High-Earner Lemons

Martin Karlsson, Technische Universität Darmstadt
Nicolas Ziebarth, Cornell University
Chicago, 5 January 2012

Outline

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Outline

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Adverse selection is potentially a serious problem in markets for health insurance.
- The theoretical literature suggests severe welfare losses might occur as a result (Rotschild & Stiglitz, 1976).
- The empirical literature has mainly been concerned with the identification of adverse selection:
  - A positive correlation between coverage and risk may be due to moral hazard, or adverse selection, or both.
  - Risk is not the only source of heterogeneity: other dimensions might give rise to advantageous selection.
  - Methods are typically simple and results not necessarily useful for policy purposes.
- Recent contributions use structural estimation to identify underlying preference and risk parameters (Einav et al, 2009).
- The identification problem remains, but estimates more useful and informative.

Aim of this paper

TECHNISCHE UNIVERSITÄT DARMSTADT

Aim of the paper: To estimate the distribution of risk and preference parameters among German holders of PHI.

- The German system is dual: 20 % of population can opt out of public system.
- Insured individuals may choose coinsurance rate, monetary deductible, and service package.
- The choice of parameters reveals information on individual risk and preferences.
- We consider two stages of the decision:
    1. **Ex ante**: The choice of coinsurance reveals information.
    2. **Ex post**: These coinsurance parameters rule out some ‘corner solutions’.
- Theoretical model: Additive CARA utility function.
- We do not achieve point identification, but the distribution of parameters is identified already from ex ante information – thus not contaminated by moral hazard.

The utility function

TECHNISCHE UNIVERSITÄT DARMSTADT

We use a simple additive CARA utility function:

$$U\left(c, m \mid n\right) = -\exp\left(-\gamma c\right) - \beta \exp\left(-\gamma\left(m - n\right)\right) \tag{1}$$

where

- m is the consumption of medical care services.
- c is the consumption of other goods and services: $c = y - p - z\left(m \mid \alpha, D\right)$
- $z\left(m \mid \alpha, D\right)$ is the out-of-pocket payment for someone with coinsurance rate α and deductible D.
- γ is the relative risk aversion.
- β is the preference for consumption of health care.
- n is the severity of illness (exponentially distributed with parameter θ).
- p is the insurance premium.

Optimal Consumption of health care

TECHNISCHE UNIVERSITÄT DARMSTADT

Inserting the budget constraint and solving, we get

$$m_{-}^{*}(n) = \frac{y - p + n - \frac{1}{\gamma}\ln\left(\frac{1+\delta}{\beta}\right)}{2+\delta}$$

$$m_{+}^{*}(n) = \frac{y - p - \alpha D + n - \frac{1}{\gamma}\ln\left(\frac{1-\alpha+\delta}{\beta}\right)}{2-\alpha+\delta}$$

...from which we get two 'corner solutions':

1. At $\tilde{n}$, $m_{-}^{*}(n) = 0$.
2. At $\bar{n}$, consumer is indifferent between $m_{-}^{*}(n)$ and $m_{+}^{*}(n)$.

Identifying information:

- If $0 < m < D$, then $m < m_{-}^{*}(\bar{n})$.
- If $m = 0$, then $\tilde{n} \geq 0$.
- If $m > D$, then $m \geq m_{+}^{*}(\bar{n})$

The first stage decision

TECHNISCHE UNIVERSITÄT DARMSTADT

Assuming individuals know their risk parameter θ, we consider first order conditions for the choice of α and D.

The ex ante expected utility equals

$$V = -\int_0^\infty f_n(n)\left[\exp\left(-\gamma c^*(n)\right) - \beta\exp\left(-\gamma\left(m^*(n) - n\right)\right)\right] dn. \qquad (2)$$

Necessary conditions for an optimum are

$$\frac{\partial V}{\partial \alpha} \geq 0, \frac{\partial V}{\partial D} \geq 0 \qquad (3)$$

From these FOC:s, we get partial identification of parameters:

1. **Case 1** ($\alpha < 1, D > 0$): Interval identification of γ, point identification of (θ, β).
2. **Case 2** ($\alpha = 1, D > 0$): Interval identification of θ, point identification of β.
3. **Case 3** ($\alpha < 1, D = 0$): Interval identification of θ, point identification of β.
4. **Case 4** ($\alpha = 1, D = 0$): Interval identification of β.

Distributional Assumptions

TECHNISCHE UNIVERSITÄT DARMSTADT

If we are willing to make assumptions concerning the distribution of γ, θ and β, we may estimate the parameters of this distribution.

Assumptions

- *The parameters* $\psi = (\gamma, \theta, \beta)'$ *take on a log-normal distribution.*
- *The logarithm vector* $\ln(\psi)$ *has mean vector* $\mu = \left(\mu_\gamma, \mu_\theta, \mu_\beta\right)'$ *and covariance matrix* Σ*:* $\ln(\psi) \backsim \mathcal{N}(\mu, \Sigma)$
- *The parameter* δ*, reflecting the shadow cost of health care consumption, is the same for everyone.*

We model the means $\mu = \left(\mu_\gamma, \mu_\theta, \mu_\beta\right)'$ as linear functions of characteristics X.
These characteristics should include any information available to the insurer.
In that case, the covariance matrix Σ captures the degree of asymmetric information.

Specifications

TECHNISCHE UNIVERSITÄT DARMSTADT

Estimation is done using maximum likelihood, with numerical integration over intervals.

We (will) consider three specifications:

1. Using *ex ante* information only:

$$L_i^1\left(\mu, \Sigma, \delta | \alpha_i, D_i, X_i\right) = \Pr\left(\alpha_i, D_i | \mu, \Sigma, \delta, X_i\right).$$

   These estimates cannot possibly be affected by moral hazard, but identifying information is weak.

2. Combining *ex ante* and *ex post* information:

$$L_i^2\left(\mu, \Sigma, \delta | m_i, \alpha_i, D_i, X_i\right) = \Pr\left(m_i | \alpha_i, D_i, \mu, \Sigma, \delta, X_i\right) \Pr\left(\alpha_i, D_i | \mu, \Sigma, \delta, X_i\right).$$

   This approach allows for much more precise estimates, but weaker identification.

3. Using only *ex post* information:

$$L_i^3\left(\mu, \Sigma, \delta | m_i, \alpha_i, D_i, X_i\right) = \Pr\left(m_i | \alpha_i, D_i, \mu, \Sigma, \delta, X_i\right).$$

   May serve as a test whether standard correlation tests deliver biased estimates.

Data

We use the German Socio-Economic Panel for estimation: it is a household survey that is representative for Germany.

The waves 2007 and 2008 contains information on (almost) everythhing we need:

- Insurance parameters (*p*, α, *D*),
- Income, age, gender, location.
- Consumption of medical care: tricky
  - We have (annual) days spent in hospital and (quarterly) visits to doctors.
  - Amounts spent had to be imputed using national averages.
- All information is self-reported.
- *p* assumed to be log-linear function of α and *D*.

After deleting individuals with item nonresponse, we were left with a sample of $2,363$ individuals.

Descriptive Statistics 1

TECHNISCHE UNIVERSITÄT DARMSTADT

Variable	Mean	Std. Dev.	N
m	1,526.146	3,834.939	2,363
α	0.977	0.071	2,363
D	283.667	601.5	2,363
p	4,932.711	2,126.193	2,363
y	31,913	21,507	2,363
year	0.485	0.5	2,363
age	47.058	10.802	2,363
sex	0.34	0.474	2,363
East	0.163	0.369	2,363

Descriptive Statistics 2

TECHNISCHE UNIVERSITÄT DARMSTADT

Table: Summary statistics of cost-sharing parameters

	D = 0			D > 0		
$\alpha = 1$	N = 1902 (56 per cent)			N = 1037 (31 per cent)		
		D	alpha		D	alpha
	Min	0	1	Min	10	1
	Mean	0	1	Mean	798	1
	Max	0	1	Max	6000	1
	SD	0	0	SD	808	0
$\alpha < 1$	N = 397 (12 per cent)			N = 31 (1 per cent)		
		D	alpha		D	alpha
	Min	0	0.50	Min	40	0.60
	Mean	0	0.79	Mean	668	0.83
	Max	0	0.99	Max	2500	0.95
	SD	0	0.11	SD	663	0.08

Determinants of insurance premium

TECHNISCHE UNIVERSITÄT DARMSTADT

	(1)	(2)
	$\ln p_{2007}$	$\ln p_{2008}$
α	0.124	0.153
	(1.01)	(1.17)
D	-0.000153***	-0.000138***
	(-8.45)	(-8.04)
age	0.0602***	0.0795***
	(12.29)	(16.38)
age2	-0.000452***	-0.000627***
	(-9.85)	(-13.74)
female	0.135***	0.144***
	(6.21)	(6.74)
East	-0.164***	-0.127***
	(-5.29)	(-4.09)
_cons	4.007***	3.467***
	(22.26)	(19.38)
N	1,592	1,561

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Influence of Observable Characteristics

TECHNISCHE
UNIVERSITÄT
DARMSTADT

	(1)	(2)	(3)
	ln γ	ln θ	ln β
δ	30.799***		
	(0.516)		
constant	-12.6114***	-10.4343***	2.9244***
	(0.069)	(0.072)	(0.038)
year	-0.4015***	-0.4003***	0.0216*
	(0.022)	(0.027)	(0.012)
age	0.0384	0.013	0.1549***
	(0.118)	(0.128)	(0.054)
female	0.0601**	0.1589***	0.109***
	(0.024)	(0.027)	(0.013)
east	0.0773*	0.0839*	-0.0074
	(0.046)	(0.051)	(0.024)

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Covariance Matrix

	(1)	(2)	(3)
	ln γ	ln θ	ln β
ln γ	0.1834***		
	(0.01)		
ln θ	-0.1117***	0.4622***	
	(0.01)	(0.013)	
ln β	-0.1143***	0.1877***	0.1191***
	(0.005)	(0.005)	(0.001)

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Conclusions

TECHNISCHE UNIVERSITÄT DARMSTADT

- The distribution of unobservables amongst privately insured may be detected from choice of coinsurance parameters.
- Further, these coinsurance parameters introduce non-convexities in the budget set, which also carry identifying information.
- Based on a simple CARA utility function, we estimated the distribution of unobservables, and their determinants.
- Combining ex ante and ex post information gives relatively precise information on unobservables, but point identification not achieved.
- Our estimates suggest there is considerable scope for selection, but the direction not yet clear.
- Clearly, the quality of the data is one main limitation of this study.