Beyond Statistics: The Economic Content of Risk Scores

Liran Einav (Stanford and NBER), Amy Finkelstein (MIT and NBER), Raymond Kluender (MIT), and Paul Schrimpf (UBC)

AEA Meeting, Boston
January 5, 2015

Overview

- IT and “big data” had a huge impact on markets in recent decades
- Insurance and credit market are leading examples
- Key observations:
    - Risk adjustment reimburses insurers based on predicted medical spending under a given contract
    - Realizations of predictions is partially driven by economic responses to a (potentially different) contract
    - Heterogeneity in such behavior would make a statistically “perfect” risk score in one context imperfect for another
- Implication: Strategic incentives for cream-skimming can still exist even with "perfect" risk scoring under a given contract.

Outline

- Simple model of consumer medical spending decisions
- Draw from Medicare Part D to present empirical evidence on central premise of paper:
  - Two dimensions of individuals heterogeneity are clearly visible:
    - Heterogeneity in health
    - Heterogeneity in response to price ("moral hazard")
  - Existing risk scores (predictions of utilization) do not capture heterogeneity in moral hazard (only in health)
- Use a stylized model in the context of Medicare Advantage to illustrate key implications:
  - With heterogeneity in behavioral response as well as health, private providers' strategic incentives for cream skimming can still exist even with risk scores that are "perfect" in a statistical sense.

Large literature on risk adjustment in health insurance

- Strong emphasis on statistical modeling techniques
- Attention to cream-skimming and "gaming" in the presence of *imperfect* predictions of individual risks
  - Glazer and McGuire (2000): optimal risk-adjustment to minimize cream skimming in presence of imperfect prediction
  - Recent empirical work analyzing provider strategic responses to imperfect risk scoring in Medicare Advantage (Brown et al. 2012, Newhouse et al. 2012)
- Focus has been on one-dimensional heterogeneity and imperfect risk scoring
  - In existing framework, all issues go away with “perfect” scoring capturing all residual heterogeneity *under a given contract*
- Key distinction with our work
  - Outcomes predicted by risk scores partially reflect (economic) choices and heterogeneity in behavioral response, cream-skimming incentives cannot be eliminated solely by statistics

Large literature on "moral hazard" in health insurance

- Primarily focused on average behavioral responses
- Focus here is on potential *heterogeneity* in behavioral response and its implications (in this case, for risk scoring)
- Builds on previous work on heterogeneity (and selection on) moral hazard (Einav, Finkelstein, Ryan, Schrimpf, and Cullen, 2013)
  - Same basic model
  - Focus on different conceptual issue (risk adjustment)

(Stylized) Theoretical Framework

- Individual defined by two-dimensional type (λ, ω)
  - $\lambda \geq 0$ denotes underlying health
  - $\omega \geq 0$ denotes price sensitivity of demand for medical care ("moral hazard")
- Individual chooses medical spending m to trade off health h and money y

$$u(m; \lambda, \omega) = \underbrace{\left[(m-\lambda) - \frac{1}{2\omega}(m-\lambda)^2\right]}_{h(m-\lambda;\omega)} + \underbrace{[y - c(m)]}_{y(m)}$$

  - Health h depends on underlying health λ and is increasing and concave in medical spending m
    - Concavity of h in m reflects diminishing returns to medical spending
  - Higher ω individuals have higher relative weight on health
  - $c(m)$ maps medical spending into out of pocket spending
    - $c \in [0, 1]$ indicates amount individual pays per dollar of m

(Stylized) Theoretical Framework (continued)

- Optimal spending m^* given by

$$m^*(\lambda, \omega) = \arg\max_{m \geq 0} u(m; \lambda, \omega)$$

- With linear contracts ($c(m) = c \cdot m$) yields the first order condition:

$$m^*(\lambda, \omega) = \lambda + \omega(1 - c)$$

- Individual spending as both a "level" term λ and a "slope" term ω
    - With no insurance ($c = 1$) spend λ
    - With full insurance ($c = 0$) spend $\lambda + \omega$

(Stylized) Theoretical framework (continued)

$$m^*(\lambda, \omega) = \lambda + \omega(1 - c)$$

- This ω term typically referred to as "moral hazard"
  - Reflect preferences over health and income, as well as how discretionary health spending is
  - Can be correlated with various components of health
- Key point: individuals have two-dimensional types (λ, ω), but only their spending $m^*(\lambda, \omega)$ is observed
  - Risk scores try to predict m^* using data from one environment
  - An individual's m^* in another contact will depend on their ω

Empirical Strategy

- Goal: Illustrate that individuals are heterogeneous both in their underlying health (λ) and in their response to insurance contracts (ω)
- Research design: exploit the famous "donut hole" in Medicare Part D prescription drug coverage
  - Insurance becomes discontinuously less generous on the margin
  - Examine behavioral response to this change in coverage across individuals

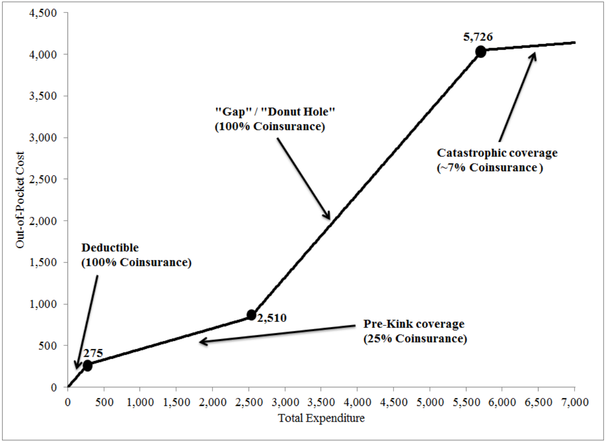
Setting

- Medicare Part D introduced in 2006, covering approximately 30M individuals (about 60% of Medicare beneficiaries)
  - Accounts for about 11% of total Medicare spending
- Enrollees can choose among different prescription drug plans offered by private insurers
  - Government sets a standard plan but actual plans often modify this
  - Medicare reimburses private plans as a function of the "Part D risk scores" for their enrollees, which predict drug spending as a function of demographics and prior medical diagnoses

Medicare Part D Risk Scores

- Centers for Medicare & Medicaid Services (CMS) started risk adjusting in 2004 to set reimbursement rates for Medicare Advantage
- Model uses medical diagnoses, age, and gender to predict medical expenditures
  - Designed to encourage specific coding of diagnoses and not reward coding proliferation
  - Aggregate 14,000 ICD-9 diagnosis codes to 189 hierarchical condition categories ("HCCs")
- Model was expanded to predict prescription drug spending for Part D
- HCC and demographic coefficients are added up to create risk score

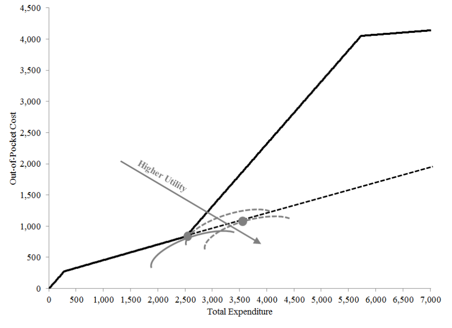
Standard Coverage in Medicare Part D (in 2008)



- In the data, price increases at the kink on average from 34 to 93 cents

Empirical Exercise

- Builds on Einav, Finkelstein, and Schrimpf (2013) - analyze spending response to non-linear contracts
- Standard economic theory: with convex preferences smoothly distributed in population, should see bunching at the convex kink



- Bunching provides opportunity to observe price response to prescription drug insurance contracts

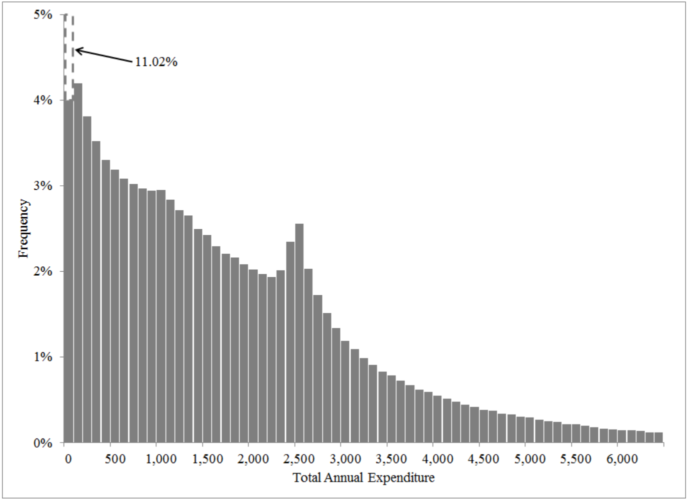
Data

- Use 20% random sample of all Part D insurees (2007-2009)
- Data include
  - Basic demographic information (e.g. age and gender)
  - Detailed information on plan characteristics
  - Detailed, claim-level information on utilization (2006-2010) both for prescription drugs (Part D) as well as in-patient, emergency room and (non-emergency) outpatient (covered by Part A and B)
  - Mortality through 2010

Sample and Summary Statistics

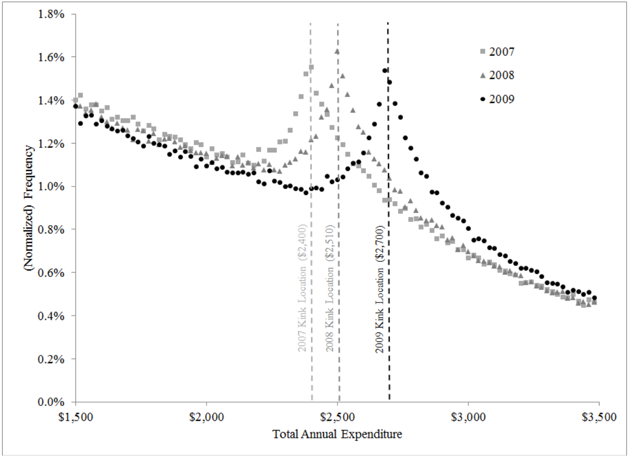
- Restrict sample to about one-quarter of full sample
  - Restrict to age 65+ with no low income subsidies (including Medicaid) in stand-alone PDPs
  - Focus on beneficiaries who were enrolled in Medicare at least one year (allows us to calculate risk scores)
- Final sample: 3.7M beneficiary-years (1.6M unique beneficiaries), average age of 76, about 2/3 female
- Average drug spending just over $1,900 dollars, roughly $800 paid out of pocket
- Spending is skewed: 5% spend nothing, median about $1,400, and 90th pctile around $4,000.
  - Kink hits at about the 75th pctile

Bunching at Kink I: 2008 Spending Distribution



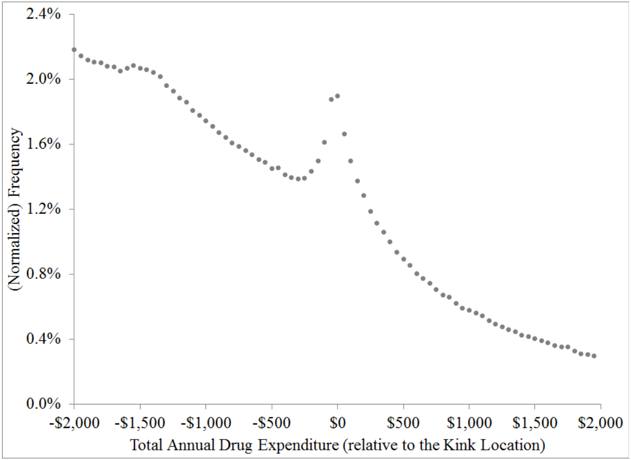
- From Einav, Finkelstein, and Schrimpf (2013).

Bunching at Kink II: Changes Across Years



- From Einav, Finkelstein, and Schrimpf (2013).

Bunching at Kink III: Pooling Across Years

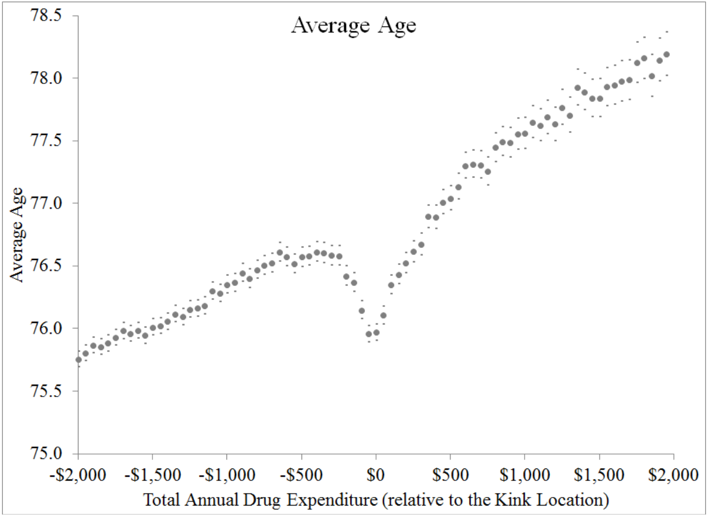


- From Einav, Finkelstein, and Schrimpf (2013).

Detecting Heterogeneity in Moral Hazard

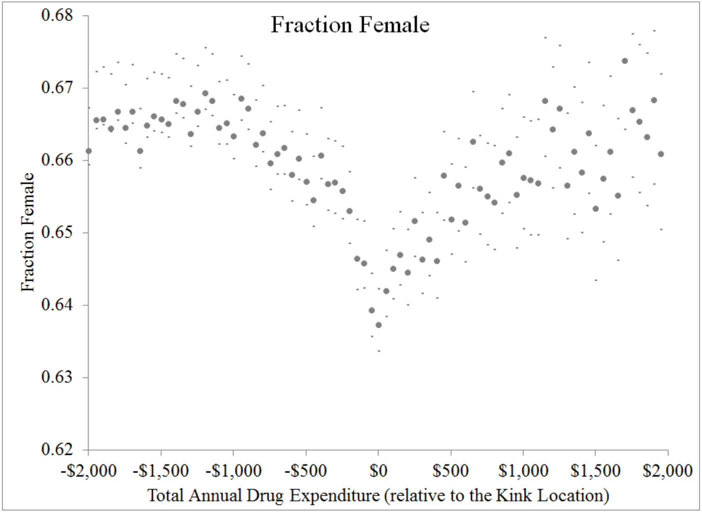
- Detect heterogeneity in the response to contract by documenting sharp changes in composition of sample around kink
- Over-representation of characteristics (e.g. being male, having diabetes) around the kink indicate that individuals with these characteristics have greater price sensitivity (c)
    - Likewise an individual characteristic that is under-represented at the kink are less responsive

"Bunchers" are younger

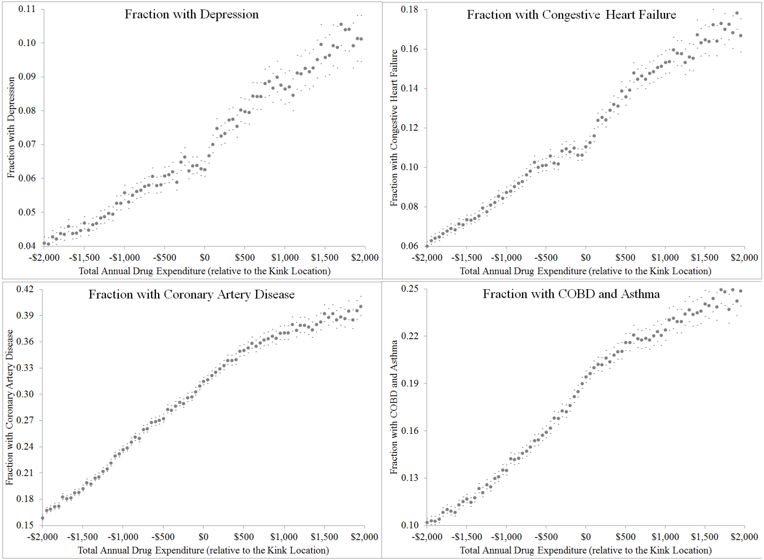


- Figure shows heterogeneity with age in both λ and ω

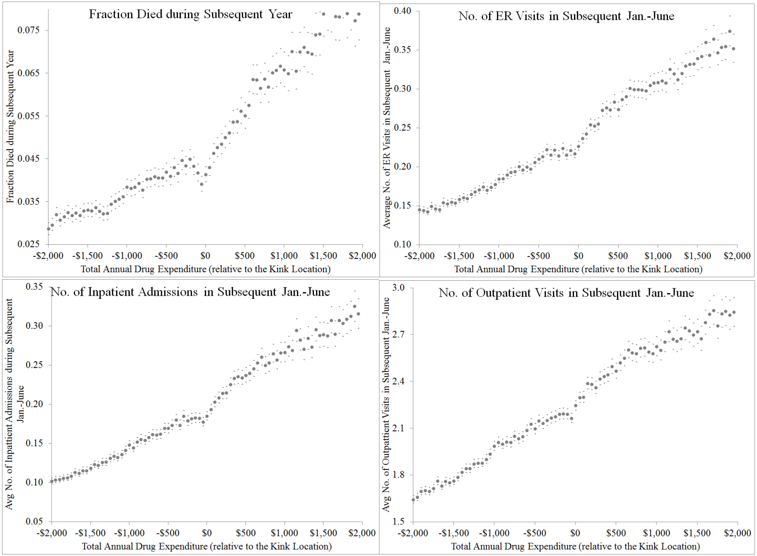
"Bunchers" are more likely to be male



"Bunchers" appear slightly healthier (HCCs)



"Bunchers" have better subsequent outcomes

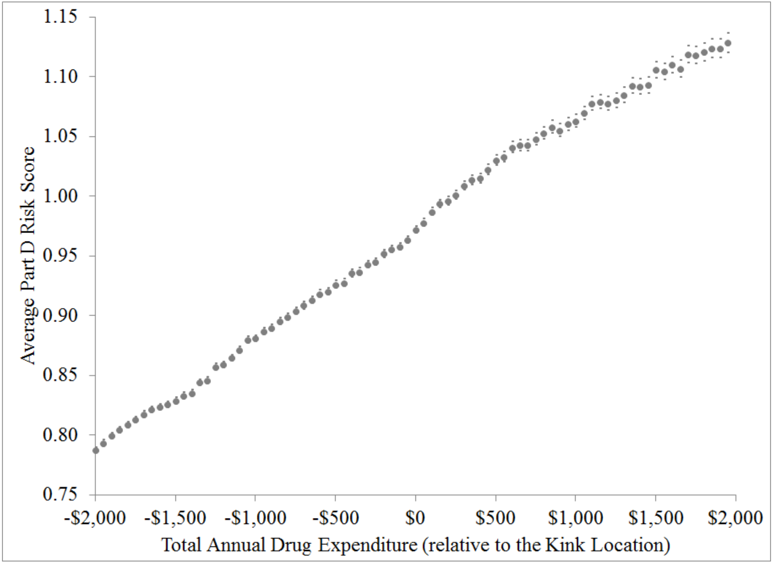


- We interpret this as additional evidence that individuals who are more

Recap

- First result: two-dimensional heterogeneity
  - Heterogeneity in health (λ): No surprise
  - Heterogeneity in moral hazard (ω): Males, younger, and healthier beneficiaries are more price sensitive
- In our context, results also suggest that ω and λ negatively correlated (Healthier have a larger behavioral response)
- Next result: risk scores do not capture this second dimension of heterogeneity

Risk Scores Reflect Only One Dimension



Risk scores are smooth through the kink

- Average risk score monotone in average annual spending
- Risk scores predict spending under observed contract
- Risk scores do not capture the behavioral responsiveness (ω) dimension of individual heterogeneity

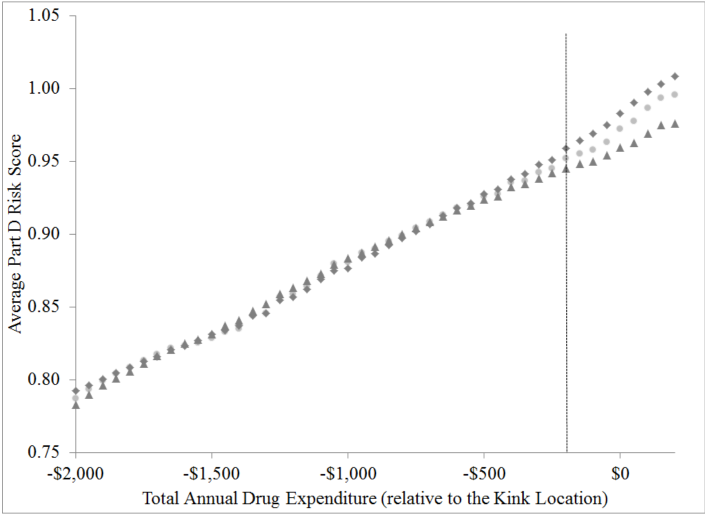
Side note: If bunchers are healthier why not show up in risk scores?

- We saw that bunchers tended to be healthier, but don't have lower predicted risk scores
- Two possible explanations:
    - Demographics that we saw changing at kink aren't quantitatively important in generating risk scores
    - Other components of risk score move in opposite direction, to offset
- Interpretation not important for our main point (current risk scores do not capture behavioral response)

Investigating offsetting effects

- Predict each component of risk score (age, gender, specific HCCs) by fitting line to left of kink (-$2,000 to -$200) and predicting in the (-$200 to +$200) kink range
- Split components into those that exhibit excess bunching (relative to kink) and those that exhibit dip in kink (relative to prediction)
- Then generate two sets of predicted risk scores
  - Use predicted values for components that bunch and actual for rest
  - Use predicted values for components that dip and actual for rest
- If components that bunch and components that dip do not do so in manner that is quantitatively important for risk scores, would expect two versions to lie close to one another

Offsetting effects



Primary Contributing Risk Score Components

	Incidence around the kink			Share of Risk-
	Actual	"Predicted"	Difference	Score difference
Top 10 components with positive kink incidence				
Chronic Obstructive Pulmonary Disease and Asthma	0.1908	0.1784	0.0124	20.50%
Diabetes with Complications	0.0908	0.0816	0.0091	19.13%
Breast, Lung, and Other Cancers and Tumors	0.0582	0.0520	0.0062	10.72%
Alzheimer's Disease	0.0203	0.0179	0.0024	9.32%
Diabetes without Complications	0.2020	0.1962	0.0058	8.45%
Esophageal Reflux and Other Disorders of Esophagus	0.2146	0.2082	0.0064	7.26%
Inflammatory Bowel Disease	0.0107	0.0093	0.0014	3.21%
Diabetic Retinopathy	0.0278	0.0237	0.0041	3.20%
Parkinson's Disease	0.0127	0.0119	0.0009	3.06%
Major Depression	0.0196	0.0187	0.0010	2.26%
Top 10 components with negative kink incidence				
Hypertension	0.6531	0.6735	0.0203	33.48%
Disorders of Lipoid Metabolism	0.7344	0.7530	0.0186	21.63%
Osteoporosis, Vertebral and Pathological Fractures	0.1730	0.1874	0.0144	13.13%
Open-Angle Glaucoma	0.0918	0.0999	0.0081	11.28%
Atrial Arrhythmias	0.1361	0.1460	0.0099	5.99%
Congestive Heart Failure	0.1117	0.1148	0.0031	5.40%
Thyroid Disorders	0.2525	0.2596	0.0071	2.64%
Coronary Artery Disease	0.3107	0.3116	0.0009	1.23%
Depression	0.0659	0.0668	0.0009	1.20%
Cerebrovascular Disease, Except Hemorrhage or Aneurysm	0.1513	0.1522	0.0009	0.59%

Recap

- Individuals vary significantly and predictably in their price response to coverage (ω)
- Risk scores do not reflect this heterogeneity
  - This is by design: They are only supposed to reflect spending in a particular contract environment
- This could have implications for cream-skimming incentives under alternative contracts
- We construct one potential example next

Illustrative Example in the (Stylized) Context of Medicare Advantage

- Medicare beneficiaries selecting inpatient and outpatient coverage (Parts A & B)
  - Option A: Public Plan–original Fee for Service
  - Option B: Private Medicare Advantage coverage
- Government reimburses private insurers based on beneficiaries' risk scores which predict spending under the public plan
- Private insurers observe beneficiary risk scores and decide which plans to offer conditional on risk scores to elicit truthful revelation of ω
- Beneficiary chooses between public and private options. If private, self-reports ω.

Stylized Example Preliminaries: Individual Spending Decision

- Use earlier model to represent an individual's behavior and spending. Recall:

$$m^*(\lambda, \omega) = \lambda + \omega(1 - c)$$

$$u^*(\lambda, \omega) = u(m^*(\lambda, \omega); \lambda, \omega) = y - c \cdot \lambda + \frac{1}{2}(1 - c)^2 \omega$$

- Assume the government offers a default contract which provides full insurance ($c = 0$)
  - Government spending is given by $r_i = \lambda_i + \omega_i$
  - Note: There is no residual uncertainty!
  - r_i is a "perfect" risk score under government contract

Stylized Example Preliminaries: Private Insurer Set Up

- Private monopolist offers a contract that competes to attract beneficiaries from the default contract
  - Has a technology to "detect" ω-related spending, and not cover it, but provides full coverage for the rest
  - Medical spending is given by λ_i
  - Government pays private insurer $g_i = r_i + s(r_i)$ for covering individual i
- Note: Very stylized example; an extreme form of a more realistic situation

Stylized Example Preliminaries: Total Surplus

- Firm profits π_i from covering an individual: risk adjusted transfer $g(r_i)$ minus the cost to the insurer of covering i (which are λ_i by assumption)
- Total surplus: sum of consumer surplus + producer surplus - government spending (and associated costs)

$$TS_i = u_i^* + \pi_i - (1+k)g_i$$

- $k \geq 0$ denotes the marginal cost of public funds

Setting Summary

	Public Coverage	Private Coverage
1. Individual medical spending	$\lambda_i + \omega_i$	λ_i
2. Individual optimized utilty (u_i*)	$y_i + 0.5 \cdot \omega_i$	y_i
3. Government spending (g_i)	$\lambda_i + \omega_i$	$\lambda_i + \omega_i + s_i$
4. Profits (π_i)	N/A	$\omega_i + s_i$
5. Total Surplus (TS_i)	$y_i - (1+k)\lambda_i - (0.5+k)\omega_i$	$y_i - (1+k)\lambda_i - k\omega_i - ks_i$

- Higher-ω individuals prefer the public coverage
- Private insurer has incentive to cover everyone, but especially to cream-skim higher-ω individuals
- Efficiency gain from allocating beneficiaries to private plans, especially higher-ω individuals
  - Choice of subsidy trades off efficient allocation vs. cost of public funds

Private Insurer's Problem

- Modeled as a standard optimal contracting model with incomplete information
- Observes risk score r (spending under the public contract)
- Offers a family of contracts $p(r, \omega')$ that depend on r and on a self-reported type ω'
- Beneficiary chooses contract
  - Contract $p(r, \omega')$ associated with beneficiary premium $p(r, \omega')$ and private insurer covers medical spending $\lambda' = r - \omega'$
  - Note that this is efficient amount of medical spending for type ω'
- Assume individuals know their true type (λ, ω) when choosing insurance plans
- Insurers design contracts to incentivize truthful revelation

Private Insurer's Problem (continued)

- Consider utility of beneficiary (λ, ω) from private contract $p(r, \omega')$
- Recall that $r = \lambda + \omega$ is observed, and individual medical spending under private contract given by $\lambda' = r - \omega' = \lambda + \omega - \omega'$
- Individuals' utility given by

$$
\begin{aligned}
u(m; \lambda, \omega) &= \underbrace{\left[(m - \lambda) - \frac{1}{2\omega}(m - \lambda)^2\right]} + \underbrace{[y - c(m)]} \\
u(\lambda, \omega; \omega') &= \left[(\omega - \omega') - \frac{1}{2\omega}(\omega - \omega')^2\right] + y - p(r, \omega')
\end{aligned}
$$

Private insurer's problem (continued)

- Private insurer's problem is to choose menu $p(r, \omega')$ in order to maximize

$$\max_{p(r,\omega)} \pi(r) = \int [p(r,x) + s(r) + x] \, dF_{\omega|r}(x)$$

subject to IC (truth-telling)

$$u(\lambda, \omega; \omega) = y - p(r, \omega) \geq u(\lambda, \omega; \omega') \ \forall \omega'$$

and IR (to opt into private coverage)

$$u(\lambda, \omega; \omega) = y - p(r, \omega) \geq y + \omega/2$$

Equilibrium Cream-Skimming

- The IC implies that $-1 - \partial p/\partial \omega = 0$, or that

$$p(r, \omega) = t(r) - \omega$$

  where $t(r)$ is the integration constant
- Substituting this schedule into the IR, we obtain $y - (t(r) - \omega) \geq y + \omega/2$, or equilibrium selection into private coverage for:

$$\omega \geq 2t(r)$$

- This results in cream-skimming, for every risk score r, of higher-ω beneficiaries
  - Movement toward the efficient allocation
  - Some fraction of beneficiaries still inefficiently covered by the public plan
  - Key Result: Cream-skimming is still there, even though risk scoring is "perfect"

Relationship to Monopolistic Pricing

- Equilibrium selection rule implies profit max given by

$$\max_{t(r)} \pi(r) = (t(r) + s(r)) \left[1 - F_{\omega|r}(2t(r))\right]$$

- Monopolist therefore sets $t^*(r)$ to solve the FOC

$$t^*(r) = \frac{1 - F_{\omega|r}(2t^*(r))}{2f_{\omega|r}(2t^*(r))} - s(r)$$

- Analogous to a textbook monopolist's pricing problem $\pi = (p - c)D(p)$, with price $t(r)$, marginal cost given by $-s(r)$, and demand $D(p)$ given by $1 - F_{\omega|r}(\cdot)$
  - Monopolist chooses $t(r)$ to trade off price vs. quantity
  - Key primitive is the demand curve, or $1 - F_{\omega|r}(\cdot)$
  - Private provider does not observe ω and cream-skimming incentive (profit obtained from higher ω beneficiary) exactly offset by increased incentive of higher ω beneficiary to remain in the public plan

Implications for Designing Risk Adjustment

$$\max_{t(r)} \pi(r) = (t(r) + s(r)) \left[1 - F_{\omega|r}(2t(r))\right]$$

- Government's instrument is the subsidy $s(r)$ which shifts monopolist's marginal costs
- Government would optimally set $s(r)$ to resolve a trade-off:
  - $(+)$ Higher $s(r)$ will get passed through to lower $t(r)$ (monopolist price to beneficiary)
  - (-) Higher $s(r)$ will cost more due to cost of public funds
- If $k = 0$, should set subsidy $s(r)$ high enough that private provider sets $t(r) = 0$ and everyone in private plan.
- With $k > 0$, optimal subsidy resolves tradeoff between more people in private plan and higher cost of public funds.
- Knowledge of primitives ($F_{\omega|r}(\cdot)$ and k) guide optimal choice of $s(r)$
  - Earlier analysis suggested that ω and λ are negatively correlated, at least around the donut hole, which may help identify $F_{\omega|r}$ (recall, $r = \omega + \lambda$)

Summary

- Central Takeaway: Risk scores are not only statistical objects; they are generated by economic behavior
  - May have consequences when risk scores are used "out of sample"
- Illustrated the point empirically in Medicare Part D and theoretically using a stylized model
  - Cream-skimming incentives arise even when risk scoring is "perfect" statistically (no residual uncertainty)
- Described implications for optimal risk adjustment
- An alternative: Move beyond a one-dimensional risk score and customize formula to specific applications
  - Requires a research design and an economic framework; predictive modeling is unlikely to suffice