14.472 Public Finance II

Asymmetric Information: Welfare Analysis If Demand Does not Reveal Value

Amy Finkelstein

Fall 2024

Road Map

- Thus far have estimated welfare costs of adverse selection using approaches that rely on observing demand and using revealed preference
- Two additional topics to consider in welfare analysis of insurance markets
  - [Up next] What if observed demand does not reveal value?
  - What if market doesn't exist / has completely unraveled and we therefore cannot observe demand. How do we recover preferences?

Outline

- Theory: Why might demand not reveal value and implications for existing welfare analyses
- Empirics:
  - Using data to identifying behavioral models
  - Welfare analysis in the presence of behavioral frictions

Why Might Demand Not Reveal Value?

- Economic constraints: Liquidity constraints (NO!)
- Timing of measurement of demand - after information revealed / risk resolved
- Behavioral constraints: misperception, inattention, inertia, cognitive limitations

Liquidity constraints

- Definition: inability to borrow against future income at market rate of interest
- Insurance products often have a temporal dimension (Casaburi and Willis AER 2018)
  - pay premiums up front
  - therefore insurance transfers income across time as well as states
- This means that in terms of PDV lifetime budget, might want to purchase insurance at existing price but do not want to purchase insurance out of current income
  - Cost of borrowing is higher than market interest rate

Demand reveals value in presence of liquidity constraints

- If WTP for health insurance is low because indivdiuals have a high value of current cash due to liquidity constraints, this means they prefer other consumption to health insurance
- It doesn't mean their choice does not reveal their valuation
    - That valuation itself is affected by the liquidity constraints

Why Might Demand Not Reveal Value?

- Economic constraints: Liquidity constraints (NO!)
- **[Up Next]** Timing of measurement of demand - after information revealed / risk resolved
- Behavioral constraints: misperception, inattention, inertia, cognitive limitations

Timing of measurement of demand

- Hendren (2021 ReStud) "Measuring Ex Ante Welfare In Insurance Markets"
- Key idea: observed demand does not capture value of insurance prior to when demand is measured
- By time demand is measured may have already learned something about your type (that generates adverse selection) which destroys some of the insurance value
- Implication: EFC (2010) may systematically under-state welfare cost of adverse selection!
- Consider extreme example: if demand is measured at the point where individuals know their costs, demand equals cost and private market would unravel
  - If use observed demand and cost curves to measure welfare loss, would find no loss - willingness to pay does not exceed costs for anyone
  - But what about individuals' willingness to pay prior to learning their costs?

Toy (extreme) example

- Individuals have \$30 and face a uniformly distributed risk of losing between \$0 and \$10. How much would they be willing to pay ($D^{ex-ante}$)?

$$u(30 - D^{ex-ante}) = \int_0^{10} u(30 - x)dx$$

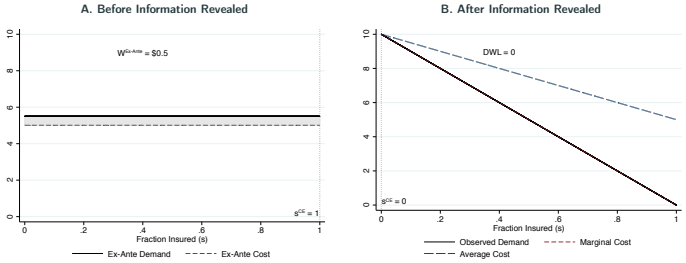
- Assuming coefficient of relative risk aversion of 3, $D^{ex-ante}$~\$5.50 , so indifferent between \$24.50 with certainty or a uniformly distributed consumption between \$20 and \$30
- Expected cost to insurer of insuring everyone is \$5, so insurance delivers a surplus of $W^{ex-ante} = \$0.50$

example (con't)

- But if demand is observed after individuals have learned their loss m with certainty
  - Then WTP will equal cost: $D(s) = m(s)$
  - Given uniform distribution of risk this generates a linear demand curve falling from \$10 at $s = 0$ to \$0 at $s = 1$ (where $s \in [0, 1]$ denotes fraction insured)
  - demand always equals marginal cost (not willing to pay above MC because no uncertainty)
  - Insurance would completely unravel because average cost of insuring fraction s in market always exceeds demand
- Using observed demand would not measure any welfare loss (0 DWL bc D=MC)
  - But recall from ex-ante perspective, welfare loss from not having insurance was \$.50

example (con't)

Figure 1: Example Demand and Cost Curves



How do we measure ex ante welfare?

- Need to know
  - Extent of ex ante heterogeneity - i.e. distribution of risk types before information is revealed ("behind the veil of ignorance")
  - How risk averse are individuals (curvature of utility function)
- Key insight: slope of the *observed* demand and cost curves reveal extent of ex ante heterogeneity
  - If ex ante everyone were same, demand and cost curves would be flat
  - Ex-post cross-sectional heterogeniety (slope of demand and cost curves) is ex ante risk heterogeneity
- Where do we get risk aversion?
  - Calibrate (i.e. assume) it from other estimates - very standard (but problematic)
  - Or back it out from difference between observed demand and cost curve (requires assumptions like e.g. no moral hazard)

Aside: Reclassification risk

- Ex-ante perspective related to issue of relcassification risk ("premium risk")
- In dynamic (multi-period) context, individuals benefit not only from period-by-period "event" insurance but also from insurance against becoming a bad risk and being reclassified into a higher risk group with a higher premiun
- Problem is one of *symmetric* information:
  - How to "insure" information that is known at time of contracting (but from an earlier perspective one faced ex-ante risk)
  - e.g. risk of being (or becoming) a bad driver
- Will discuss more in a few lectures
  - For those interested, two great references are Hendal and Lizzeri (QJE 2003) and Handel, Hendel and Whinston (EMA 2015)

Why Might Demand Not Reveal Value?

- Economic constraints: Liquidity constraints (NO!)
- Timing of measurement of demand - after information revealed / risk resolved
- **[Up Next]** Behavioral constraints: misperception, inattention, inertia, cognitive limitations

Spinnewijn (2017)

- Imagine there is some "non-welfarist constraint" that affects insurance demand but not insurance value
- Example. Discrepancy between perceived and actual risks
- Key: Creates a wedge between *actual value* of insurance and value of insurance *revealed* by individual demand
- In this setting, revealed preference approach likely *systematically understates the welfare cost of adverse selection*

Key selection effect

- If there is discrepancy between perceived and actual risks, on average those who select insurance will tend to over-estimate value of insurance and those who don't buy will under-estimate it
  - Note: can get this even if beliefs are accurate *on average* as long as there is some distribution of gap between perceived and actual risk
- As a result, demand curve overstates surplus for insured and understates potential surplus for uninsured
- If we treat demand curve as value curve (i.e. use revealed preference) get unambiguous sign to bias
  - Under-estimate welfare cost of selection; under-estimate welfare gain from mandate
- EFC (2010) may systematically under-state welfare cost of adverse selection!

Perceived vs "true" value

- Individuals differ on a vector of characteristics ζ
- $v(\zeta)$: *true* value of insurance (relevant for welfare)
- $\widehat{v}(\zeta)$: *perceived* value of insurance (Determines demand)
- noise term ϵ drives wedge between true and perceived value

$$\widehat{v}(\zeta) = v(\zeta) + \epsilon(\zeta) \text{ with } E_{\zeta}(\epsilon) = 0$$

- e.g. Noise term is positive if over-estimate risk, negative if under-estimate risk
- Key insight: even if noise cancels out across entire population (so true and perceived value are equal on average), since demand for insurance depends only on perceived value, true and perceived value may differ substantially conditional on insurance decision

Demand curve vs. value curve

- Demand: Buy if perceived value exceeds price: $\widehat{v}(\zeta) \geq p$

$$D(p) = 1 - F_{\widehat{v}}(p)$$

- Demand curve reveals WTP of marginal buyers at different prices. Price reveals perceived value for marginal buyer at that price:

$$p = E_{\zeta}(\widehat{v}|\widehat{v} = p)$$

- For welfare, what is relevant is expected true value of marginal buyers

$$MV(p) \equiv E_{\zeta}(v|\widehat{v} = p)$$

Demand curve vs. value curve

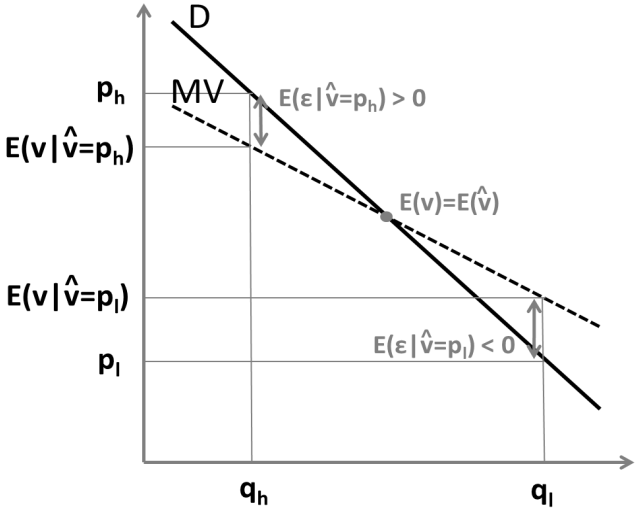


Figure 1: The Demand Curve and the Value Curve.

Implications for inferring insurance value

- (Proposition 1): If true value v and noise term ε are independent, demand curve overestimates insurance value for insured and underestimates the insurance value for the uninsured
- Simple selection effect: those who buy are selected for positive ε (and those who do not for negative ε):

$$E_\zeta(\varepsilon|\widehat{v} \geq p) \geq 0 \geq E_\zeta(\varepsilon|\widehat{v} < p) \quad \text{for any } p$$

- What if noise is not independent of true value?
  - (Proposition 2) If true and perceived value are normally distributed, sign of bias remains same as long as the correlation between noise term and true value is not "too negative".
- Naive policy maker (using demand curve) will overestimate insurance value for insured and underestimate insurance value for uninsured when true value changes less than one for one with perceived value

Implications for cost of adverse selection

- Assume Adversely seleted equilibrium generates too little insurance, measured wrt the marginal value (MV) curve
- Welfare cost of under-insurance depends on difference beween MV curve and MC curve for those not insured in equilibrium (demand below AC) but efficient to insure (MV above MC)
- If instead use demand curve to estimate welfare cost of adverse selection estimate welfare cost of adverse selection as difference bewteen *demand* and MC for those not insured in equilibrium (demand below AC) but efficient to insure (*demand* above MC)
- Two causes of under-estimating welfare loss from under-insurance (from using Demand curve instead of MV curve):
  - Misidentify set whom it is efficient to insure
  - Misidentify the welfare loss for those inefficiently uninsured

Under-estimating cost of AS using revealed preference

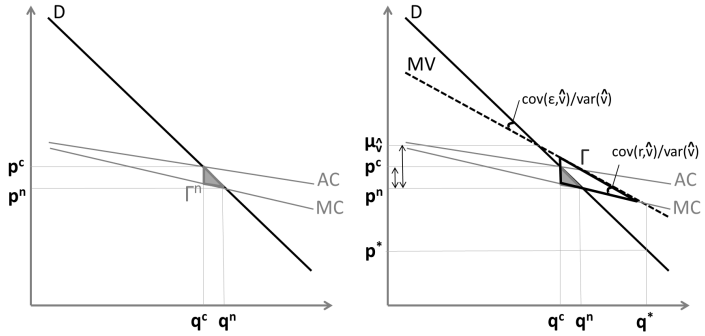


Figure 2: Adverse Selection: the naively estimated cost Γ^n vs. the actual cost Γ.

Comments

- Nice, and likely important insight
- What are the sources of "noise" (ε)
  - Uses example of perceived vs actual risk
  - Other behavioral constraints: inattention, cognitive inability, inertia
- Welfare can become trickier. Why is $v(p)$ relevant for welfare instead of $\widehat{v}(p)$? Welfare from whose perspective?

Key challenge: how implement empirically?

- In addition to the demand and cost curves needed for EFC (2010), need one additional statistic:
  - share of the variation in insurance demand - left unexplained by heterogeneity in risks – that is driven by non-welfarist constraints rather than by heterogeneous preferences
- Need additional data to disaggregate revealed value of insurance into true value and constraints
- Two components
  - Testing: How do we identify these "behavioral" constraints empirically? Will now discuss. There has been progress but scope for more.
  - Quantifying: How do we estimate the value curve - i.e. what the demand curve would be in absence of all behavioral frictions?
    - A key - and ongoing – challenge

Recap / Up Next

- [Done] Observed demand may systematically understate welfare cost of selection
  - Hendren (2021)
  - Spinnewijn (2017)
- [Up next]: The challenge (and the frontier) I: Using data to identify "behavioral" departures
- [Then] The challenge (and the frontier) II: Incorporating frictions into welfare analysis

Recall identification challenge from Cohen and Einav

- Can rationalize choices with sufficiently flexible joint distribution of risk type and risk aversion
  - e.g. if see someone with history of low claims choosing low deductible plan, must be very risk averse (and/or high risk and lucky)
- First generation of behavioral models impose restrictions / functional form on joint distribution of risk type and risk preferences and call what they can't explain 'mistakes'
  - Not very appealing
  - But neither is assuming no mistakes and rationalizing everything with sufficiently flexible distribution

Using Data To Identify the Behavioral Model

- 'Inertia' (Handel AER 2013)
- Violations of Slutsky Symmetry (Abaluck and Adams-Prassl (QJE 2021)
- Dominated Choices (Handel AER 2013, Bhargava et al. QJE 2017)

Handel (2013)

- Standard economic theory: choice is good
  - Competition / productive efficiency
  - Preference heterogeneity / allocative efficiency
- Thus far, have considered two separate factors that can mitigate against value of choice
  - Adverse selection (potential welfare improving role for mandates)
  - "Mistakes" / choice inconsistencies
- Handel (2013) now combines them
  - Investigates consumer inertia in health insurance markets where adverse selection is a potential concern

Setting and Data

- Employee menu of health insurance options, choices, and claims over seveal years in a large firm
- Very similar set up to Alcoa data in Einav, Finkelstein and Cullen (2010)
- Key feature: Changes in menu
  - Firm significantly altered menu of plans, forced employees out of old plans (no longer offered) and required them to make an active choice from new menu (no stated default)
  - In subsequent years, options remain same but premiums changed a lot and if no active choice, defaulted into prior year's choice
- Key identifying feature for inertia: when employees join firm relative to when menu or price changes occur

Overview of Approach

- Descriptive evidence of inertia
  - Comparison of choices made by different cohorts of new employees (very different choice environemnt; otherwise appear quite similar)
  - Dominated Choices
- Model (a la Cohen and Einav 2007) to recover distribution of risk type, risk preferences and switching costs. Can be used to
  - Quantify the extent of switching costs
  - Model plan choice and welfare under counterfactual policies (such as forced active choice / no inertia by construction)
- NB: Very nice pairing
  - Descriptive evidence on key feature of model (relatively model free)
  - Additional modeling assumptions allow him to ask questions (counterfactual choice; welfare) that you can't get from the reduced form

Overview of findings

- Substantial inertia from descriptive evidence
  - As plan prices and choice environment change over time, incoming cohorts of new employees make active choices that reflect updated setting while prior cohorts make very different choices that reflect past setup (cohorts look otherwise similar)
  - Some options become *dominated* and yet most consumers stay with them (NB: strict dominance doesn't require modeling assumptions about e.g. risk type or preferences)
- Counterfactual results from model: inertia ameliorates adverse selection and improves welfare
  - Reduces adverse selection pressure (i.e. healthiest dropping out)
  - Application of theory of the second best

Descriptive Part I: New employees and inertia

- Key idea: new employees forced to make active choices (vs prior cohorts)
- Compare how choices vary for new employees vs old (confirming that demographics don't vary across cohorts)
- Notation:
  - t_0 = year of menu change (everyone has to make an active choice)
  - t_1 : menus don't change (so can be passive) but large price changes
- Examines: how do t_1 choices vary for those who enter at t_1 (active choices) vs. those already in at t_0 (potentially passive)

Table 2

New Enrollee Analysis	New Enrollee t_{-1}	New Enrollee t_0	New Enrollee t_1
N, t_0	1056	1377	-
N, t_1	784	1267	1305
t_0 **Choices**			
PPO_{250}	259 **(25%)**	287 **(21%)**	-
PPO_{500}	205 **(19%)**	306 **(23%)**	-
PPO_{1200}	155 (15%)	236 (17%)	-
HMO_1	238 (23%)	278 (20%)	-
HMO_2	199 (18%)	270 (19%)	-
t_1 **Choices**			
PPO_{250}	182 **(23%)**	253 **(20%)**	142 **(11%)**
PPO_{500}	201 **(26%)**	324 **(26%)**	562 **(43%)**
PPO_{1200}	95 (12%)	194 (15%)	188 (14%)
HMO_1	171 (22%)	257 (20%)	262 (20%)
HMO_2	135 (17%)	239 (19%)	151 (12%)
Demographics			
Mean Age	33	33	32
Median Age	31	31	31
Female %	56%	54%	53%
Manager %	20%	18%	19%
FSA Enroll %	15%	12%	14%
Dental Enroll %	88%	86%	86%
Median (Mean) Expense t_1	844 (4758)	899 (5723)	-
Income Tier 1	48%	50%	47%
Income Tier 2	33%	31%	32%
Income Tier 3	10%	10%	12%
Income Tier 4	5%	4%	4%
Income Tier 5	4%	5%	5%

Welfare relevance?

- Is this inertial / lack of attention or just a good-old fashioned neoclassical switching cost
  - Matters for welfare
  - Handel (2013) admits agnosticism and explores alternative assumptions
- How could we tell?
  - Enter Abaluck and Adams-Prassl (2021)

Violations of Slutsky Symmetry

- Abaluck and Adams-Prassl (2021)
- 'Consideration set models': generalization of discrete-choice models to relax the assumption that individuals consider all goods
  - Could arise from inattention / bounded rationality, search costs, unobserved constraints on options
  - goal: simulate how consumers would choose if they were informed about relevant options
- Two key classes of consideration models:
  - Default specific consideration model: consumers are either 'asleep' and choose default or 'wake up' and make an active choice
  - Alternative specific consideration model: Each good has an independent consideration probability which depends on characteristics of that good
- Key identification challenges: if changes in prices or product characteristics affect choices, is this via consideration or via utility?
  - i.e. to what extent to observed choice probabilities reflect consideration vs preferences?

Violations of Slutsky Symmetry

- Their proposal: use violations of Slutsky symmetry to identify consideration sets / ask how consumers would choose with full consideration
- Key insight: imperfect consideration breaks symmetry between cross-price responses
  - symmetry in cross price demand responses: demand response of plan A to change in price of plan B same as demand response of plan B to change in price of plan A
  - But if inattentive consumers choose default option and changes in price perturbs consideration by causing some consumers to 'wake up and pay attention', then demand for nondefault plan will be more sensitive to price of the default plan than demand for default plan is to price of nondefault plan
- Violations of Slutsky symmetry identify the probability that consumers consider various subsets of products

Application

- Choice of Medicare Part D prescription drug plan
- Treat individual's prior year plan as 'default'
- Separately identify inattention from utility-driven switching costs by asymmetries in how the decision to remain with default plan depends on default characteristics versus rival plan characteristics
- Key finding: switching decisions significantly more sensitive to default premiums and characteristics than rivals
    - Consistent with a model in which many consumers do not actively search each period but are induced to make an active choice if the default plan becomes bad enough

Using Data To Identify the Behavioral Model

- 'Inertia' (Handel AER 2013)
- Violations of Slutsky Symmetry (Abaluck and Adams-Prassl (QJE 2021)
- **[Up Next]** Dominated Choices (Handel AER 2013, Bhargava et al. QJE 2017)

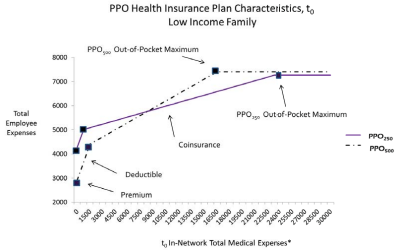
The Appeal of Dominated Choices

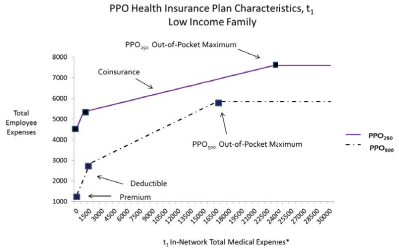
- Recall central identification challenge: observe risk realization not underlying risk type
  - so can rationalize all choices with flexible enough distribution of risk type and risk preferences
  - is the individual making a mistake when he has had no health care claims for years and is buying very comprehensive health insurance?
    - Or is he very risk averse?
    - Or does he have private information that he's higher risk than we (the econometrician) think?
- But if one option is strictly dominated, don't need to model individual's risk type!
  - Strict dominance: no matter what claims they have, total out of pocket spending (premiums + cost sharing) lower under one plan than another

Handel 2013 Descriptive Part II: Dominated choices

- Look at what happens when one option becomes dominated due to a price change over time
- In t_1 firm increased the premium for the (more comprehenisve) \$250 deductible plan (PPO_{250}) and decreased the premium for the (less comprehensive) \$500 deductible plan (PPO_{500})
  - For some combinations of family size and income (which determine employee premium contributions) PPO_{250} became *strictly dominated* by PPO_{500}.
  - Strict dominance: for any level and type of total medical expenditures, PPO_{500} leads to lower employee expenditures (premium plus out-of-pocket) than PPO_{250}
- Attraction of strict dominance: don't have to model individual's risk type

Strict dominance: illustration





Dominated choices: Findings

- Many people remain in dominated choices
- Of those whose choice becomes dominated in t_1, only 11% switch. Even by t_2 only 25% total have switched out of dominated option.
- Why do people choose dominated plans?

More on dominated choices: consumers

- Bhargava et al. (QJE 2017) present evidence from a large firm of majority of employees (and particularly lower income ones) choosing dominated plans
- Non-trivial consequence: estimate dominated choice results in excess spending equal to about 25% of chosen plan premium
- Conduct choice experiments that suggest
  - simplifying and shortening menu doesn't have much effect
  - clarifying economic consequences of plan choice does reduce dominated choices substantially
- Conclude: reflects fundamental lack of understanding of health insurance
- "Our findings challenge the standard practice of inferring risk preferences from insurance choices, and raise doubts about the welfare benefits of reforms that give consumers more choice"

More on dominated plans: firms

- Why do firms offer dominated plans? (Liu and Sydnor AEJ Policy 2022)
  - Provide descriptive evidence that dominated plan offerings are common
  - High deductible plan often dominates lower deductible plan
  - They propose it's due to selection: high deductible attracts better risks, and lower costs are then passed onto enrollees
- Are dominated plans actually dominated (Ericson and Sydnor NBER WP 2018)?
  - In presence of liquidity constraints, individuals may rationally prefer "dominated" plan
  - Compare low vs high deductible plan where premiums + oop for low plan exceed high for any realization of health care spending
  - But liquidity constrained individual may prefer a series of small payments to one large lumpy expenditure

Recap / Up Next

- Observed demand may systematically understate welfare cost of selection
- The challenge (and the frontier) I: Using data to identify "behavioral" departures
- [Up Next] The challenge (and the frontier) II: Incorporating frictions into welfare analysis

Handel (2013): Moving past the descriptive

- Descriptive evidence of inertia is compelling.
- What is the point of the model?
  - Quantify amount of intertia
  - Perform counterfactuals: how would choices (and welfare) change if we reduce inertia?

Choice model

- Model: essentially Cohen and Einav (2007) + inertia
  - Three dimensions of heterogeneity: risk type, risk preference, inertia
  - Same sort of modeling choices with respect to risk type and risk preferences that we have discussed
- Inertia modeled as an incremental (monetary) cost η that is paid to switch plans (structural interpretation is that of a switching cost). Has direct negative impact on utility

Choice model

$$U_{kjt} = \int_0^\infty f_{kjt}(OOP)\, u_k(W_k, OOP, P_{kjt}, \mathbf{1}_{kj,t-1})\, dOOP$$

- k is family unit, j is plan choice, and t is one of three years (t_0 to t_2)
  - t_0 everyone forced to make new, "active" choice
  - t_1 large relative price changes
- 3 plan choices: PPO_{250}, PPO_{500}, PPO_{1200}
  - differ only in financial aspects (premiums and cost sharing)
  - PPO_{1200} includes HSA (save tax-free for later medical expenses)
- OOP is realization of medical expenses from $F_{kjt}(\cdot)$
- W_k denotes family-specific wealth
- P_{kjt} is family x time specific premium contribution for plan j)
- $\mathbf{1}_{kj,t-1}$ is an indicator for whether family was enrolled in plan j in previous time period

Choice model

- CARA assumption
  - for a given ex-post consumption level x:

$$u_k(x) = -\frac{1}{\gamma_k(X_k^A)} e^{-\gamma(X_k^A)x}$$

- γ_k is a family-specific risk preference parameter (known to family; unobserved to econometrician)
  - γ_k is a random coefficient, assumed to be normally distributed (truncated just above zero) with a mean that is linearly related to observable characteristics X_k^A (employee age and income)
    - $\gamma_k(X_k^A) \tilde{} N(\mu_\gamma(X_k^A), \sigma_\gamma^2)$
    - $\mu_\gamma(X_k^A) = \mu + \beta(X_k^A)$
- Note: CARA assumption means don't need to observe wealth because level of absolute risk aversion $\frac{-u''}{u'} = \gamma$ which is constant with respect to level of x

Choice model

- family's consumption x, conditional on a draw OOP from $F_{kjt}(\cdot)$ is given by:

$$x = W_k - P_{kjt} - OOP + \eta(X^B_{kt}, Y_k)\mathbf{1}_{kj,t-1} + \delta_k(Y_k)\mathbf{1}_{1200} + \alpha H_{k,t-1}\mathbf{1}_{250} + \epsilon_{kjt}(Y_k)$$

- Inertia (η) modeled as an implied monetary cost / reduction in consumption (structural interpretation similar to a tangible switching cost)
  - depends on linked choice ($\mathbf{1}_{kj,t-1}$) and on demographic variables (X^B_{kt} and Y_k)
  - $\eta(X^B_{kt}, Y_k) = \eta_0 + \eta_1 X^B_{kt} + \eta_2 Y_k$
  - X^B_{kt} contains potentially time varying variables that may affect inertia (e.g. income, health status, change in predicted medical expenditures etc)
  - Y_k is family status (single vs dependents)

Choice model: inside the sausage factory

- family's consumption x, conditional on a draw OOP from $F_{kjt}(\cdot)$ is given by:

$$x = W_k - P_{kjt} - OOP + \eta(X^B_{kt}, Y_k)\mathbf{1}_{kj,t-1} + \delta_k(Y_k)\mathbf{1}_{1200} + \alpha H_{k,t-1}\mathbf{1}_{250} + \epsilon_{kjt}(Y_k)$$

- δ_k is an unobserved family-specific intercept for $PP0_{1200}$
  - expect non-zero δ_k because HSA in $PP0_{1200}$ is horizontally differentiated
  - so makes sense (and i'm guessing also helps fit the data better)
- $H_{k,t-1}$ is a binary variable for family above 90th pctile of cost distribution last year
  - α measures an intrinsic preference of a high cost family for PPO_{250}
  - "Intended to proxy for empirical fact that almost all families with very high expenses choose PPO_{250} whether or not it is the best plan for them".

Choice model: inside the sausage factory

- family's consumption x, conditional on a draw OOP from $F_{kjt}(\cdot)$ is given by:

$$x = W_k - P_{kjt} - OOP + \eta(X^B_{kt}, Y_k)\mathbf{1}_{kj,t-1} + \delta_k(Y_k)\mathbf{1}_{1200} + \alpha H_{k,t-1}\mathbf{1}_{250} + \epsilon_{kjt}(Y_k)$$

- ϵ_{kjt} is a family-plan-time specific idiosyncratic preference shock
  - assume probit error term, distributed $i.i.d$ for each j with zero mean and variable $\sigma_{e_j}(Y_k)$
- Standard thing to do (makes it a lot easier to rationalize the data) but kind of strange when choices differ purely on financial characteristics (conditional on the modeled PPO_{1200} differentiation)
  - Particularly unappealing if heterogeneity in preferences (or joint distribution of unboserved heterogeneities) is a focus
  - Einav et al. (2013 AER, selection on moral hazard) are focused on joint distribution of unobserved preferences, risk type and moral hazard type
    - Do not include this additional error term / "preference shock"
    - And incur much pain and suffering as a result

Cost model (for distribution of OOP)

- Model individual's (ex-ante - i.e. at time of insurance choice) expected future spending at time of plan choice using past diagnostic, demographic and cost information
  - generate ex-ante distribution faced by individual by grouping individuals into bins based on mean predicted future spending and estimate spending distribution for upcoming year based on ex-post observed cost realizations
- Impose two restrictions:
  - No moral hazard (total expenditures do not vary with j)
  - No private information about health conditional on model above
    - Question to class: so how can he estimate / study adverse selection?

Identification (loosely)

- Risk type modeled directly from (rich) information not only on claims but on "risk score" (past spending and diagnoses) which is used to group indivdiuals into cells for whom spending distribution is computed
  - No additional unobserved heterogeneity (or moral hazard).
- Risk aversion identified by choice between PPO_{500} and PPO_{250}
  - Identifed by active choices at t_0
- Inertia identified by choice movement (or lack thereof) over time as plan values change due to changes in price or health status
- Of course requires (some) parametric assumptions
  - Described specific choices above
  - Explores robustness

Identification (con't)

- Central (pervasive) challenge in many applications: separating path-dependence from serial correlation / persistence of types
  - e.g. argument over whether welfare "creates dependency" / reduces labor market potential. How separate path-dependence from persistence of types (does welfare erode human capital or do people with low human capital end up on welfare?)
- Here, fundamental challenge is to separately identify "inertia" from persistent, unobserved preference heterogeneity
  - inertia = state-dependence. if you randomly assigned someone to a plan they would be more likely to still be in it the subsequent year.
- Key to their approach: Changes in prices and health status over time identify inertia separately from risk preference levels and risk preference heterogeneity

Findings

- Large (and heterogeneous) inertia
  - Average employee to forgo ~ $2,000 annually (sd is $446)
  - Relative to average family spending of ~$4,500
- Counterfactual policies that "reduce inertia" (from η_k to $Z\eta_k$) where Z is some fraction
  - As Z goes to 0, eliminate inertia
  - Considers welfare as the certainty equivalent that equates expected utility under a health plan choice with a certain monetary payment such that individual indifferent between losing that amount for sure and obtaining the risky payoff from enrolling in the plan

Model findings (con't)

- How to think about η
  - Do you count reduction in η as "direct" welfare benefit. Depends on underlying source of inertia (e.g. real tangible switching cost vs. some abstract psychic force causing delay?)
  - Tries allowing for various fractions of η reduction to "count" in welfare
- Two main counterfactuals as reduce inertia
  - Partial equilibrium / naive: Changes in plans and welfare, holding premiums fixed
  - Allow supply side response: prices adjust as people move across policies (need model of supply side)

Model findings (con't)

- Counterfactual: Reducing inertia by three-quarters (not counting η directly in welfare)
- Partial equilibrium
  - 44% increase in fraction enrolling in PPO_{500} at t_1 (recall big decrease in relative premium)
  - Increase in welfare of about 5% of premiums
- Allowing supply side response of premiums:
  - Still improves plan choices conditional on prices (recall too few were choosing PPO_{500} at t_1) but now exacerbates adverse selection leadings to a *reduction* in welfare.

Model findings (con't)

- Why does reducing inertia reduce welfare once account for supply side response of premiums?
- Reduced inertia / choice frictions causes more people to re-optimize
  - leads to more enrollment in PPO_{500} when relative price decreases
  - On the margin it is the healthier ones who choose this lower coverage plan (PPO_{500})
  - So this drives up the price in PPO_{250} as it becomes more adversely selected
  - Over time, counterfactuals suggest PPO_{250} could experience a death spiral (a la Cutler and Reber 1998)

Comments

- Two reasons now in insurance markets that greater choice may not improve welfare
  - Selection
  - "Behavioral" issues / "bad choices"
- But what is "inertia"?
  - Matters crucially for welfare analysis (as paper realizes)
  - Modeled as a real switching cost (but baseline welfare analysis assumes it's not directly affecting utility)
  - Are search costs "behavioral"?
  - See Abalack and Adams-Prassl on distinguishing consideration sets from utility

Comments (con't)

- Handel (2013) finds that inertia ameliorates adverse selection in this setting once one accounts for supply side premium response
- But there is no general theorem. (e.g. "Anything that gums up choices ameliorates adverse selection").
  - although paper is often (mis-) interpreted this way
- Polyakova (AEJ:Applied 2016) finds for Medicare Part D switching costs **help** sustain an adversely selected equilibrium.
- Depends crucially on "where you start"
  - In Handel setting, inertial consumers respond little to the relative premium decrease for the low coverage (PPO_{500}) plan (even though it made PPO_{250} dominated in some cases).
  - Recall adverse selection creates problem of too little insurance / above MC pricing in higher coverage plans
  - If the price change had been relative premium decrease for high coverage (PPO_{250}) plan, inertia would have exacerbated adverse selection

Polyakova (2016): "Active" choices when enter at 65

Table 3: **Evidence of switching costs: choice patterns in 2006-2009 tracked for cohorts entering in different years**

Cohorts of 65 year olds whose incumbent plans were not re-classified into a different type by the insurer										
	65 y.o. in 2006				65 y.o. in 2007			65 y.o. in 2008		65 y.o. in 2009
A. *Enrollment shares*	2006	2007	2008	2009	2007	2008	2009	2008	2009	2009
Contracts of type 1	22 %	22 %	19 %	17 %	17 %	15 %	14 %	10 %	11 %	12 %
Contracts of type 2	73 %	73 %	77 %	79 %	72 %	75 %	77 %	82 %	82 %	82 %
Contracts of type 3	4 %	5 %	5 %	4 %	11 %	11 %	10 %	7 %	7 %	5 %
N	37,500	37,500	37,500	37,500	35,759	35,759	35,759	40,960	40,960	43, 520
B. *Incremental premium*	**in year 2006**				**in year 2007**			**in year 2008**		**in year 2009**
Contracts of type 2	\$138				\$125			\$54		\$37
Contracts of type 3	\$375				\$360			\$410		\$469

Shows up in differential price elasticities

Table 4: **Evidence of switching costs: price sensitivity estimates for individuals with and without incumbent plans**

	Age of beneficiaries					
Price coefficient [p-value]	65	66	67	68	69	70
	Baseline	Interaction	Interaction	Interaction	Interaction	Interaction
2006	**-0.003**	0.0001	0.0002	0.0006	-0.0001	0.0006
	[0.000]	[0.809]	[0.683]	[0.386]	[0.876]	[0.321]
2007	**-0.003**	**0.0018**	**0.0012**	**0.0011**	**0.0013**	**0.0010**
	[0.000]	[0.002]	[0.035]	[0.031]	[0.002]	[0.040]
2008	**-0.003**	**0.0022**	**0.0023**	**0.0021**	**0.0019**	**0.0020**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]
2009	**-0.010**	**0.0072**	**0.0085**	**0.0090**	**0.0085**	**0.0084**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

The price coefficients are estimated using the following random utility specification:

$$\begin{aligned} u_{ij} &= -\alpha_{65} p_{ij} + \alpha_{66} p_{ij} \mathbf{1}\{Age = 66\} + \alpha_{67} p_{ij} \mathbf{1}\{Age = 67\} + \\ &+ \alpha_{68} p_{ij} \mathbf{1}\{Age = 68\} + \alpha_{69} p_{ij} \mathbf{1}\{Age = 69\} + \alpha_{70} p_{ij} \mathbf{1}\{Age = 70\} + brand_j + \epsilon_{ij} \end{aligned}$$

Other interesting aspects of inertia / switching costs I

- How does it affect firm pricing?
- Ho, Hogan and Morton (2017 RAND) "Impact of consumer inattention on insurer pricing"
  - Theoretically unclear whether equilibrium prices higher or lower with strategic (dynamic) pricing behavior
    - Competing goals: invest (lower prices) vs harvest (raise prices)
  - Descriptive evidence that firm pricing reflects strategic response to inertia (e.g. increasing over time)
  - Explore implications for counterfactual pricing and welfare with non-strategic (static) pricing
- See also: bonuses for credit card switching or signing up for bank accounts..

Other interesting aspects of inertia / switching costs II

- Distributional consequences
  - Inertia redistributes from those with high inertia to low inertia
  - Under what social welfare functions is this desirable or undesirable?
    - Assume that social welfare function weights low income more highly and that inertia costs are *not* real (simplest case)
    - If market wage is positively (negatively) correlated with inertia cost then relative to a no inertia benchmark welfare is higher (lower)
- See Handel et al. (AER:I forthcoming) on distribution of behavioral frictions by SES

The challenge (and frontier) II: Estimating the value curve

- Thus far we have talked about more or less data driven ways to identify behavioral frictions in health insurance markets
  - More work certainly needed!
- Key open question: When behavioral frictions exist, when and why are they quantitatively important for welfare? for policy?
  - Can imagine that depending on correlation between behavioral friction, costs, and willingness to pay, could get any type of rotation of demand curve relative to value curve
  - Similar point in intellectual history in the adverse selection literature. Now need to move beyond testing.

Estimating the value curve: a few initial thoughts

- I am not sure how to do this!
- Handel, Kolstad and Spinnewijn (2019)
  - Take frictions estimates from H&K "Health Insurance for Humans" + modeling assumptions to try to estimate Spinnewijn's value curve (vs demand curve)
- Presumably a key object is how the "value" curve varies with price
  - i.e. in addition to cost and demand curve, also want to know how "behavioral stuff" (e.g. choice of dominated plan) varies with price
  - If behavioral factors are flat with respect price is it less consequential for welfare analysis using demand vs value curve?
- Do you have to take a stand on the behavioral model?

Recap: key challenges (and opportunities!)

- Detecting departures from neoclassical model: data vs. modeling assumptions
- Welfare analysis: what do we call the error term?
  - Preference heterogeneity vs. mistakes?
  - How do we get away from ad hoc decisions / make it more data driven?
- Thus far we have seen work exploiting:
  - Changes in menus for different cohorts ("inertia")
  - Dominated choices
  - Ripe for additional work!
- Key challenge: how do we identify the value curve (i.e. the demand curve in the absence of any behavioral frictions)?

The behavioral public "cookbook"

- Approach 1: Specify a behavioral model (e.g. Handel's inertia cost + $H_{k,t-1}$)
- Approach 2: Measure preferences using undistorted choices
  - Active choice (Handel 2013, Chetty, Looney and Kroft 2016)
  - Unbiased group (popular group is experts: Dube, Gentzkow and Shapiro's "Do Pharmacists Buy Bayer?")
  - Connected to the sufficient statistics approach in EFC (2010).
    - Researcher does not need a fully-specified model of biases, which enables model structure like the ξ's for unobserved heterogeneity in EFC.
- Discussion q – why is approach 2 more challenging *for insurance*?