets and create scope for welfare improving government intervention
- Lays out an empirical agenda:
  - Does selection exist in a particular market
  - What are the efficiency costs of any detected selection
  - What are the welfare costs of alternative potential government interventions?

Lecture(s) road map

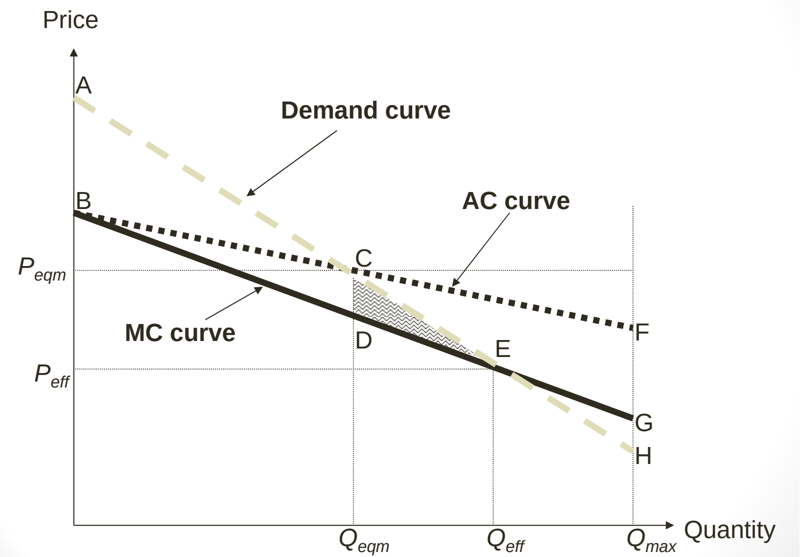
- Testing for selection
- Empirical welfare analysis I: Using data on choices and claims
- Empirical welfare analysis II: What happens when you can't use choice data
    - Don't trust revealed preference
    - Markets don't exist

Testing for Selection

Some perspective

- Seminal theoretical work on adverse selection in 1970s
    - Very influential in policy
    - 2001 Nobel prize
- Empirical testing only emerges several decades later!

Adverse selection: downward sloping marginal cost curve



Positive correlation test

- Testing for adverse selection essentially requires testing whether MC curve downward sloping
- Making inferences about marginal individuals can be difficult
- Initial work focused on testing equilibrium prediction of "positive correlation" (Chiappori and Salanie 2000)
  - Reject null of symmetric information if there is a positive correlation between insurance coverage and ex-post risk occurrence
  - Are average costs of insured higher than average costs of uninsured?
  - At any given price, and in particular at the equilibrium price, adverse selection implies that average cost of insured individuals is higher than average costs of uninsured individuals

Example: Annuitant vs. Population Mortality (US 2007)

	Annuitant Mortality		Population Mortality	
	Men	Women	Men	Women
65	1.02%	0.57%	1.72%	1.16%
75	2.98	1.61	4.29	2.98
85	8.06	5.08	11.35	8.54

Limitations to positive correlation test

1. Need to observe (customized) prices for products in the choice set
2. How to interpret a null result:
   - In the presence of multiple dimensions of heterogeneity, inability to detect a positive correlation does not preclude presence of selection
   - If risk aversion is negatively correlated with risk, lower-risk individuals can purchase more insurance in equilibrium even when there is private information about risk type (e.g. Finkelstein and McGarry 2006; Fang et al. 2008)
3. How to identify selection separately from a direct impact of contract on risk ("moral hazard")

Moral hazard also generates positive correlation

- Adverse selection: those with private information they are high expected cost self select into insurance market
- Moral hazard: individuals identical before purchasing insurance; those with greater coverage have less incentive to take actions to reduce their expected costs ex post
- Different policy implications

Distinguishing selection from moral hazard

- Key point: need exogenous variation in contracts
  - Basic problem: distinguishing treatment (moral hazard) from selection (selection!)
- Variety of sources of variation
  - Quasi-Exogenous variation (examples)
    - border design / geographic variation in health insurance premiums (Panhans 2019)
    - policy reforms with differential effects on flood insurance prices by house type (Wagner 2021)
  - Randomized experiments (examples)
    - health insurance in rural Pakistan (Fischer et al. 2018)
    - consumer loans in South Africa (Karlan and Zinman 2009)

Karlan and Zinman 2009

- Setting: Consumer lender (South Africa)
- Randomized offer interest rate and contract rate on loan
- Selection: compare repayment rate of those offered different rates (but receiving same rate)
- MH: compare repayment rates of those responding to same high offer rate but facing different contract rates

	High Contract Rate	Low Contract Rate
High Offer Rate	Moral Hazard	Adverse Selection
Low Offer Rate	N/A	

Cost curve test

- Einav, Finkelstein and Cullen 2010
- Slope of MC curve provides a direct test of existence and nature of selection
    - Reject null of no selection if reject null of constant MC curve
    - Slope of cost curve indicates if selection is adverse or advantageous
- Implementation: estimate average cost curve on sample who (endogenously) purchase the same contract:

$$c_i = \gamma + \delta p_i + \epsilon_i$$

  where c_i is insurable costs (claims), and p_i
- Estimating how costs change for the endogenously selected sample of those who stay insured as price (exogenously) varies
    - data requirements of positive correlation test plus variation in prices that is exogenous to demand

Aside: Selecting on endogenous outcome

- Useful if you want to understand the characteristics of those who respond to the intervention
- Other examples:
  - What type of DI applicants deterred from hassles (Deshpande and Li 2019)?
  - Who is the marginal child when abortion is legalized (Gruber, Levine and Staiger 1999)?
- More generally: "characterizing the compliers" (Abadie 2002).

Example: Colorado Health Insurance Exchange

- Panhans (2019)
- Colorado Health Insurance Exchange 2014
  - Created by Affordable Care Act (ACA)
  - Subsidized for low income individuals
- Statewide data on premiums, claims, insurance coverage (exchange coverage vs. not) .
- Source of premium variation: geographic discontinuities in insurance premiums at boundaries of "rating areas" established by law
  - Premiums change discretely at "artificial" boundaries of rating areas
  - Compare costs of those enrolled on either side of the border (fixed effect for each zip code pair $\varphi_{g(k)}$)

$$c_i = \gamma + \delta p_{ik} + \varphi_{g(k)} + \epsilon_i$$

Border Design

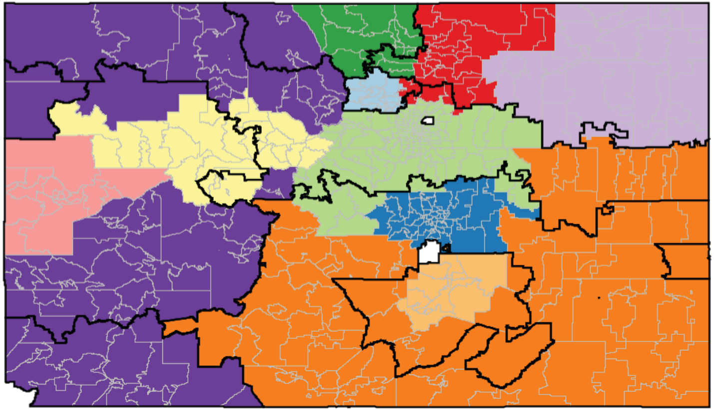


FIGURE 2. 2014 RATING AREAS IN COLORADO

Notes: Five-digit zip codes are shown grouped into rating areas based on color. The outlines designate the grouping of zip codes into medical markets, here defined as the Hospital Referral Region (HRR).

Border-induced premium variation

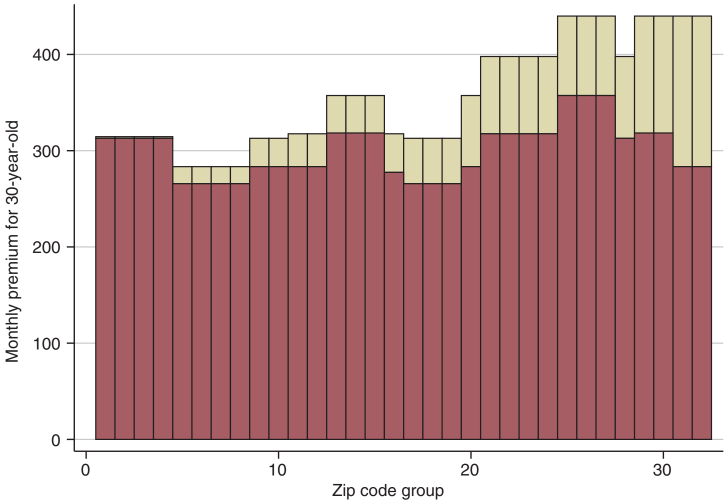


FIGURE 3. CHANGE IN PREMIUM ACROSS RATING AREA BOUNDARY

Cost curve indicates selection

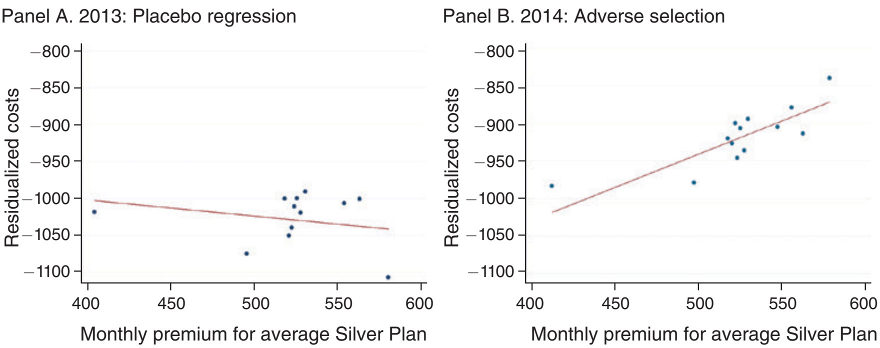


Figure 5. Binned Scatterplot of Selection Regression

Notes: Panel A presents graphically the results from the placebo regression in column 1 of Table 4. Panel B presents the results from the main OLS results in panel A, column 1 of Table 3, which indicate adverse selection. The sample means of premiums have been added back in to the premium residuals before plotting.

Other consequences of selection

- Most existing work looks at impact of adverse selection on (mis-) pricing and insurance coverage
- Selection may also give insurers incentives to distort plan benefits (Rothschild-Stiglitz 1976)
- Relatively little existing work looking at impact of selection on contract / benefit design
    - Cost-sharing to discourage high cost enrollees
        - e.g. high cost-sharing for HIV drugs in Part D when new drug development makes existing diagnosis-based risk adjustment inadequate (Carey 2017 AEJ:Policy)
    - Bundling gym memberships with health insurance plans attracts lower cost enrollees (Cooper and Trivedi 2012 NEJM)
    - Broader networks attract higher cost enrollees (Shepard, AER 2022)

Some open testing questions

- Impact of selection on contract design
- Many markets where selection has had relatively little examination
  - e.g. Disability insurance. There's a lot of public policy (and research on that public policy) but not on potential underlying market failure
- Why don't insurance companies price on more observable characteristics?

Empirical Welfare Analysis I: Choice Data

Empirical welfare analysis I

- 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) and Einav, Finkelstein and Mahoney (IO Handbook 2021)

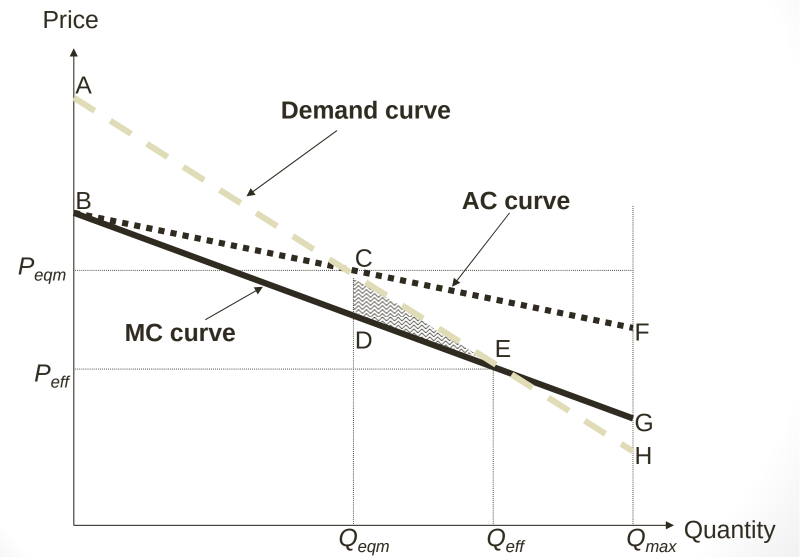
Welfare analysis: emphasized in PF

- 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 equbm?
  - 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)

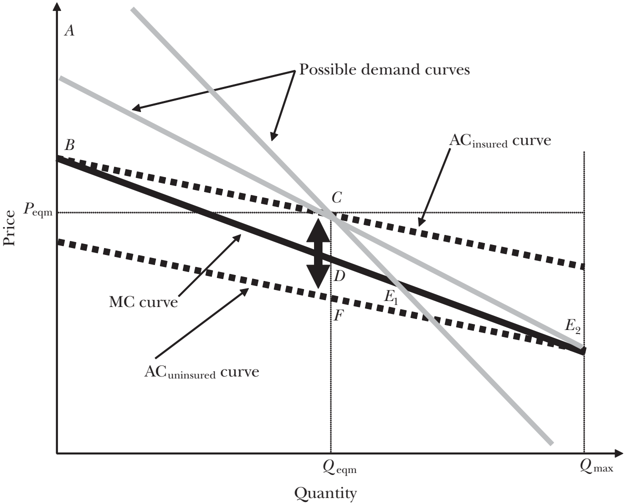
Welfare an empirical question



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?



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 (sufficient statistics)?
    - 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
  - ζ_i includes preferences, information set (i.e. expected claims) etc.
  - ζ_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 ζ

Setup and notation (con't)

- $\pi(\zeta_i)$ is willingness to pay for H (i.e., $v_H(\zeta_i, \pi(\zeta_i)) = v_L(\zeta_i)$)
- c (ζ_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)

Theory: Demand, Supply, and Equilibrium

- Demand:

$$D(p) = \Pr\left(\pi(\zeta_i) \geq p\right)$$

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

$$AC(p) = E\left(c(\zeta) | \pi(\zeta) \geq p\right)$$

  - Marginal cost (MC):

$$MC(p) = E\left(c(\zeta) | \pi(\zeta) = p\right)$$

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

$$p^* = \min\left\{p : p = AC(p)\right\}$$

Welfare definitions

- Total surplus from allocating H to individual i is

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

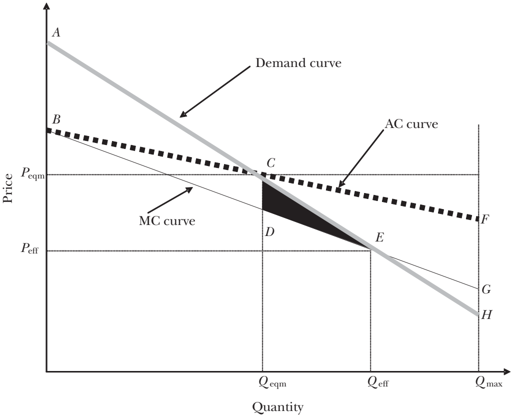
- First best allocation: individual i purchases insurance if and only if

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

- 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

$$\pi(\zeta_i) \geq E(c(\widetilde{\zeta})|\pi(\widetilde{\zeta}) = \pi(\zeta_i))$$

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...

Sufficient statistics: demand and cost curve

- 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
- 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?

Estimation

- Sufficient statistics for welfare analysis are:
  - the demand curve $D(p)$
  - the average cost curve $AC(p)$
- Estimation:

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

- 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-parametrically. In practice, likely that need to make functional form assumptions.

Estimation (con't)

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

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

- Conceptually, variation in p identifies all curves non-parameterically. 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

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
  - Recall: this speaks to optimal level of private or social insurance

Moral hazard

- 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 and AC^L) and two marginal cost curves (MC^H and MC^L)
- To explicitly recognize moral hazard in preceding analysis, replace c, AC, and MC with superscript "H"
    - Recall that cost curve estimated on sample of individuals who endogenously choose H
- Estimate a second pair of cost curves AC^L and MC^L. Difference between MC^H and MC^L measures moral hazard for each customer time.

- What would we do if we wanted to actually estimate moral hazard in an application?
- Regress costs on insurance coverage
    - Instrument for insurance coverage using exogenous variation in prices

Moral hazard: implications for welfare analysis of selection

- 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 (π)
  - (Caveat: when L is partial coverage, need to account for any "moral hazard externality". see e.g. Cabral and Mahoney (AEJ 2019))

Final comment on moral hazard

- What if what creates the downward slope of the cost curve is heterogeneous moral hazard?
  - i.e. those who have high WTP for insurance have higher behavioral response to the contract
- "Selection on moral hazard" (Einav et al. 2013)
  - implications for current welfare analysis based on mispricing
  - implications for combatting adverse selection through monitoring

Empirical application

- In paper: variation in relative price charged for high vs low deductible plans offered to different business units within Alcoa
- Rich data, including all relevant aspects of option set (vs. e.g. "networks")
- What is the source of price variation and the identifying assumptions?
    - How can we investigate validity?
    - What do we think about the empirical strategy?
- Learning about welfare consequences of adverse selection between no deductible and $500 deductible
    - Lampost problem?

Setting: employer provided health insurance

- Most (~90%) of private health insurance is provided through employers
  - economies of scale; pooling mechanism
  - tax subsidy to employer-provided health insurance
- Employer contributions to employee health insurance premiums are not taxed as income to employee
  - Single largest federal tax expenditure ($173 billion for FY 2019)
- Consider $x Worker compensation paid in cash (wages) vs employer contributions to health insurance premiums
  - Worker gets X(1-τ) in wages, but X in premiums
- Encourages provision of overly generous health insurance (Feldstein 1973)
  - "Cadillac tax" under the Affordable Care Act

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 uinits 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"
  - Borne out by data: prices are not correlated with observables of our sample of salaried workers (see Table 2)

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

Specification

- We estimate (using OLS):

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

  recall $c_i = c(m_i; H) - c(m_i; L)$
- Marginal cost derived from these without additional estimation

Raw data with basic findings

(Relative) price ($) (1)	Number of employees (2)	Fraction chose contract H (3)	Average incremental cost ($) for those covered under: Contract H (4)	Contract L (5)
384	2,939	0.67	451.40	425.48
466	67	0.66	499.32	423.30
489	7	0.43	661.27	517.00
495	526	0.64	458.60	421.42
570	199	0.46	492.59	438.83
659	41	0.49	489.05	448.50

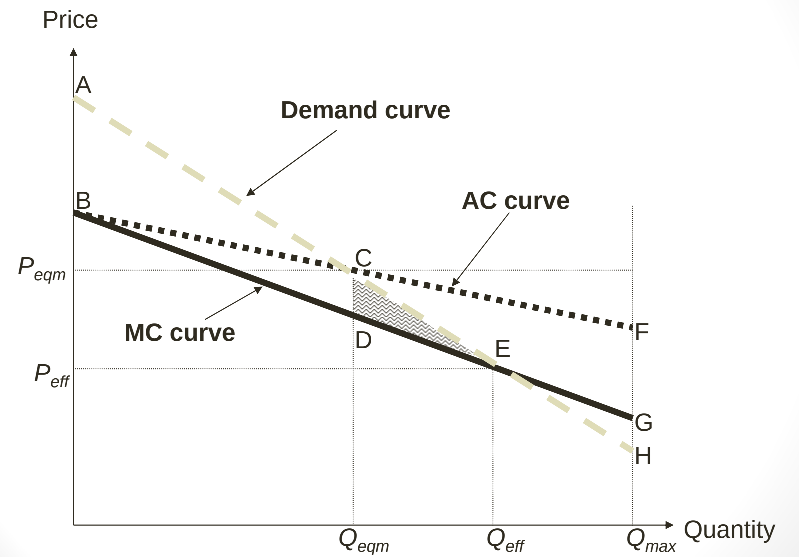
Results: estimates

Dependent variable (sample)	1 if chose High (both High and Low) (1)	Incremental cost (only High) (2)
	Panel A: Estimation results	
Relative price of High (US$)	−0.00070	0.15524
	(0.00032)	(0.06388)
	[.034]	[.021]
Constant	0.940	391.690
	(0.123)	(26.789)
	[.000]	[.000]
Mean dependent variable	0.652	455.341
Number of observations	3,779	2,465
R^2	.008	.005

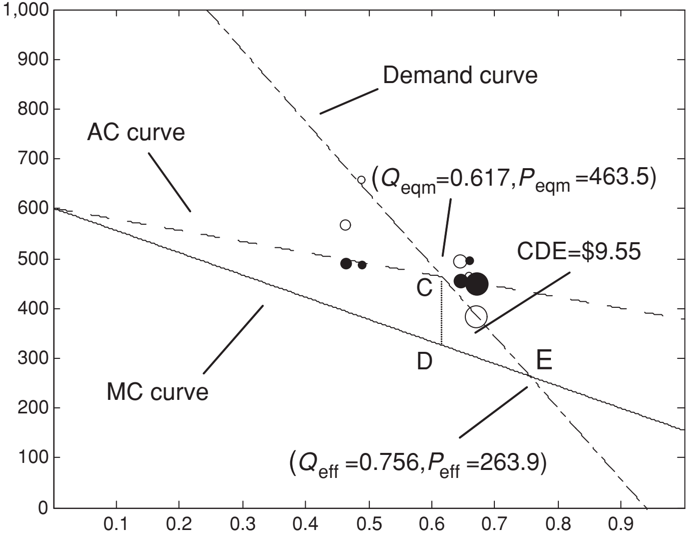
Standard errors (in parentheses) clustered on state

p-values in [square brackets]

The welfare cost of adverse selection



Results: graphical illustration



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 ˜3% of total surplus at stake from efficient pricing

Many possible applications

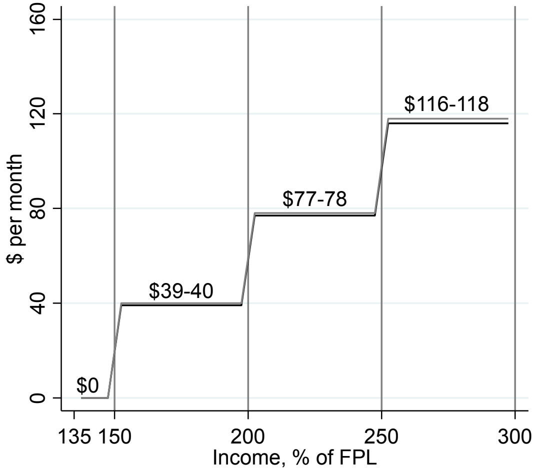
- Relatively little work estimating welfare costs of selection (vs. testing for selection)
- Many (better!) possible sources of pricing variation, including
  - sharp pricing changes over time: Landais et al. (2021) estimate welfare consequences of choice vs mandate for supplemental UI in Sweden
  - regulatory induced discontinuities in pricing - Finkelstein, Hendren and Shepard (2019)

Application: Health insurance subsidies for low income adults

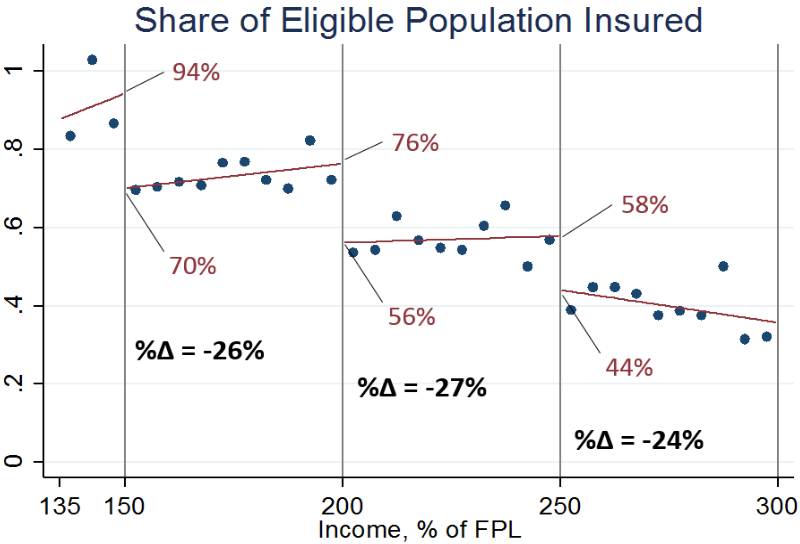
- Finkelstein, Hendren, Shepard (2019)
- Subsidized health insurance exchange in MA introduced in 2006 "Romneycare" reform
  - Precursor to ACA exchanges
  - Subsidies for low-income, non-elderly uninsured adults between 133-300% of FPL
- Quasi-random pricing across individuals
  - Public subsidies - designed to make insurance "affordable"
    - Change at discrete income bins
  - Regression discontinuity design

Quasi-Random Variation in Price

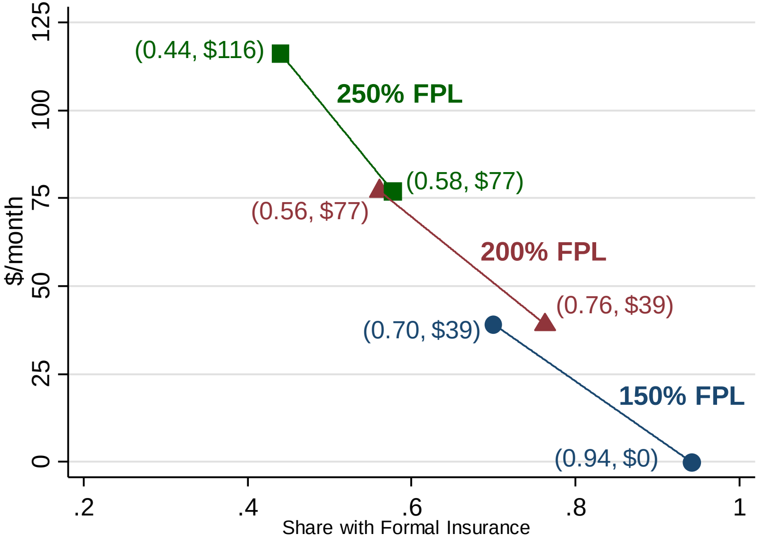
Panel A: Premiums for Cheapest Plan (2009-2013)



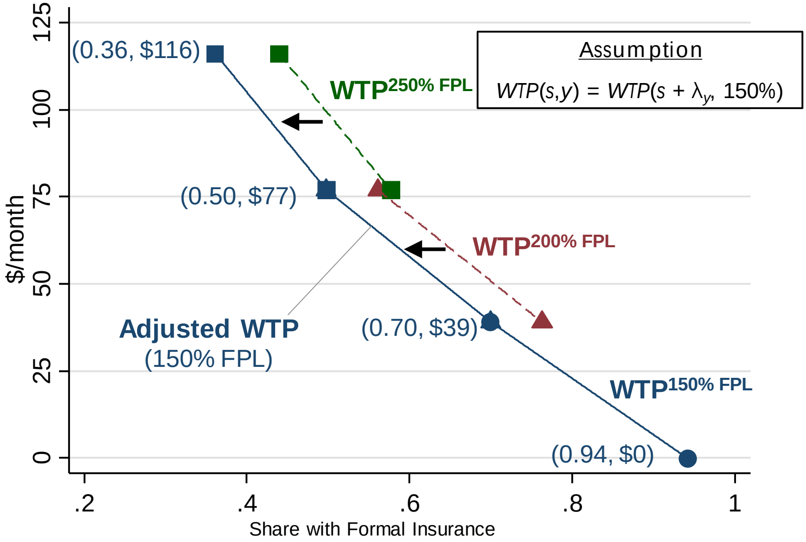
Price Changes Prompt Coverage Changes



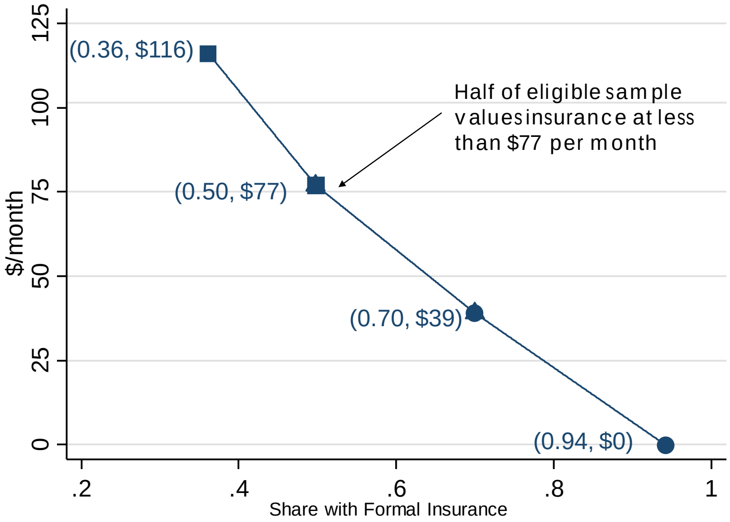
Observed Demand Points



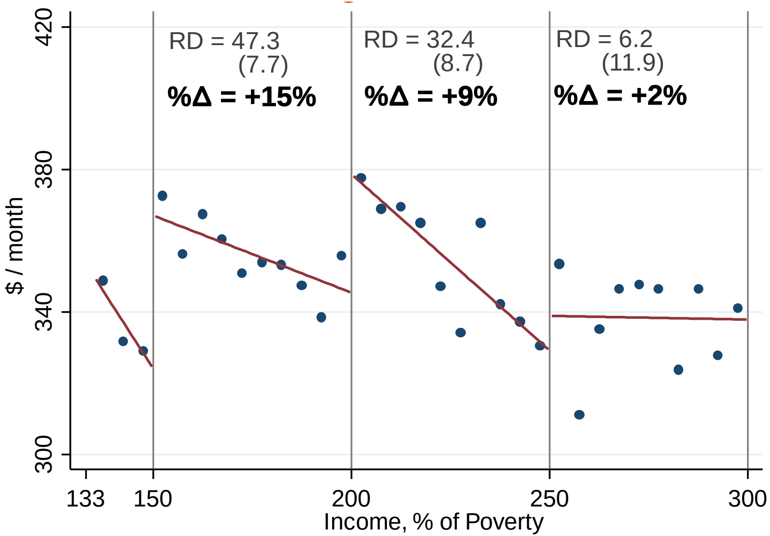
WTP (Adjusted to 150% FPL)



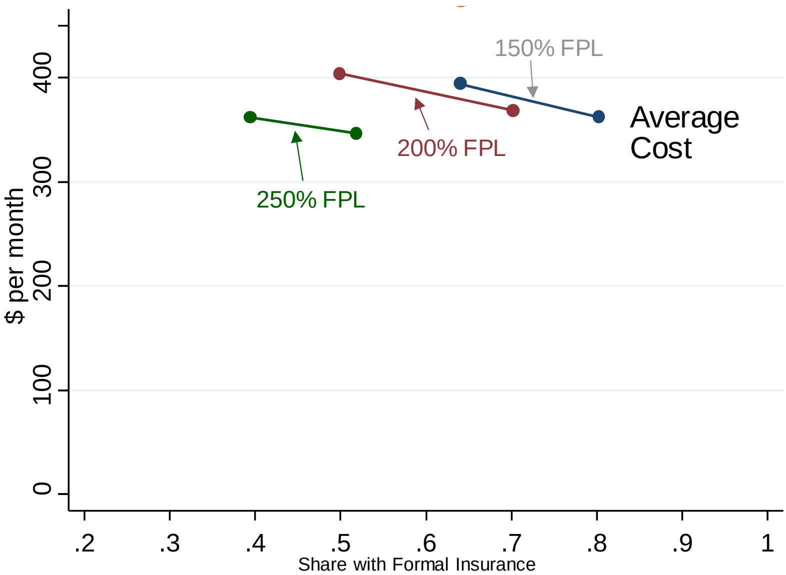
Demand Curve



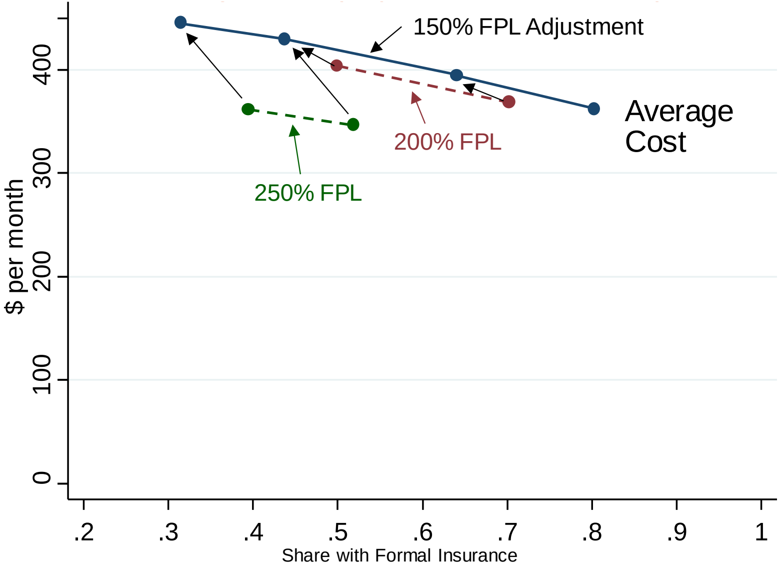
Average Insurer Costs



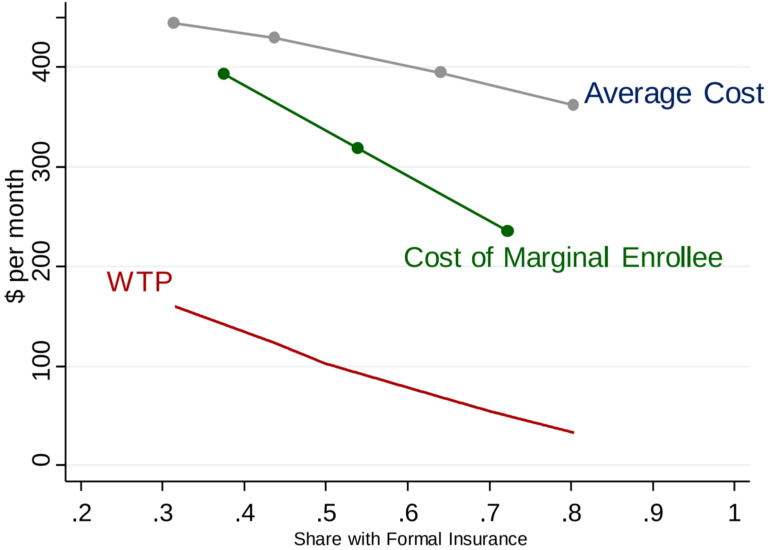
Observed Average Costs



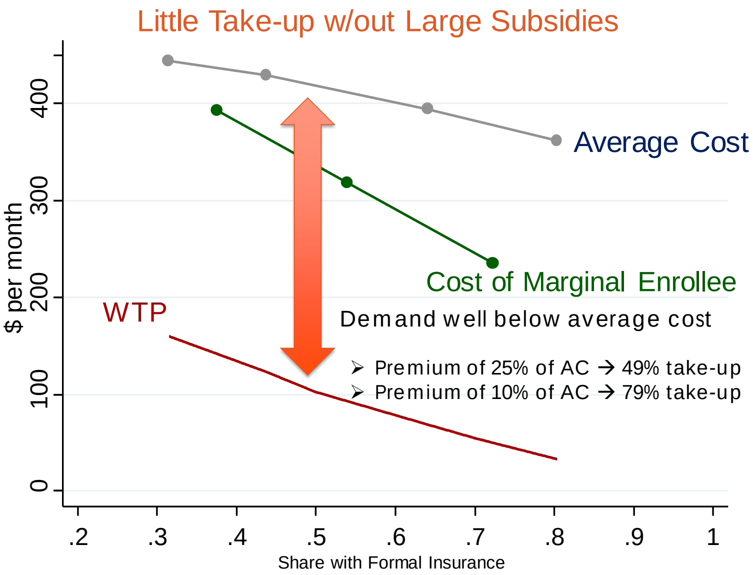
Average Cost (Adjusted to 150% FPL)



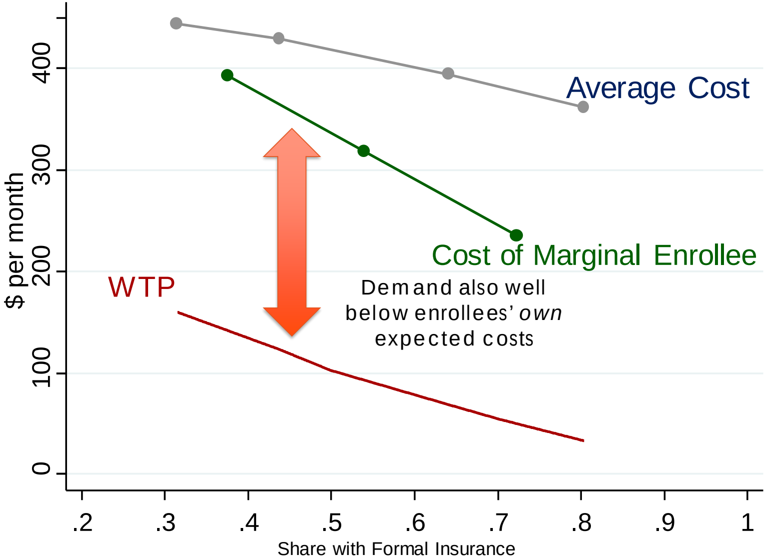
Final WTP and Cost Curves



Little Take-up without Large Subsidies



Adverse Selection Alone Cannot Explain Low Coverage



So Why is WTP Below Own Costs?

Application: Credit Markets

- Analagous problem: insurance premium ˜ interest rate on loan
- Additional challenges:
  - Assumption that premiums affect contract choice but not costs conditional on product choice allows separation of selection from moral hazard
    - insurance: premiums are sunk, so only threat is income effects
    - credit markets: interest rates can affect downstream probability of default
  - Demand curve estimates maximum quoted rate borrowers are willing to accept
    - Need to adjust down for borrowers' expected default; absent moral hazard adjustment factor based on MC curve (= expected charge off rate)
    - With moral hazard, this is an overadjustment / will overestimate WTP (lower price to get marginal borrower, which lowers default rate on inframarginal borrowers)
- Application: DeFusco et al. (2021) RCT of interest rates for large fintech lender in China

Application: Extensive vs Intensive Margin Selection

- Geruso, Layton, McCormack, and Shepard (2019)
- "Trade-offs between Extensive and Intensive Margin Selection in Competitive Insurance Markets"
- Use EFC (2010) framework to make a simple, important point:
  - once you have more than two choices (e.g. insured vs not and within insured high vs low coverage), then policies that work to reduce selection on one margin can worsen it on the other
  - e.g. insurance mandate penalty
    - can reduce selection on extensive margin
    - but worsen on intensive margin, by bringing in healthy people who lower cost of low coverage plan, can get people dropping out of high coverage plan
- Show calibrated results using demand and cost system from Finkelstein, Hendren and Shepard (2019)
  - Nice example of re-using existing estimates for another purpose

Extension: 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 equilibirum
- Could "easily" extend welfare analysis under a different specific assumption about competition
  - Mahoney and Weyl (2017) develop this formally

Mahoney and Weyl (2017)

- 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 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**

The Theory of the Second Best

- Lipsey and Lancaster (ReStud 1956)

  "It is well known that the attainment of a Paretian optimum requires the simultaneous fulfillment of all the optimum conditions. The general theorm of the second best optimum states that if there is introduced into a general equilibrium system a constraint which prevents the attainment of one of the Paretian conditions, the other Paretian conditions, although still attainable are, in general, no longer desirable.... From this theorm there follows the important negative corollary that there is no a priori way to judge as between various situations in which some of the Paretian optimum conditions are fulfilled, while others are not."

- Translation: Full employment program for empirical economists

Discussion of EFC: Attractions

- Model demand and costs but not their primitives (ζ_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 of EFC: Limitations

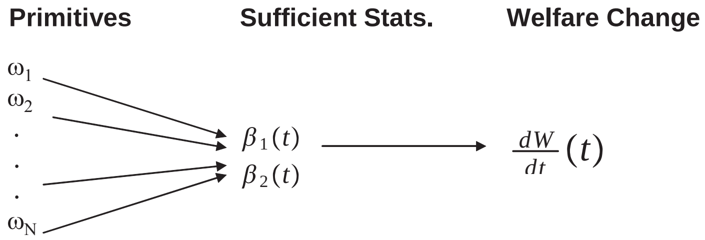
- Requires good price variation – not always easy to find (but see many possibilities...!)
- Fixed contracts assumption
    - Cannot evaluate welfare from introducing contracts not observed
        - Requires underlying structural primitives (as in EFS EMA 2010)
        - Welfare analysis limited to 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
        - Product-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.

Discussion: key assumptions of framework

- "Valid" pricing variation
- 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 (Chetty 2009)

- Approach by which (hopefully well-identified) reduced form parameters can be mapped - via a model - into economic objects of interest such as welfare



ω=preferences, constraints	$\beta = f(\omega,t)$ $y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	dW/dt used for policy analysis
ω not uniquely identified	β identified using program evaluation	

Sufficient statistics

- 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
  - Sufficient given the model (e.g. fixed contracts)

EFS (2010) – The big picture

- 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
  - The art of the counterfactual

Recovering risk and preferences from claims and choices

- 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 potential rationales for Social Security:
  - forced savings (paternalism)
  - redistribution based on lifetime (rather than annual) income (Akerlovian tag)

Welfare enhancing potential of annuities

- 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:

$$\min_{c_1,A,B} c_1 + A + B \;\; s.t. U(c_1, R_A A + R_B B) \geq U\_bar$$

- 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 indivdiuals 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).

Davidoff, Diamond and Brown (AER 2005)

- 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 actuarialy 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

End of Annuity Interlude

- We now all understand how fascinating and imporant 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
    - Side note: We are studying a semi-compulsory annuity market
    - Recently, there has been a lifting of the compulsory requirements - potential topic to study?!

Model and estimation

- 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

Guarantee Length	60 Females	65 Females	60 Males	65 Males
0	0.1078	0.1172	0.1201	0.1330
5	0.1070	0.1155	0.1178	0.1287
10	0.1049	0.1115	0.1127	0.1198

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 α_i
  - Time separable CRRA utility

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

- Heterogeneity in
  - risk type, α_i – mortality rate
  - preferences, β_i – weight placed on wealth at death
- Given α_i, β_i individual chooses annuity contract that maximizes lifetime utility (given optimal consumption path)
  - Optimal guarantee length increases with mortality (α_i) and preference for wealth after death (β_i)

Additional Assumptions

- Gompertz survival function with shape parameter λ and shift parameter α
  - Individual hazard rate as function of age (t) given by $\psi^i(t) = \alpha_i e^{\lambda t}$
- α and β 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 β (bequests? ex ante regret? etc.)
- Relatedly, note that β is not separately identified from risk aversion (γ), discount rate (δ), 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 β is similar to allowing heterogeneity in other preference parameters
    - Also try alternative model in which allowing for heterogeneity in other parameter (e.g. γ), rather than β

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)

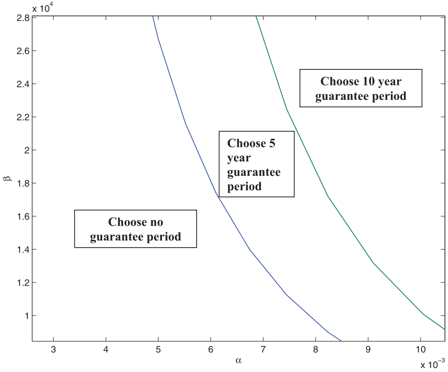
- Nature of ex ante information about risk type
  - We assume perfect information about mortality type (individuals know their own α)
- Identifying private information about mortality requires modeling assumptions
  - Although not for existence. See conditional correlation between guarantee and mortality (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 λ.
    - 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 μ_α, average slope affects estimate of λ, and concavity affects estimate of σ_α (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) λ using mortality data
- Calculate cutoff given λ using guarantee choice model (essentially no data yet)
- Estimate (by ML) distribution of α and β using cutoffs, guarantee data, and mortality data



Data

- 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

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

- 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 gaurantee (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:

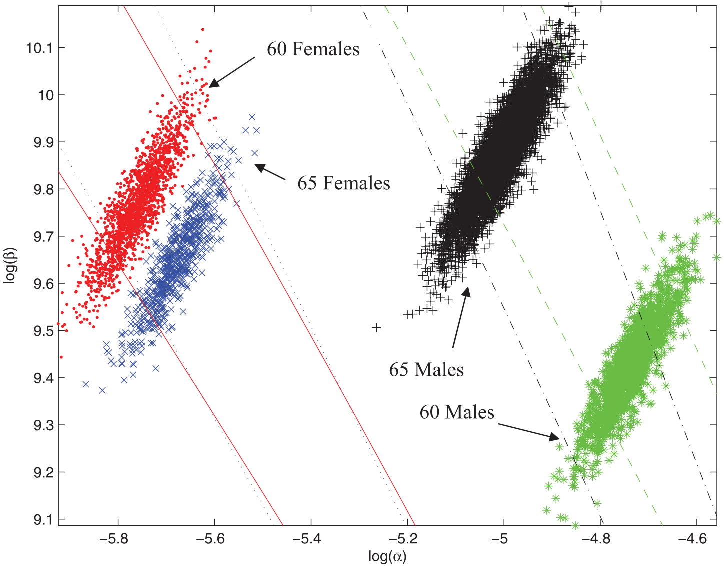
Guarantee Length	60 Females	65 Females	60 Males	65 Males
0	0.1078	0.1172	0.1201	0.1330
5	0.1070	0.1155	0.1178	0.1287
10	0.1049	0.1115	0.1127	0.1198

Estimates

Parameter Estimates

		Estimate	Std. Error
μ_α	60 Females	–5.76	(0.165)
	65 Females	–5.68	(0.264)
	60 Males	–4.74	(0.223)
	65 Males	–5.01	(0.189)
σ_α		0.054	(0.019)
λ		0.110	(0.015)
μ_β	60 Females	9.77	(0.221)
	65 Females	9.65	(0.269)
	60 Males	9.42	(0.300)
	65 Males	9.87	(0.304)
σ_β		0.099	(0.043)
ρ		0.881	(0.415)
No. of obs.		9364	

Graphical illustration



Model Fit

- 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)

Welfare analysis

- 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

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

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)

Discussion: Strengths

- 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 (α) and preferences (β)
  - 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

Comment: Empirical welfare analysis of contract design

- Area of opportunity / very little work
- EFC (2010) fixes contract design. EFS (2010) allows analysis of alternative contracts (if you are willing to impose all that structure)
- In practice, relatively little work
  - Recall small empirical literature testing impact of contract design on selection (e.g. Shepard 2022).
  - This is the first step...

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
  - Cutler and Reber (QJE 1998), Carlin and Town (2010), Bundorf, Levin and Mahoney (AER 2012), Einav, Finkelstein, Ryan, Schrimpf and Cullen (AER 2013), Handel (AER 2013).
  - 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.
  - Perhaps where it WAS a big deal for welfare, that's where policy solutions emerged
    - Tax subsidies for employer-provided health insurance
    - Mandates
    - Social safety net / publicly provided insurance
  - 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 (2019) on extensive margin insurance choice
    - What about welfare costs from complete unraveling of market (ultimate distortion of contract space)?