Selection on Moral Hazard in Health Insurance

Liran Einav[1] Amy Finkelstein[2] Stephen Ryan[3]
Paul Schrimpf[4] Mark R. Cullen[5]

[1]Stanford and NBER

[2]MIT and NBER

[3]MIT

[4]UBC

[5]Stanford School of Medicine

May 18, 2012

Motivation

- Sources of market failure in insurance markets:
  - Selection (on health status)
  - Moral hazard (responsiveness of health spending to insurance coverage)
- Often analyzed separately
- Important interactions with implication for:
  - Monitoring
  - Offering higher cost sharing options

Overview

- Utility-maximizing model of health insurance plan choice and subsequent spending
  - Choices and spending determined by individuals'
    1. Health (λ)
    2. Moral hazard (ω)
    3. Risk aversion (ψ)
  - Spending $= \lambda + (1 - c)\omega$
- Data on health insurance options, choices, and utilization of Alcoa's employees
  - Panel
  - Change in set of insurance options
- Difference in difference estimates
  - Average level of moral hazard
  - Suggestive of heterogeneity in and selection on moral hazard
- Structural estimates
  - Recover joint distribution of health, moral hazard, and risk aversion
  - Quantify importance of each to choices and spending

Related Literature

- Modeling approach --- utility maximizing model of health plan choice that accounts for selection on health
  - Cardon and Hendel (2001); Bajari et. al. (2010); Carlin and Town (2010); Handel (2011)
  - New: focus on heterogeneity and selection on moral hazard
- (Quasi-)experimental estimates of effect of cost sharing on health spending
  - Manning et. al. (1987); Newhouse (1993); review by Chandra, Gruber, and McKnight (2010)
- Insurance choice and multi-dimensional heterogeneity
  - Risk aversion: Finkelstein and McGarry (2006); Cohen and Einav (2007)
  - Cognition: Fang, Keane, and Silverman (2008)
  - Bequest prefernce: Einav, Finkelstein, and Schrimpf (2010)
- Selection with heterogeneous treatment effects
  - Björklund and Moffitt (1987); Heckman, Urzua, and Vytlacil (2006)

Model

- Focus on three determinants of insurance choice
  1. Risk aversion (ψ)
  2. Health (λ)
  3. Moral hazard (ω)
- Two stage model
  1. Choose insurance plan based on ψ, ω, and F_λ
  2. Choose spending based on chosen plan, λ, and ω

Model: Spending

- Utility separable in health and spending:

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

  - m = spending
  - j = chosen plan
  - $h(m-\lambda;\omega)$ = monetized utility from health
  - $c_j(m)$ = out of pocket health expenditure
  - p_j = premium for plan j
  - $y(m)$ = residual income
- Chosen spending and realized utility

$$m^*(\lambda,\omega,j) = \arg\max_m u(m;\lambda,\omega,j)$$
$$u^*(\lambda,\omega,j) = \max_m u(m;\lambda,\omega,j)$$

- With linear coverage, i.e., $c_j(m) = (1-c_j)m$,

$$m^*(\lambda,\omega,j) = \max[0, \lambda + (1-c_j)\omega]$$

Model: Coverage Choice

- Coverage choice:

$$j^*(F_\lambda(\cdot), \omega, \psi) = \arg\max_{j \in J} - \int e^{-\psi u^*(\lambda, \omega, j)} dF_\lambda(\lambda)$$

  - J = set of available plans
  - ω and ψ known to agent
  - λ unknown, and has distribution F_λ
- Willingness to pay for more coverage increasing in risk aversion ψ, risk F_λ, and moral hazard ω

Setting and Data

- Employee-level data from 2003-2006 on U.S.-based employees of Alcoa, Inc.
- Data include:
  - The menu of health insurance options available to each employee
  - The premium associated with each option
  - Employees' choices
  - Employees' (and dependents) subsequent medical expenditure (claim by claim)
  - Rich demographics, including risk scores

Key Variation

- New set of health insurance options introduced beginning in 2004
  - Old benefits were relatively cheap and provided very generous coverage
  - New benefits are less generous and priced higher
  - New and old options primarily differ in cost sharing
  - Had to make an active choice---could not stay with old plan and no default option
- Focus on unionized hourly employees
  - Face new options only when labor contract expires. Different unions had contracts that expired in different years.
- Focus on 2003-2004
  - Premiums change and in 2006 some plans were completely dominated, yet chosen
  - Abstract from inertial behavior

Old and New Options

Single coverage (N=1,679)

	Original Plan Options			New Plan Options				
	1	2	3	1	2	3	4	5
Plan features:								
Deductible	1,000	0	0	1,500	750	500	250	0
OoP Max	5,000	2,500	1,000	4,500	3,750	3,500	2,750	2,500
Avg. Share OoP	0.580	0.150	0.111	0.819	0.724	0.660	0.535	0.112
Premium	0	351	1,222	0	132	224	336	496
Percent Choosing	3.3%	63.5%	33.2%	14.1%	0.0%	2.2%	37.8%	45.9%

Non-single coverage (N=5,895)

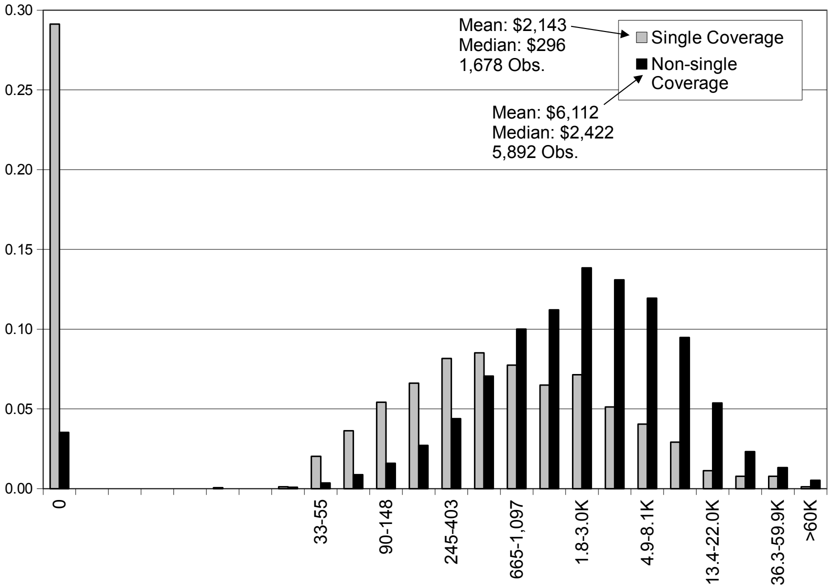
	Original Plan Options			New Plan Options				
	1	2	3	1	2	3	4	5
Plan features:								
Deductible	2,000	0	0	3,000	1,500	1,000	500	0
OoP Max	10,000	5,000	2,000	9,000	7,500	7,000	5,500	5,000
Avg. Share OoP	0.495	0.130	0.098	0.732	0.600	0.520	0.387	0.111
Premium	0	354	1,297	0	364	620	914	1,306
Percent Choosing	0.6%	56.1%	43.3%	3.9%	0.6%	1.8%	24.4%	69.3%

- 10% coinsurance rate between deductible and maximum out of pocket in all plans.

Summary Statistics

	All	Switched			
		in 2004	in 2005	in 2006	after 2006
Obs.	3,995	682	974	1,075	1,264
Age	41.3	44.5	39.7	38.3	43.3
Income	31,292	39,715	25,532	29,952	32,316
Tenure	10.2	15.5	8.2	5.7	12.7
Male	0.84	0.96	0.73	0.86	0.85
White	0.72	0.85	0.44	0.82	0.79
Single	0.23	0.21	0.25	0.23	0.22
Risk Score	0.95	1.06	0.91	0.86	1.01
Family Size	2.8	2.7	2.8	2.9	2.6
Medical Spending	5,283	5,194	5,364	5,927	4,717

Observed Distribution of Spending



Plan Transitions

Coverage Choice in 2003	Coverage Choice with *Old* Options in 2004		Coverage Choice with *New* Options in 2004	
	Highest	Other	Highest	Other
Highest	40.0%	0.5%	32.0%	15.8%
Other	0.6%	58.9%	27.8%	24.5%

Descriptive Evidence of Moral Hazard

	Single Coverage			Non-single Coverage		
	Count	Mean	Median	Count	Mean	Median
Original Options						
Highest Coverage	512	$3,130	$557	2,318	$6,635	$2,670
Other Coverage	1031	$1,795	$233	3,035	$5,768	$2,288
New Options						
Highest Coverage	62	$1,650	$447	375	$6,858	$2,630
Other Coverage	73	$560	$52	164	$3,405	$1,481

- More comprehensive coverage $\Rightarrow$ higher spending
- Evidence of selection and/or moral hazard

Average Moral Hazard

- Difference in difference specification:

$$y_{it} = \alpha_{g(i)} + \delta_t + \beta New_{g(i)t} + x_{it}\gamma + \epsilon_{it}$$

	2003-2004 Sample		2003-2006 Sample	
	OLS in levels	OLS in logs	OLS in levels	OLS in logs
$\widehat{\beta}$	-297.2	-0.35	-591.8	-0.175
S.E.	(753.7)	(0.19)	(264.2)	(0.12)
p-value	[0.70]	[0.08]	[0.034]	[0.17]
Elasticity	-0.07	-0.45	-0.14	-0.23
Observations	7,570	7,570	14,638	14,638

Suggestive Evidence of Heterogeneity

- Difficult to separate heterogeneity in moral hazard from:
  - Interaction between health status and moral hazard
  - Heterogeneity of treatment --- nonlinear health insurance coverage means that people with different health status face different treatments from change in options
- Nonetheless, evidence suggestive of moral hazard heterogeneity:
  - Diff-in-diff estimates differ among different groups of workers
    - ★ Larger for: old than young; sicker than healthier; female than male; low income than higher income
    - ★ But imprecise
  - Quantile diff-in-diff estimates range from zero for low quantiles to -1826 for the 0.9 quantile
    - ★ But we'd expect this to differ even with homogeneous moral hazard due to non-linearity

Econometric Specification

- Estimate model to quantify extent of heterogeneity in moral hazard and its importance
- Recall model:
  - Utility from health:

$$u(m;\lambda,\omega,j) = \left[(m-\lambda) - \frac{1}{2\omega}(m-\lambda)^2\right] + \left[y - c_j(m) - p_j\right]$$

  - Spending:

$$\begin{aligned} m^*(\lambda,\omega,j) &= \arg\max_m u(m;\lambda,\omega,j) \\ &= \max[0, \lambda + (1 - c_{j,s^*})\omega] \text{ (for piecewise linear coverage)} \end{aligned}$$

    where c_{j,s^*} is the coinsurance rate on the segment chosen
  - Choice:

$$j^*(F_\lambda(\cdot),\omega,\psi) = \arg\max_{j\in J} - \int e^{-\psi u^*(\lambda,\omega,j)} dF_\lambda(\lambda)$$

- Want to estimate joint distribution of ψ, ω, F_λ conditional on covariates X, $G(\psi,\omega,F_\lambda|X)$

Identification

- Given panel data on choices and spending with an exogenous change in the choice set, need to recover joint distribution of ψ, ω, F_λ
- Consider ideal data with infinite panel before and after change and, for simplicity, ignore truncation of spending at 0
- Assume:
  1. ψ_i and ω_i are constant over time
  2. $F_{\lambda,it}$ can vary with t, but the distribution of $F_{\lambda,it}$ before and after the choice set change is the same
  3. $F_{\lambda,it}$ identifiable from observations of $\{\lambda_{it}\}_{t=-\infty}^{\infty}$. e.g., rational expectation and λ_{it} ARMA
  4. $\mathrm{E}[c_{j,s^*}|i, \text{before}] \neq \mathrm{E}[c_{j,s^*}|i, \text{after}]$ almost surely

Identification

- Identify ω_i from

$$\mathrm{E}[m_{it}|i,\text{after}] - \mathrm{E}[m_{it}|i,\text{before}] = \omega_i \left(\mathrm{E}[c_{j,s^*}|i,\text{before}] - \mathrm{E}[c_{j,s^*}|i,\text{after}]\right)$$

- Given ω_i, can construct $\lambda_{it} = m_{it} + (1 - c_{j,s^*,i})\omega_i$, so distribution of $F_{\lambda,it}$ recoverable
- Choices identify distribution of ψ:

$$\mathrm{P}(j_{it}|J_{it},\omega_i,F_{\lambda,it}) = \mathrm{P}\left(\psi : j_{it} = \arg\max_{j\in J_{it}} - \int e^{-\psi u^*(\lambda,\omega_i,j)} dF_{\lambda,it}(\lambda) \middle| \omega_i, F_{\lambda,it}\right)$$

  - Probability of ψ being in $|J|$ regions, so can only parametricly identify $F(\psi|\omega, F_\lambda)$ unless J continuous

Parameterization

- $\lambda_{it} \sim$ shifted log normal, i.e.,

$$\log(\lambda_{it} - \kappa_{\lambda,i}) \sim N(\mu_{\lambda,it}, \sigma^2_{\lambda,i})$$

  - Shifted to rationalize zero spending
  - Higher μ_λ, σ_λ, or $\kappa_\lambda \Rightarrow$ worse health
- $\sigma^{-2}_{\lambda,i} \sim \Gamma(\gamma_1, \gamma_2)1\{\sigma^2_{\lambda,i} \leq \overline{\sigma}^2\}$, a truncated inverse gamma
- $\kappa_i \sim N(\overline{x_i}\beta_\kappa, \sigma^2_\kappa)$
- $\mu_{\lambda_{it}} = \overline{\mu_{\lambda,i}} + (x_{it} - \overline{x_i})\beta_\lambda + \epsilon_{\lambda,it}$
- $$\begin{pmatrix} \overline{\mu_{\lambda,i}} \\ \log \omega_i \\ \log \psi_i \end{pmatrix} \sim N\left(\begin{pmatrix} \overline{x_i}\beta_\lambda \\ \overline{x_i}\beta_\omega \\ \overline{x_i}\beta_\psi \end{pmatrix}, \begin{pmatrix} \sigma^2_{\bar{\mu}} & \sigma_{\mu,\omega} & \sigma_{\mu,\psi} \\ \sigma_{\mu,\omega} & \sigma^2_\omega & \sigma_{\omega,\psi} \\ \sigma_{\mu,\psi} & \sigma_{\omega,\psi} & \sigma^2_\psi \end{pmatrix} \right)$$
- $x =$ treatment group, coverage tier, age, gender, tenure, income, health risk score

Estimation

- No plan-specific *i.i.d.* error term makes maximum likelihood difficult
- MCMC Gibbs sampler
- Hierarchical model
  - Parameters $\theta_1 = \{\beta, \sigma, \gamma\}$
  - Latent variables $\theta_2 = \{\lambda_i, \mu_{\lambda,it}, \sigma_{\lambda,i}, \kappa_{\lambda,i}, \omega_i, \psi_i\}_{i=1,t=1}^{N,T}$
  - $F(\theta_1|\theta_2, \text{data}) = F(\theta_1|\theta_2)$ is tractable
  - Conditional on θ_1 can always find latent variables θ_2 that rationalize the data

Parameter Estimates: β

	μ_λ (Health risk)	κ_λ (Health risk)	$\log(\omega)$ (Moral hazard)	$\log(\psi)$ (Risk aversion)
Constant	6.11 (0.14)	-389 (73)	5.31 (0.24)	-5.57 (0.10)
2004 Time dummy	-0.12 (0.02)	--	--	--
Coverage tier				
Single	(omitted)	(omitted)	(omitted)	(omitted)
Family	0.19 (0.08)	57 (51)	-0.58 (0.18)	-0.88 (0.07)
Emp+Spouse	0.27 (0.09)	44 (53)	-0.66 (0.22)	-0.95 (0.07)
Emp+Children	0.24 (0.08)	185 (47)	-0.28 (0.21)	-0.91 (0.06)
Treatment group				
Switch 2004	-0.01 (0.07)	-278 (43)	-0.24 (0.11)	-0.31 (0.05)
Switch 2005	-0.10 (0.06)	-78 (38)	0.07 (0.12)	-0.23 (0.05)
Switch 2006	0.12 (0.07)	-94 (37)	0.01 (0.12)	-0.07 (0.05)
Switch later	(omitted)	(omitted)	(omitted)	(omitted)
Demographics				
Age	-0.01 (0.003)	-5 (1.8)	-0.01 (0.006)	0.01 (0.002)
Female	0.18 (0.08)	94 (39)	-0.08 (0.13)	-0.07 (0.06)
Job Tenure	0.002 (0.003)	-2.3 (1.6)	0.002 (0.004)	0.003 (0.002)
Income	0.003 (0.002)	6 (0.9)	0.001 (0.003)	-0.0003 (0.001)
Health risk score				
1st quartile ($<$ 1.119)	(omitted)	(omitted)	(omitted)	(omitted)
2nd quartile (1.119 to 1.863)	0.91 (0.07)	305 (59)	0.13 (0.29)	-0.41 (0.06)
3rd quartile (1.863 to 2.834)	1.48 (0.08)	242 (81)	1.79 (0.27)	-0.66 (0.06)
4th quartile ($>$ 2.834)	2.05 (0.09)	-416 (120)	3.38 (0.22)	-0.89 (0.07)

Parameter Estimates, Continued

Variance-covariance matrix

	$\overline{\mu_\lambda}$	$\log(\omega)$	$\log(\psi)$
$\overline{\mu_\lambda}$	0.20 (0.03)	-0.03 (0.04)	-0.12 (0.02)
$\log(\omega)$	--	0.98 (0.08)	-0.01 (0.03)
$\log(\psi)$	--	--	0.25 (0.02)

Additional parameters

σ_ϵ	0.33 (0.03)
σ_κ	290 (12)
γ_1	0.04 (0.004)
γ_2	15 (1.2)

Parameter Estimates: Implied Quantities

	$E[\lambda]$	ω	ψ
Average	4,340 (200)	1,330 (59)	0.0019 (0.00002)
Std. Deviation	5,130 (343)	3,190 (320)	0.0020 (0.00007)

Unconditional correlations

	$E[\lambda]$	ω	ψ
$E[\lambda]$	1.00	0.24 (0.03)	-0.36 (0.01)
ω	--	1.00	-0.15 (0.01)
ψ	--	--	1.00

Model Fit: Choices

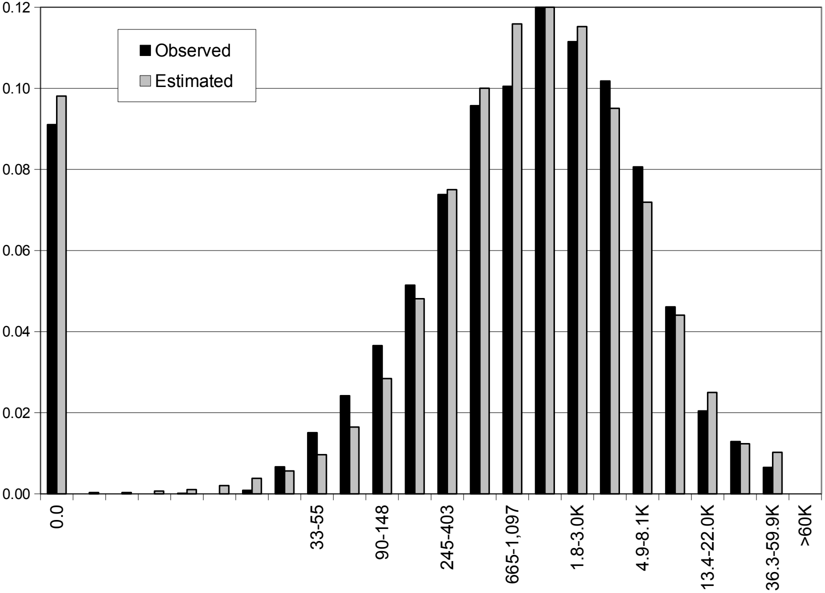
Original options (N = 6,896)

Plan	Data	Model
Option 1	1.2%	2.0%
Option 2	58%	57%
Option 3	41%	41%

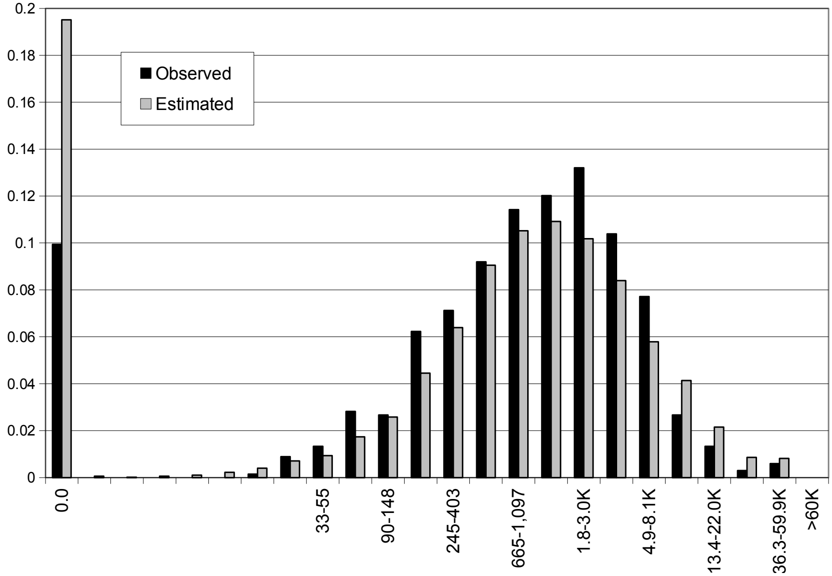
New options (N = 674)

Plan	Data	Model
Option 1	5.9%	5.0%
Option 2	0.5%	5.0%
Option 3	1.9%	1.0%
Option 4	27%	11%
Option 5	65%	76%

Model Fit: Spending with Old Options



Model Fit: Spending with New Options



Quantifying Moral Hazard

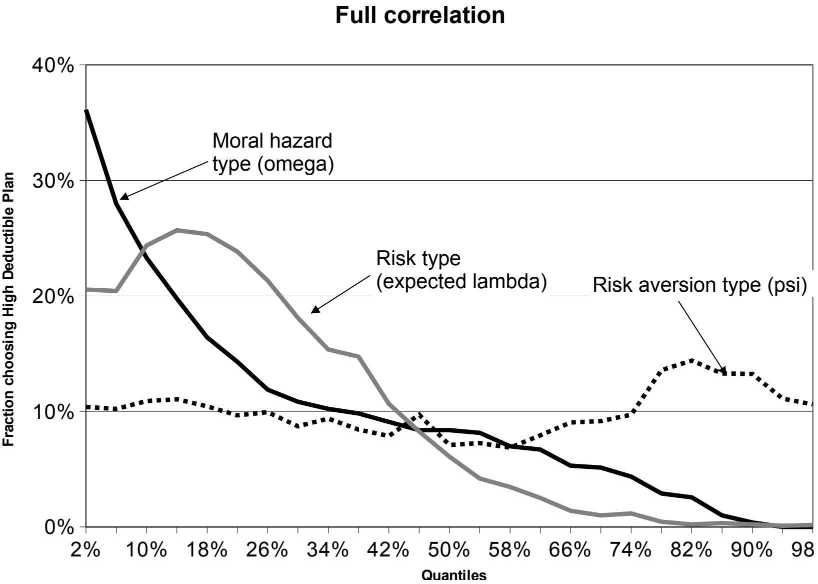
- Average ω = 1,330 or 30% of expected health risk, $\mathrm{E}[\lambda]$
- Standard deviation of ω = 3,200

			Percentiles				
	Mean	Std. Dev.	10th	25th	50th	75th	90th
Spending difference from no to high deductible	348	749	0	0	48	316	1,028
Spending difference from full to no insurance	1,273	3,181	0	86	310	1,126	3,236

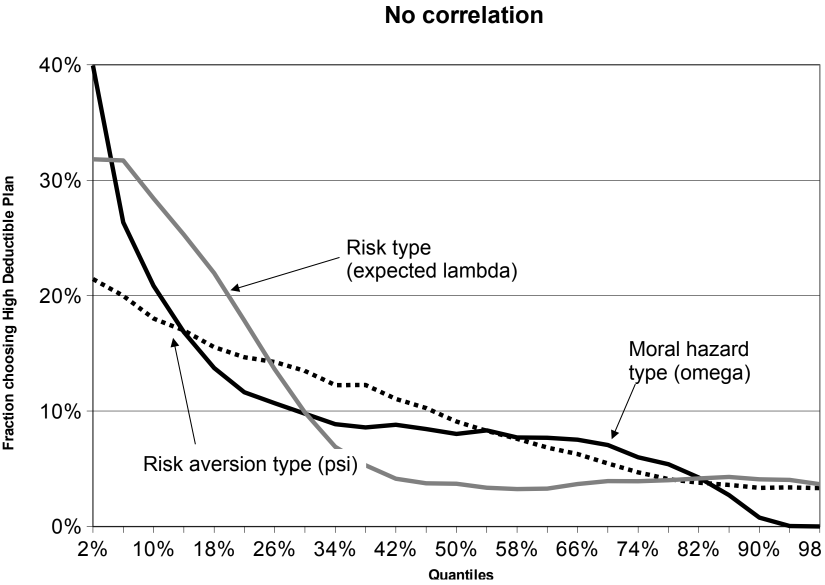
Selection on Moral Hazard, Health, and Risk Aversion

- To quantify relative importance of moral hazard, health status, and risk aversion for insurance plan choice we:
  - Simulate model with only plans 1 (high deductible) and 5 (no deductible) available and premiums set so that 10% of sample chooses high deductible plan
  - Report percent choosing the high deductible plan as a function of their percentile in the distributions of ω, $\mathrm{E}[\lambda]$, and ψ

Selection on Moral Hazard, Health, and Risk Aversion



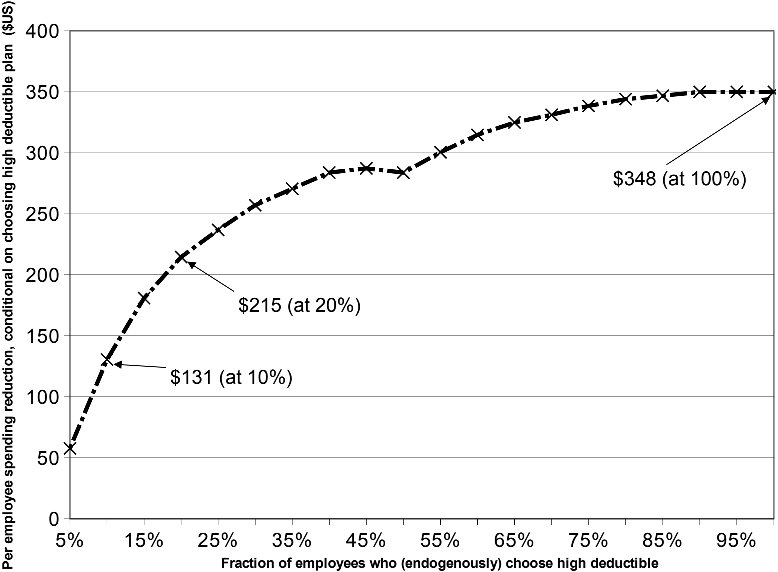
Selection on Moral Hazard, Health, and Risk Aversion



Implications for Spending

- Offering a plan with higher cost sharing is a common policy to reduce spending
- Selection on moral hazard implies that the effect of offering such a plan for those who endogenously choose it may be very different from the average effect of assigning such a plan
- To quantify:
  - Simulate model offering only plans 1 (high deductible) and 5 (no deductible) from the new options
  - Adjust premiums to shift the portion choosing the high deductible plan from 0 to 1
  - Report average spending reduction conditional on choosing high deductible plan as a function of percent choosing the high deductible plan

Implications for Spending



Illustrative Welfare Analysis

- Two possible source of efficiency gains:
  1. Improved screening: allow premium to depend (perhaps noisily) on F_λ and/or ω
  2. Improved monitoring: reimburse based on health realization λ instead of spending
- Monitoring and screening can interact --- monitoring changes expected utility, so affects pattern of selection and gains from screening
- To compare gains from screening vs monitoring:
  - Simulate model with high and no deductible plans
  - Premiums set so that incremental price of no deductible plan equals its incremental cost (consistent with perfect competition among providers of incremental coverage)
  - Measure welfare as consumers' surplus (certainty equivalent) plus providers' profits

Spending and Welfare Effects of Asymmetric Information

	Average equilibrium (incremental) premium	No deductible plan share	Expected spending per employee	Total welfare per employee
"Status quo": no screening or monitoring	1,568	0.90	5,318	normalized to 0
"Perfect screening": premiums depend on F_λ and ω	1,491	0.91	5,248	52
"Imperfect screening": premiums depend on ω (but not on F_λ)	1,523	0.88	5,265	34
"Perfect monitoring": contracts reimburse only "λ-related" spending	1,139	0.94	4,185	490
"Imperfect monitoring": perfect monitoring assumed for choice (but not for utilization)	1,139	0.94	5,327	25

Conclusions

- Empirical analysis of selection on moral hazard
- Selection on moral hazard about as important as selection on health status; both more important than selection on risk aversion
- Ignoring selection on moral hazard leads to an overestimate of the spending reduction from introducing a high deductible option
- Results specific to our sample; future work could look at selection on moral hazard in other contexts
- Our counterfactuals limited to set of contracts observed; could be interesting to look at contract design