"Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," by Cardon and Hendel (*Rand*, 2001)

- More structural model of adverse selection.
- Again, find no evidence for adverse selection in health insurance.
- As we go, note issues about identification and think about how the structural model relates to the more reduced-form tests of Chiappori and Salanie.

Data

- Survey of 13,000 individuals from 1987.
- Data include: demographics, contract choice, contract menu, and health expenditure.
- Restrict the sample to 826 single individuals as families face complex choice sets. See Tables 1-3.

TABLE 1 **Weighted Means**

	All	Insured	Uninsured
Age	34.0	35.3	31.5
% female	44.7	50.4	33.3
Income	18,280	22,059	10,632
Total health care expenditure	901	1,019	660
Coinsurance rate (offered)	.12	.12	.13
Premium (offered)	360.1	170.6	745.2
Deductible (offered)	140	124	173
Total employees	638.53	878.01	153.38
Northeast (%)	20.40	23.09	14.93
Midwest	25.36	27.77	20.48
West	21.08	20.86	21.53
South	33.16	28.28	43.06
Hispanic (%)	7.35	4.88	12.38
Black (%)	12.05	9.28	17.68
Urban core (%)	32.27	33.37	30.04
Urban metropolitan area (%)	51.07	52.85	47.48
Nonurban (%)	16.66	13.78	22.45
Self-reported health state			
Excellent	34.3	38.0	26.8
Good	54.0	52.2	57.6
Fair	10.9	8.8	15.0
Poor	.8	.9	.6
Number of observations	826	516	310
Weight	100	67	33

Note: These are weighted means for the single population sample used in the estimation. Total health care expenditure includes the total cost of care no matter who paid for it and excludes insurance premia. Coinsurance, Premium, and Deductible

TABLE 2 **Means Conditional on Offered**

	Offered	Not Offered
Expenditure	990.5	718.4
Age	35.0	32.1
Income	21,774	11,216
Premium	152.2	—
Employer size	933.4	41.7
N	525	301
Weight	66.9	33.1

Note: Observations separated according to offered/not offered insurance by the employer. Employer size is the number of employees working for the employer.

TABLE 3 **Means Conditional on Offered and Insured**

	Offered		Not Offered	
	Insured	Uninsured	Insured	Uninsured
Expenditure	1,005	716	1,270	654
Age	35.2	29.5	35.9	31.7
Income	22,302	11,889	17,557	10,488
Premium	138.9	400.4	757.9	—
Employer size	924.9	1,093.3	11.2	45.2
N	492	33	24	277
Weight	65.5	3.4	3.4	29.6

Note: Similar to Table 2, but with observations further separated by insured and uninsured.

- Note:
  - Many uninsured.
  - Average expenditure much higher for insured. Why? demographics, costs (moral hazard), or adverse selection.
- Key exogenous variable: whether employer offers insurance benefits (i.e. assume employment choice is not driven by benefits). This allows identification between adverse selection and heterogeneity in moral hazard. Good assumption?

Model and Estimation

Consumers have utility $U(m_i,h_i)$ of money and health consumption, with $h_i=s_i+x_i$ so consumption can be higher either by more expenditure (x) or better health state (s).

Given a policy choice, ex-post consumption is given by:

$$\begin{aligned} U^*_{ij}(s_i) &= U^*(y_i, s_i, Z_j) = \max_{x_i} U(m_i, h_i) \\ s.t.\ m_i + C_j(x_i) &= y_i - p_j \end{aligned}$$

where some complications arise due to non-convexities of $C_j(x)$.

Model and Estimation (cont.)

When choosing a policy, an individual observes a signal ω_i which is correlated with s_i, and has some i.i.d individual-specific tastes for each option j which has no effect on the second stage behavior (what are these? why do we need it?). He then chooses j that maximizes:

$$V_{ij} = E(U^*_{ij}(s_i)|\omega_i) + a_{ij} = \int U^*(y_i, z_i, Z_j)dF(z_i|\omega_i, D_i) + a_{ij}$$

Q: What is the key parameter? What would be here the parallel to the Chiappori-Salanie test?

Empirical Model

Approximate the utility function with a quadratic form.

$$U(m_i, h_i) \approx \phi_1 m_i + \phi_2 h_i + \phi_3 m_i \cdot h_i + \phi_4 m_i^2 + \phi_5 h_i^2.$$

Assume $C_j(x)$ is linear with deductible. This makes it non-convex. Thus, they solve for each of the pieces, as well as the two corner solutions, and come up with the global maximum.

The estimates also give them price elasticities at the point of consumption, which are important to get idea about the extent of moral hazard.

The first stage is estimated assuming that utility is of the CARA from (this has no effect on second stage).

- s_i is assumed log-normal (can vary with demographics), i.e. $s_i = - exp(k(D_i)+\omega_i+\varepsilon_i)$ with both and ω_i and ε_i normal. So the key is to assess the relative variance of ω_i and ε_i.
- The a_{ij}'s are type I extreme value (logit) for convenience.

Estimation and Results

- GMM: Two types of moment conditions:
    - $P_{ij}(\theta,D_i) - I_{ij}$. Can think about it like a simple logit model with individual data.
    - $P_{ij}(\theta,D_i)E(x_{ij}(\theta,D_i)|I_{ij}=1) - I_{ij}x_{ij}$. Can think about it like a second-stage equation in a selection model. They compute the expectations by numerically integrating over the two-dimensional region.

- Combine this with a set of instruments which are unrelated to the errors. They use demographics (age, sex, race) as instruments.

- Results: Table 4. Low and insignificant value for σ_ω, i.e. little adverse selection. Column 1 is an interesting way to convince that it is not the structure that makes this: once demographics are omitted, they do find as if adverse selection exists.

- Table 5, 6 and Figures 1,2: reasonable fit.

TABLE 4 **Main Estimates**

	Parameter	Estimate (1)	t-statistic	Estimate (2)	t-statistic	Estimate (3)	t-statistic
1	σ_ω (signal)	.52	2.50	.12	.43	.06	.11
2	σ_ν (noise)	.99	2.87	1.75	7.00	1.95	4.22
3	ϕ_1 (M)	1	—	1	—	1	—
4	ϕ_2 (H)	−.94	−.22	−.70	−.23	−.61	−.20
5	ϕ_3 (M·H)	.04	.42	.11	1.40	.36	.45
6	ϕ_4 (M^2)	−.001	−.70	−.015	−13.12	−.004	−.17
7	ϕ_4 (H^2)	−7.27	−.64	−6.39	−1.13	−4.70	−.44
8	Age			−.05	−1.82	−.02	−1.07
9	Age^2			.001	3.24	.001	2.74
10	Female			.67	1.95	1.15	2.53
11	Northwestern			−.17	−.56	−.03	−.61
12	Nonmetro			−.57	−1.33	−.56	−1.22
13	Black			1.28	2.28	1.36	2.83
14	Clerical			−.23	−.32	.04	.92
15	Constant	−1.07	−1.33	−2.37	−3.26	−3.97	−4.95
16	HMO coin	2.55	1.61	2.66	3.15	2.24	3.83
17	$r(-e^{-rV})$					.04	.93
$J \sim \chi^2$		52.16		29.79		43.78	

Note: The first two rows present the estimated standard errors of the signal and of the noise around the signal. The next five rows present the coefficients of the quadratic utility function (with the linear term in M normalized to one). Rows 8 to 13 present the demographics in the function $K(D_i)$, i.e., the deterministic component of the health state. Row 15 presents the estimated HMOs "fictitious" coinsurance. Row 16 presents the CARA transformation parameter. t-statistics are based on a covariance matrix of the estimators computed using numerical gradients for a 10% increase in each parameter.

TABLE 5 **Insurance Choice Model Fit**

	Observed			
	Uninsured	Non-HMO	HMO	Total
Predicted				
Uninsured	203	92	12	307
Non-HMO	100	292	38	430
HMO	7	29	53	89
Total	310	413	103	826

Correct predictions: 66.3%

Note: Cells present the number of observation that fall into each specific predicted/observed pair. Observations in the diagonal are those with insurance status correctly predicted by the model.

TABLE 6 **Expenditures Goodness of Fit**

	Predicted	Actual
Mean expenditure	749	980
Insured	901	1,112
Uninsured	496	661
Price elasticity	.18	".2"
Income elasticity	.51	".2"

Note: Comparisons of model predictions with sample values. "Actual" elasticities are those from the RHIE.

FIGURE 1

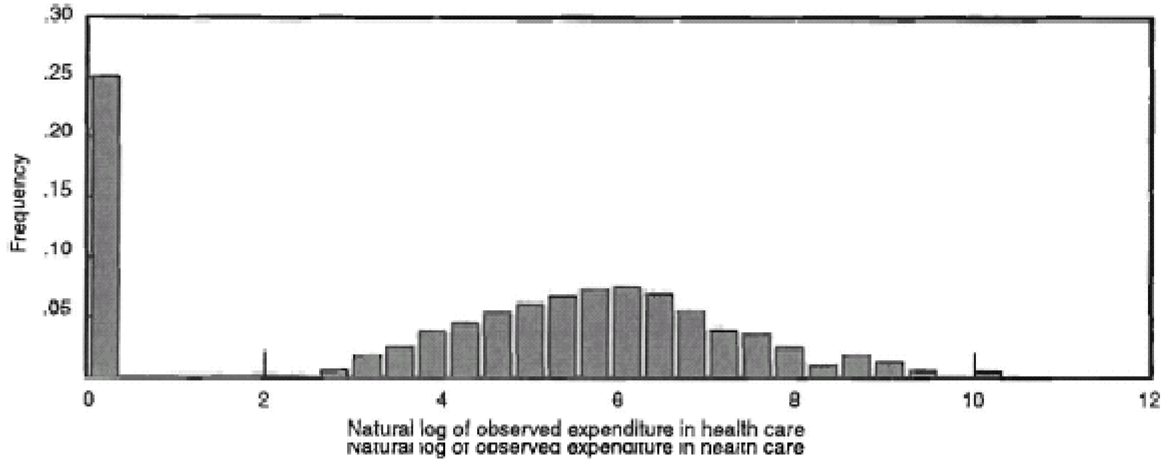
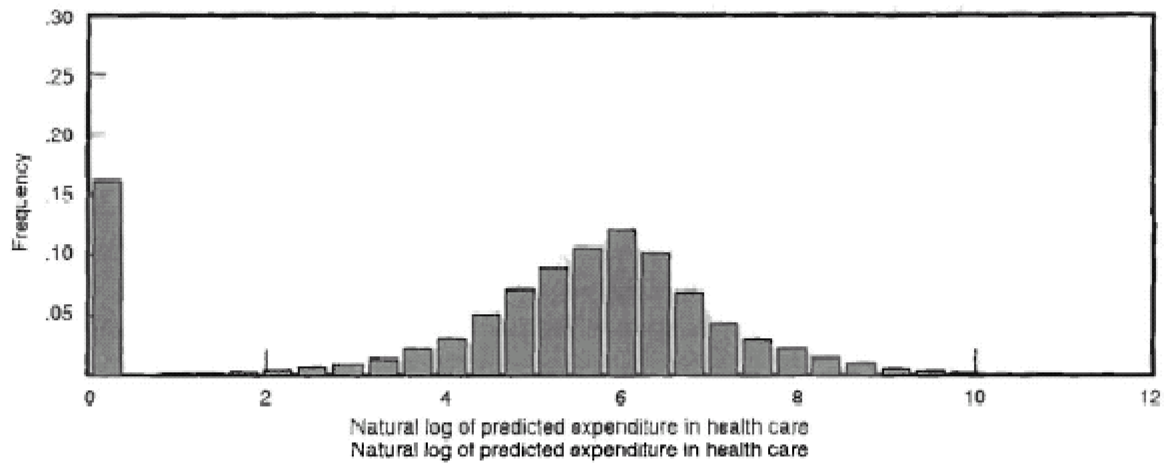


FIGURE 2



“Estimating Risk Preferences from Deductible Choice,” by Cohen and Einav (AER, 2007)

- Deductible choice in auto insurance.
- More structural approach compared to previous stuff (but simple structure).
- Cohen (2005) already found adverse selection in this data (using aversion of the Chiappori and Salanie probit test). Here we assume adverse selection, and try to use it to say something about risk aversion.

Questions:

- How big it is (most other evidence is based on experiments, surveys, financial decisions, and lotteries in TV shows)?
- How much does it vary across people, and how does it vary with observables (very little other evidence)?
- How is it related to risk? Is it indeed negatively correlated as suggested by the indirect evidence of Finkelstein-McGarry?
- How does it affect optimal insurance contracts?

Answers:

- Very large for certain fraction, but highly heterogeneous.
- Females more risk averse. Other reasonable covariates. No systematic variation with wealth/income.
- Strong positive correlation. A bit strange, but fairly robust. Will come up with potential stories later.
- Unobserved heterogeneity in risk aversion more important than that in risk.

Good application to do this, as little else that can affect one's deductible decision (compare to Cardon-Hendel, Finkelstein-McGarry, or to Chiappori-Salanie decision whether to buy insurance).

Model

All contracts offered (see later) are premium-deductible combinations.

Assumptions:

- No moral hazard.
- Any claim that's worth filing under low deductible is worth filing under high deductible (fairly consistent with the data).
- Claims are generated by a Poisson process with rate λ_i.
- We think about the length of the policy as very small: can prorate in reality, allows focus on static risk-taking, and can accommodate truncated policies.

Expected utility for individual i from contract (p_i,d_i) is (all i-specific):

$$v(p,d)=u(w-pt)-\lambda t[u(w-pt)-u(w-pt-d)]$$

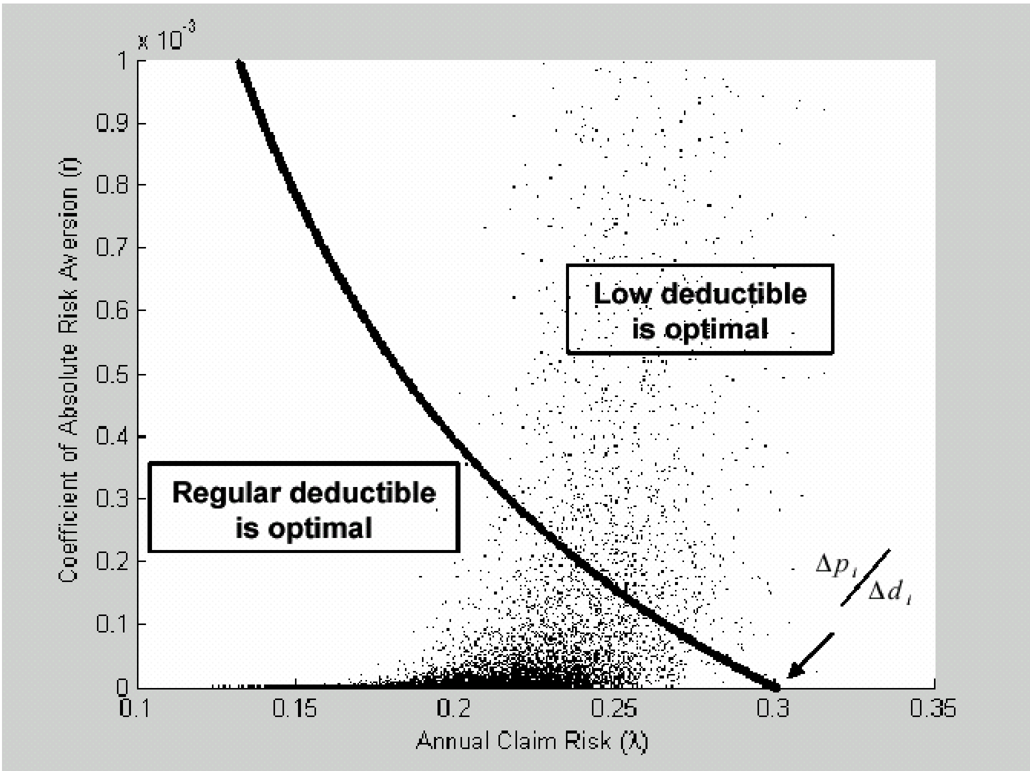
Look for the individual who is indifferent between two contracts. After taking limits, as t goes to zero, we get:

$$u'(w)\Delta p=\lambda(u(w-d^l)-u(w-d^h))$$

Approximate the utility function with a Taylor expansion: we get

$$r \equiv \frac{-u''(w)}{u'(w)} \approx \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\overline{d}}$$

where r is the coefficient of absolute risk aversion at wealth level w.



Data

- Full access to the files of a big Israeli auto insurance company from 1995 to 1999 (first five years of the company's operation).
- Restrict attention to choices of only new customers (a. potential selection issues with stayers; b. not obvious that stayers really make a new choice every time they renew).
- Total of 105,800 policies. Account for about 7% of the Israeli auto insurance market. For each policy, observe demographics, car attributes, driving experience, and, of course, the set of options, and the choice.
- Potential selection: the company is new and offers a different concept. We may select on less risk averse people or more price sensitive.
- We look at comprehensive coverage (non-mandatory, but chosen by a big fraction). The relevant claims are only those for which a deductible applies (so not at fault accidents, total-loss, and thefts are not counted). Radio, windshield, towing are covered separately.
- Less concerned about ex-post moral hazard of not filing a claim because of the features of the Israeli system.
- Just to have in mind: $1 US = 3.5 NIS on average.

The contract menu

Each customer was offered four contract choices:

- “regular” (similar to the deductible levels offered by other insurers and most policyholders chose it). The regular premium varied across individuals, and was some deterministic function (unknown to us) of all the characteristics of the policyholder. For regular premia which were not too high (see later), the level of the regular deductible was set at 50% of the (regular) premium.
- “low” deductible, set at 60% of the level of the regular deductible, with a premium equal to 106% of the regular premium.
- “high” and “very high” deductibles: rarely chosen, and almost not used for the analysis.

Those multipliers that convert the “regular” contract to any of the other contracts were fixed across individuals and over time (looks a bit stupid from a pricing standpoint ... which is why we focus on demand side).

All those contracts were subject to a uniform deductible cap, which varied with the choice, but not across people.

Sources of variation, which are independent of observables:

- Experimentation in the first year with the ratios.
- Shifts in the cap over time.

Example:

- Suppose that for a given individual the company's formula yielded a "regular" premium of 2,000 NIS. Suppose the deductible cap is 1,500 NIS (i.e. it's not binding). The menu would be: (2,000; 1,000), (2,120; 900), (1,750; 1,800), and (1,600; 2,600), respectively.
- Suppose a different individual with a higher "regular" premium of 4,000 NIS. The cap would be now binding. The menu would be: (4,000; 1,500), (4,240; 1,350), (3,500; 2,700), and (3,200; 3,900).

Table 1: Summary Statistics – Covariates

	Variable		Mean	Std. Dev.	Min	Max
Demographics:	Age		41.137	12.37	18.06	89.43
	Female		0.316	0.47	0	1
	Family	Single	0.143	0.35	0	1
		Married	0.779	0.42	0	1
		Divorced	0.057	0.23	0	1
		Widower	0.020	0.14	0	1
		Refused to Say	0.001	0.04	0	1
	Education	Elementary	0.016	0.12	0	1
		High School	0.230	0.42	0	1
		Technical	0.053	0.22	0	1
		Academic	0.233	0.42	0	1
		No Response	0.468	0.50	0	1
	Emigrant		0.335	0.47	0	1
Car Attributes:	Value (current NIS)[a]		66,958	37,377	4,000	617,000
	Car Age		3.952	2.87	0	14
	Commercial Car		0.083	0.28	0	1
	Engine Size (cc)		1,568	385	700	5,000
Driving:	License Years		18.178	10.07	0	63
	Good Driver		0.548	0.50	0	1
	Any Driver		0.743	0.44	0	1
	Secondary Car		0.151	0.36	0	1
	Business Use		0.082	0.27	0	1
	Estimated Mileage (km)[b]		14,031	5,891	1,000	32,200
	History Length		2.847	0.61	0	3
	Claims History		0.060	0.15	0	2
Young Driver:	Young		0.192	0.39	0	1
	Gender	Male	0.113	0.32	0	1
		Female	0.079	0.27	0	1
	Age	17-19	0.029	0.17	0	1
		19-21	0.051	0.22	0	1
		21-24	0.089	0.29	0	1
		>24	0.022	0.15	0	1
	Experience	<1	0.042	0.20	0	1
		1-3	0.071	0.26	0	1
		>3	0.079	0.27	0	1
Company Year:	First Year		0.207	0.41	0	1
	Second Year		0.225	0.42	0	1
	Third Year		0.194	0.40	0	1
	Fourth Year		0.178	0.38	0	1
	Fifth Year		0.195	0.40	0	1

Table 2A: Summary Statistics – Menus, Choices, and Outcomes

	Variable		Obs	Mean	Std. Dev.	Min	Max
Menu:	Deductible (current NIS)[a]	Low	105,800	875.48	121.01	374.92	1,039.11
		Regular	105,800	1,452.99	197.79	624.86	1,715.43
		High	105,800	2,608.02	352.91	1,124.75	3,087.78
		Very High	105,800	3,763.05	508.53	1,624.64	4,460.13
	Premium (current NIS)[a]	Low	105,800	3,380.57	914.04	1,324.71	19,239.62
		Regular	105,800	3,189.22	862.3	1,249.72	18,150.58
		High	105,800	2,790.57	754.51	1,093.51	15,881.76
		Very High	105,800	2,551.37	689.84	999.78	14,520.46
	$\Delta p/\Delta d$		105,800	0.328	0.06	0.3	1.8
Realization:	Choice	Low	105,800	0.178	0.38	0	1
		Regular	105,800	0.811	0.39	0	1
		High	105,800	0.006	0.08	0	1
		Very High	105,800	0.005	0.07	0	1
	Policy Termination	Active	105,800	0.150	0.36	0	1
		Canceled	105,800	0.143	0.35	0	1
		Expired	105,800	0.707	0.46	0	1
	Policy Duration (years)		105,800	0.848	0.28	0.005	1.08
	Claims	All	105,800	0.208	0.48	0	5
		Low	18,799	0.280	0.55	0	5
		Regular	85,840	0.194	0.46	0	5
		High	654	0.109	0.34	0	3
		Very High	507	0.107	0.32	0	2
	Claims per Year[b]	All	105,800	0.245	0.66	0	198.82
		Low	18,799	0.309	0.66	0	92.64
		Regular	85,840	0.232	0.66	0	198.82
		High	654	0.128	0.62	0	126.36
		Very High	507	0.133	0.50	0	33.26

Empirical Model and Estimation

Our object of interest is to estimate the joint distribution of (λ_i, r_i) in the sample, conditional on observables. The benchmark formulation assumes that (λ_i, r_i) follow a bivariate lognormal distribution. Thus, we can write the model as

$$\ln \lambda_i = x_i'\beta + \varepsilon_i$$

$$\ln r_i = x_i'\gamma + \upsilon_i$$

$$\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \stackrel{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{bmatrix} \right)$$

The interesting spin of the model is that neither λ_i nor r_i is directly observed, so they are treated as latent variables. We only observe two (integer) variables (the claims and the deductible choice) which are related to these two unobserved components.

Thus, to complete our empirical model we need to specify the relationship between those observed variables and the latent ones. This is done by making the two structural assumptions described before:

$$claims_i \sim Poisson(\lambda_i t_i)$$

$$\Pr(choice_i = 1) = \Pr\left(r_i > \frac{\frac{\Delta p_i}{\lambda_i \Delta d_i} - 1}{\bar{d}_i} \right) = \Pr\left(\exp(x_i'\gamma + \upsilon_i) > \frac{\frac{\Delta p_i}{\exp(x_i'\beta + \varepsilon_i)\Delta d_i} - 1}{\bar{d}_i} \right)$$

The last equation illustrates why the deductible choice is more than just a Probit regression. Both υ_i and ε_i (through λ_i) enter the right hand side, so we need to integrate over the two-dimensional region that predicts a choice of low deductible.

How would one estimate it:

- Maximum Likelihood (GMM, in the spirit of Cardon-Hendel, would be similar):

$$L(claims_i, choice_i|\theta) = \Pr(claims_i, choice_i|\lambda_i, r_i)\Pr(\lambda_i, r_i|\theta)$$

  Problem: need to integrate over the two dimensions for each individual separately.
- Gibbs sampler, and data augmentation.

Digression: Bayesian econometrics and Gibbs sampler

Bayesian econometrics:

Based on decision theory. The basic idea is that we think about the parameters θ has random variables. We have a prior about them, and we use the data to obtain a posterior. It is closely related to maximum likelihood because:

$$Pr(\theta|Z) = Pr(Z|\theta)Pr(\theta) \,/\, Pr(Z) \;\; \alpha \;\; Pr(Z|\theta)Pr(\theta)$$

Much focus is on “conjugate priors.”
In most cases, try to have a “flat prior” so the prior has little effect on the results.
For most standard models, a Bayesian analysis will essentially provide the same estimators as classical econometrics (e.g. OLS).

See example.

Gibbs sampler (Markov chain Monte Carlo):

Instead of doing the above all at once for all the parameters of the model, we can iterate over the parameters.
This is attractive if finding a posterior for all the parameters together is a mess, but finding a posterior for each one separately, conditioning on the others, is easy.

The basic idea: suppose we have three parameters to estimate, $\theta_1, \theta_2, \theta_3$. Instead of formulating the posterior as above, we do the following:

- Make up initial values for $\theta_1, \theta_2, \theta_3$
- Take a random draw for θ_1 from its posterior, conditional on the data and on the values of θ_2, θ_3.
- Take a random draw for θ_2 from its posterior, conditional on the data and on the most recent values of θ_1, θ_3.
- Take a random draw for θ_3 from its posterior, conditional on the data and on the most recent values of θ_1, θ_2
- Repeat the whole process long enough.

It has been shown that after a while (no real way to know when; use “eye test”) the draws will converge to draws taken from the joint posterior distribution (the draws are not independent but follow a Markov chain). Thus, we can drop the first chunk of draws and think of the reminder as draws from the posterior distribution. We can then compute posterior means, standard deviations, or anything else we are interested in.

Data augmentation:
Treat some of the latent variables as if they are part of the parameter space. They are not identified, so there is no point in reporting them, but draws from them are used to compute draws for the parameters we are interested in.

Back to the paper: Estimation of the model

Things are completely standard once we know r_i and λ_i for each individual.

Thus, we use Gibbs sampler and data augmentation to estimate the model. We treat the set of r_i and λ_i as "nuisance parameters." Given λ_i, the posterior for r_i is just a truncated lognormal. Given r_i the posterior for λ_i is a bit more complicated (truncated normal, with additional information coming from the number of claims), but it's still not too difficult to take a draw from it. Every iteration of the Gibbs, we take new draws for each individual.

The tables report the mean of the posterior and the empirical standard deviations.

Drawback: strongly rely on the distributional assumptions, and require us to impose priors.
Advantage: simple, relatively easy, relatively fast, and does not rely on asymptotic results.

Identification

Cardon-Hendel use the policy choice to figure out how much individuals know ex-ante about their ex-post risk. This takes all the variation in the data. How can we still identify unobserved risk aversion?

The key is the Poisson assumption: it is a one-parameter distribution, so knowing the risk type also implies knowing the variability of the outcome.

This allows us to use the remaining variation in policy choice to identify unobserved heterogeneity in risk aversion.

Intuition for how are the rest of the coefficients identified:

- Take identical (on observables) set if individuals.
- They vary with their number of claims. The data is the fraction of individuals in each group who choose a low deductible. Denote these fractions by p_g, where $g=0$ for the group of individuals with zero claims, $g=1$ for the group of individuals with one claim, and so forth.
- Given our structural model of deductible choice, we know which level of risk aversion is required to choose a low deductible, conditional on risk type.

- Given the posterior distribution of risk types in group 0, we can obtain the distribution of risk-aversion cutoffs that explain the deductible choice, so knowing p_0 is therefore sufficient to identify the mean level of risk aversion.
- We can repeat the same exercise for group 1. Thus, p_1 can again identify the mean level of risk aversion. In order to reconcile these two estimates, we can add unobserved variation in risk aversion, which will make the model rationalize these two observations of p_0 and p_1. Thus, the difference between p_0 and p_1 identifies the variation in unobserved risk aversion. If this variation is small, p_0 should be much smaller than p_1 due to adverse selection. As the variation increases, the effect of adverse selection on the difference between p_0 and p_1 decreases.
- Finally, if this unobserved variation in risk aversion is independent of the risk type, the model will predict some monotone relationship between p_0, p_1, and p_2. Deviations from this relationship will identify the correlation between unobserved risk and unobserved risk aversion. For example, taking the model assumptions as given, if we had p_0=0.5, p_1=0.51, and p_2=0.99 in the data, it must mean that risk aversion and risk are strongly positively correlated.

Table 4: The benchmark model

	Variable		Ln(λ) Equation	Ln(r) Equation
Demographics:	Constant		-1.5406 (0.0073)*	-11.8118 (0.1032)*
	Age		-0.0001 (0.0026)	-0.0623* (0.0213)
	Age^2		$6.24 \cdot 10^{-6}$ ($2.63 \cdot 10^{-5}$)	$6.44 \cdot 10^{-4}$ ($2.11 \cdot 10^{-4}$)*
	Female		0.0006 (0.0086)	0.2049 (0.0643)*
	Family	Single	omitted	omitted
		Married	-0.0198 (0.0115)	0.1927 (0.0974)*
		Divorced	0.0396 (0.0155)*	-0.1754 (0.1495)
		Widower	0.0135 (0.0281)	-0.1320 (0.2288)
		Other (NA)	-0.0557 (0.0968)	-0.4599 (0.7397)
	Education	Elementary	-0.0194 (0.0333)	0.1283 (0.2156)
		High School	omitted	omitted
		Technical	-0.0017 (0.0189)	0.2306 (0.1341)
		Academic	-0.0277 (0.0124)*	0.2177 (0.0840)*
		Other (NA)	-0.0029 (0.0107)	0.0128 (0.0819)
	Emigrant		0.0030 (0.0090)	0.0001 (0.0651)
Car Attributes:	Log(Value)		0.0794 (0.0177)*	0.7244 (0.1272)*
	Car Age		0.0053 (0.0023)*	-0.0411 (0.0176)*
	Commercial Car		-0.0719 (0.0187)*	-0.0313 (0.1239)
	Log(Engine Size)		0.1299 (0.0235)*	-0.3195 (0.1847)
Driving:	License Years		-0.0015 (0.0017)	0.0157 (0.0137)
	License $Years^2$		$-1.83 \cdot 10^{-5}$ ($3.51 \cdot 10^{-5}$)	$-1.48 \cdot 10^{-4}$ ($2.54 \cdot 10^{-4}$)
	Good Driver		-0.0635 (0.0112)*	-0.0317 (0.0822)
	Any Driver		0.0360 (0.0105)*	0.3000 (0.0722)*
	Secondary Car		-0.0415 (0.0141)*	0.1209 (0.0875)
	Business Use		0.0614 (0.0134)*	-0.3790 (0.1124)*
	History Length		0.0012 (0.0052)	0.3092 (0.0518)*
	Claims History		0.1295 (0.0154)*	0.0459 (0.1670)
Young Driver:	Young driver		0.0525 (0.0253)*	-0.2499 (0.2290)
	Gender	Male	omitted	-
		Female	0.0355 (0.0061)*	-
	Age	17-19	omitted	-
		19-21	-0.0387 (0.0121)*	-
		21-24	-0.0445 (0.0124)*	-
		>24	0.0114 (0.0119)	-
	Experience	<1	omitted	-
		1-3	-0.0059 (0.0104)	-
		>3	0.0762 (0.0121)*	-
Company Year:	First Year		omitted	omitted
	Second Year		-0.0771 (0.0122)*	-1.4334 (0.0853)*
	Third Year		-0.0857 (0.0137)*	-2.8459 (0.1191)*
	Fourth Year		-0.1515 (0.0160)*	-3.8089 (0.1343)*
	Fifth Year		-0.4062 (0.0249)*	-3.9525 (0.1368)*

Additional Quantities	
Var-Covar Matrix (Σ):	
σ_λ	0.1498 (0.0097)
σ_r	3.1515 (0.0773)
ρ	0.8391 (0.0265)
Unconditional Statistics:[a]	
Mean λ	0.2196 (0.0013)
Median λ	0.2174 (0.0017)
Std. Dev. λ	0.0483 (0.0019)
Mean r	0.0019 (0.0002)
Median r	$7.27 \cdot 10^{-6}$ ($7.56 \cdot 10^{-7}$)
Std. Dev. r	0.0197 (0.0015)
Corr(r,λ)	0.2067 (0.0085)
Obs.	105,800

Table 5: Risk aversion estimates

Specification[a]	Absolute Risk Aversion[b]	Interpretation[c]	Relative Risk Aversion[d]
Back-of-the-Envelope	$1.0 \cdot 10^{-3}$	90.70	14.84
Benchmark model:			
Mean Individual	$6.7 \cdot 10^{-3}$	56.05	97.22
25th Percentile	$2.3 \cdot 10^{-6}$	99.98	0.03
Median Individual	$2.6 \cdot 10^{-5}$	99.74	0.37
75th Percentile	$2.9 \cdot 10^{-4}$	97.14	4.27
90th Percentile	$2.7 \cdot 10^{-3}$	78.34	39.02
95th Percentile	$9.9 \cdot 10^{-3}$	49.37	143.27
CARA Utility:			
Mean Individual	$3.1 \cdot 10^{-3}$	76.51	44.36
Median Individual	$3.4 \cdot 10^{-5}$	99.66	0.50
Learning Model:			
Mean Individual	$4.2 \cdot 10^{-3}$	68.86	61.40
Median Individual	$5.6 \cdot 10^{-6}$	99.95	0.08
Comparable Estimates:			
Gertner (1993)	$3.1 \cdot 10^{-4}$	96.99	4.79
Metrick (1995)	$6.6 \cdot 10^{-5}$	99.34	1.02
Holt and Laury (2002)[e]	$3.2 \cdot 10^{-2}$	20.96	865.75
Sydnor (2006)	$2.0 \cdot 10^{-3}$	83.29	53.95

Figure 4: Counterfactuals – Profits

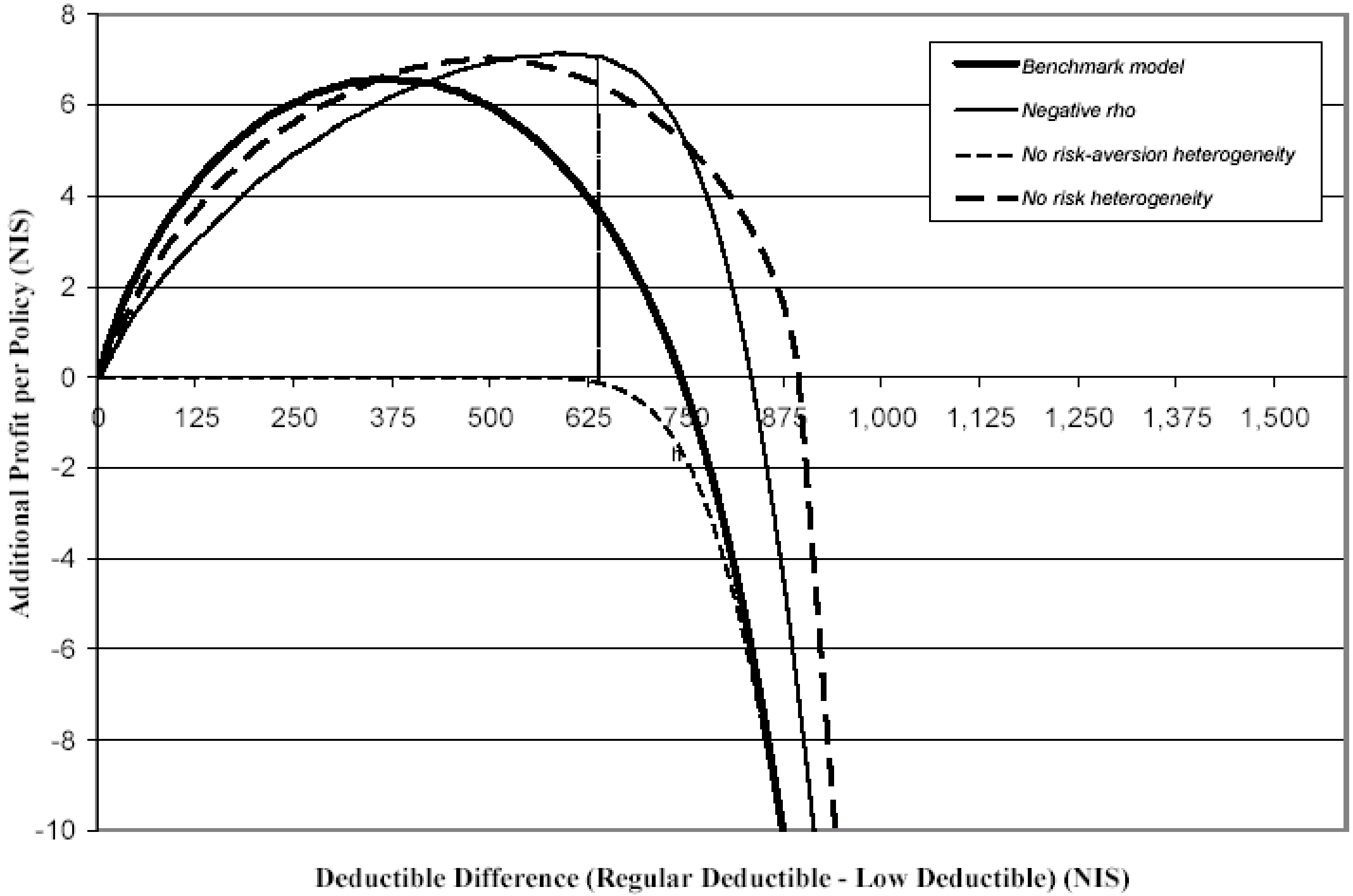
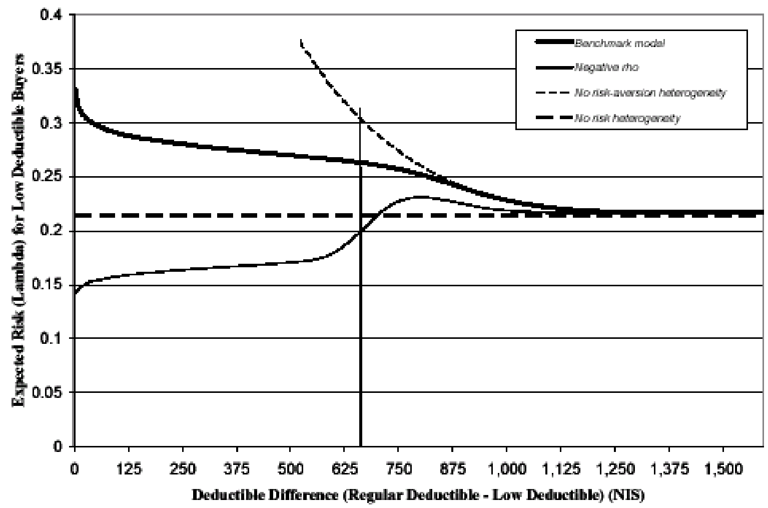
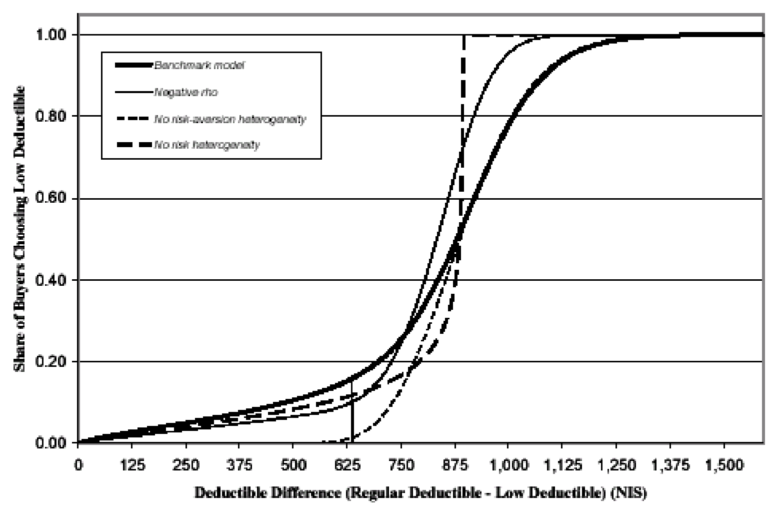


Figure 5: Counterfactuals – Selection





Welfare cost of adverse selection

- Key: overall agenda is to move beyond testing and start saying something about magnitudes.

- Still useful to test – easy to do and may provide guidance to subsequent steps (e.g., if we can't find simple evidence of adverse selection, does it make sense to quantify its costs?).

- Costs of adverse selection:

    - Mispricing (Akerlof, sort of)
    - Inefficient contracts (Rothschild and Stiglitz, sort of)
    - Disappearance of markets (empirically difficult!)

- Important to think/decide what is being held fixed. For example, what we assume about market structure and conduct may affect what we get for welfare.

"Optimal Mandates and The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," by Einav, Finkelstein, and Schrimpf (*Econometrica*, 2010)

- Similar methodology to that of Cohen and Einav, applied to guarantee choice in annuity markets:
  - Make distributional assumptions about heterogeneous mortality rates
  - Model guarantee choice using a standard "off the shelf" model

- "Big" question though: Estimating the cost of asymmetric information, and comparing them to the cost of mandates:
  - Mandates are typical solution, but far from ideal when individuals vary in prefs
  - Start by showing that structure is needed: reduced-form won't be sufficient

- Key contribution is in setting up the agenda of moving beyond "testing" and towards "quantifying."

Data

Similar to Finkelstein-Poterba, but focus on guarantee choice

	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

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

Model

Utility given by either consumption (if alive) or bequest (if dead):

$$v(w_t, c_t) = (1-\kappa_t)u(c_t) + \kappa_t b(w_t)$$

Without annuity, solve a standard consumption-saving problem:

$$V_t^{NA}(w_t) = \max_{c_t \geq 0}\left[(1-\kappa_t)(u(c_t) + \delta V_{t+1}^{NA}(w_{t+1})) + \kappa_t b(w_t)\right]$$

$$\text{s.t.} \quad w_{t+1} = (1+r)(w_t - c_t) \geq 0,$$

with terminal condition at T=100 of

$$V_{T+1}^{NA}(w_{T+1}) = b(w_{T+1}).$$

With annuity, solve:

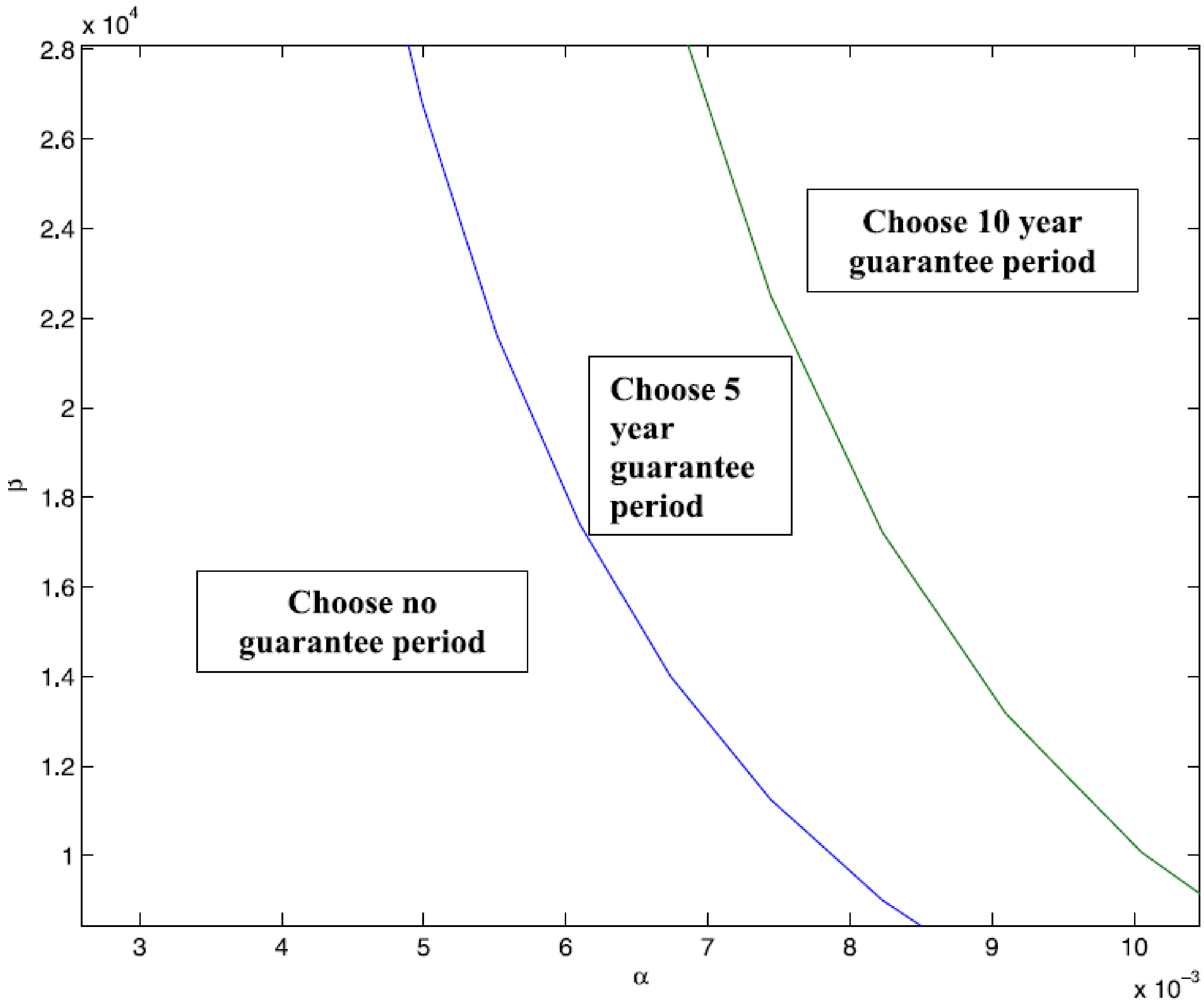
$$V_t^{A(g)}(w_t) = \max_{c_t \geq 0}\left[(1-\kappa_t)\left(u(c_t) + \delta V_{t+1}^{A(g)}(w_{t+1})\right) + \kappa_t b(w_t + Z_t(g))\right],$$

$$\text{s.t.} \quad w_{t+1} = (1+r)(w_t + z_t(g) - c_t) \geq 0,$$

where

$$Z_t(g) = \sum_{\tau=t}^{t_0+g}((1/(1+r))^{\tau-t} z_\tau(g))$$

And then choose the guarantee length *g* that maximizes $V^{A(g)}$ at time 0.


x 10^4
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
p
Choose 10 year guarantee period
Choose 5 year guarantee period
Choose no guarantee period
3
4
5
6
7
8
9
10
α
x 10^-3

Use this model to capture heterogeneity in two dimensions:

- In mortality rates (use a Gompertz assumption and alpha as a proportional shifter to it: high alpha → higher mortality hazard)
- In bequest preferences (weight on the bequest function *b*).

→ What is the analogy to Cohen and Einav?

Identification not as nice: model is more of a black box (so harder to "see" how it works), and no good price variation so stronger reliance on structure

- To deal with this, we run robustness checks and alternative specifications

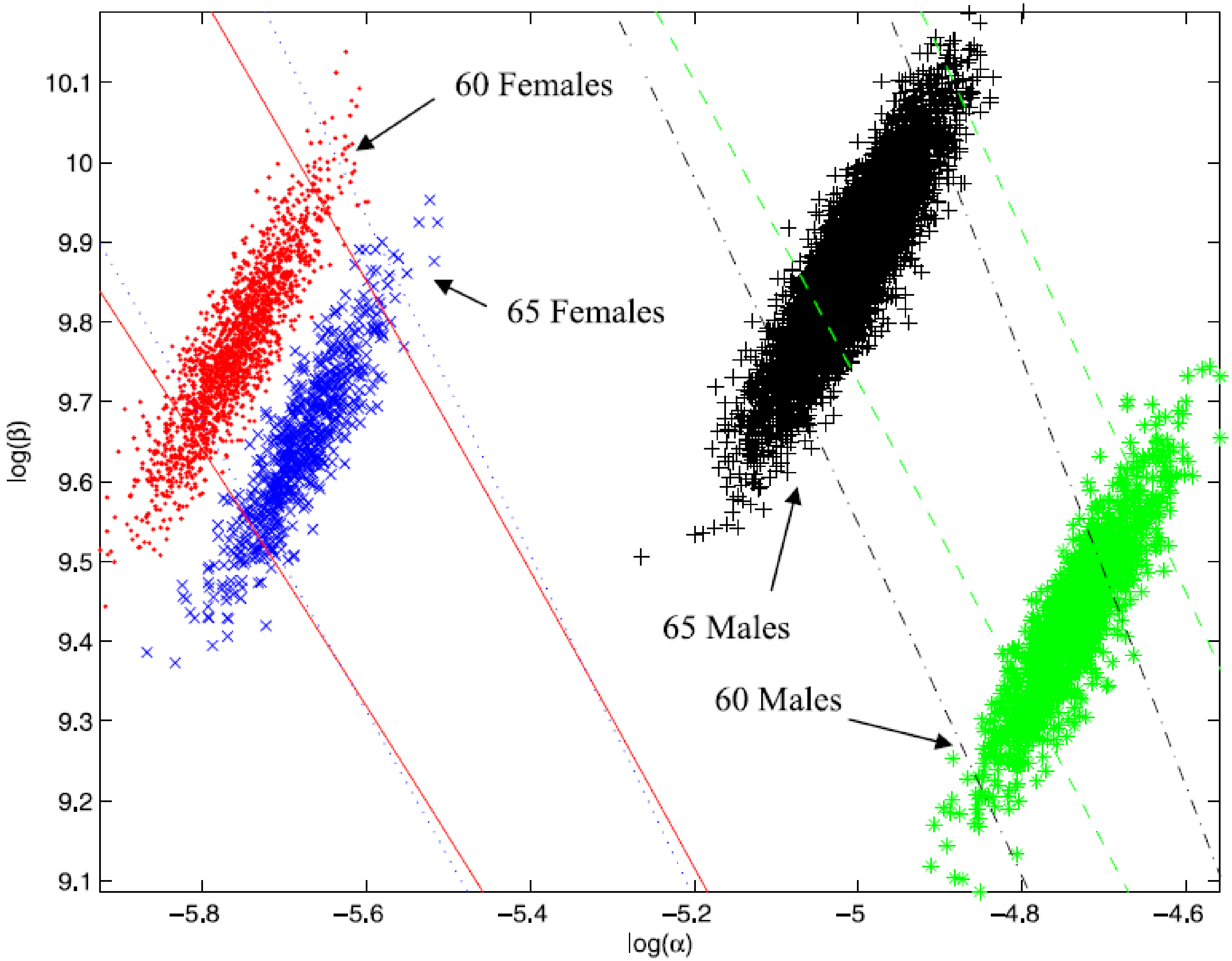
Assume joint lognormal, assume values for some other parameters (discount factor, risk aversion), and estimate using Maximum Likelihood. Then:

- Use the estimated distribution to ask what would be choices if prices were driven by individual-specific mortality rate, then convert to welfare numbers
- Repeat the same with mandates

Results

PARAMETER ESTIMATES[a]

		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	


60 Females
65 Females
65 Males
60 Males
10.1
10
9.9
9.8
9.7
9.6
9.5
9.4
9.3
9.2
9.1
log(β)
−5.8
−5.6
−5.4
−5.2
−5
−4.8
−4.6
log(α)

WELFARE ESTIMATES[a]

	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

- Large (not huge) loss from asymmetric information – around 150M pound/year
- Mandates could almost replicate first best, but other mandates could make things much worse