

1 A Canonical Insurance Model: Notation

- Agents are each characterised by a vector $\zeta = (\mathbf{x}, \nu)$, where ν is unobservable to firms;
- ϕ is the vector of insurance characteristics (e.g. what is covered and by how much);
- p is the price of the contract;
- S is the set of possible future scenarios for the agent, $s \in S$.

Agents are assumed to have **quasi-linear utility** in prices (no income effects)

$$u(s, \zeta, \phi) - p \quad (1.1)$$

they're also assumed to value contracts by their **expected utility** which, defining probabilities as $\pi(\cdot)$, is

$$v(\phi, \zeta) = \sum_{s \in S} \pi(s|\zeta) u(s, \zeta, \phi) \quad (1.2)$$

The expected outcome for a consumer characterised by vector ζ is

$$s(\zeta) = \sum_{s \in S} \pi(s|\zeta) s \quad (1.3)$$

To the insurer, the cost of agent's outcome s under contract ϕ is $\tau(s, \phi)$. This cost is observable; thus the expected cost of serving an agent is

$$c(\phi, \zeta) = \sum_{s \in S} \pi(s|\zeta) \tau(s, \phi) \quad (1.4)$$

and aggregate welfare is given by (since price is a zero-sum transfer)

$$W = \int v(\phi(\zeta), \zeta) - c(\phi(\zeta), \zeta) dG(\zeta) \quad (1.5)$$

Consumer $i \in I$ chooses contract $j \in J$ iff $v(\phi_j, \zeta_i) - p_j \geq v(\phi_k, \zeta_i) - p_k$, $\forall k \in J$; let $I(j) \subseteq I$ be the set of such agents.

Differences in (i) contract unit **costs** to the insurer and (ii) rates of **incidence** of outcomes s arise due to the irrepresentativeness of the population sample that selects into each contract j (*self-selection*). These are not necessarily the consequences of asymmetric information, as they could result from differences in \mathbf{x} ; however, if differences persist *after* conditioning on the set of characteristics observable to the insurer, selection might be the reason: that is,

$$\mathbb{E}_x[s|i \in I(j); x_i = x] \neq \mathbb{E}_x[s|i \in I(k); x_i = x] \quad (1.6)$$

is supportive of informational asymmetries between insurers and agents. Note however that asymmetric information generates two underlying effects that contribute to the difference above: (1.6) can be

These notes are based on Howard Smith's lectures from the 2018-2019 academic year. Of course, this is our interpretation of the material Howard presented. Good content is his; mistakes are ours.

rewritten as

$$\begin{aligned} \mathbb{E}_x[s|i \in I(j); \phi_j; x_i = x] - \mathbb{E}_x[s|i \in I(k); \phi_k; x_i = x] = \\ \textcolor{red}{\mathbb{E}_x[s|i \in I(j); \phi_j; x_i = x] - \mathbb{E}_x[s|i \in I(k); \phi_j; x_i = x]} + \\ \textcolor{blue}{\mathbb{E}_x[s|i \in I(k); \phi_j; x_i = x] - \mathbb{E}_x[s|i \in I(k); \phi_k; x_i = x]} \end{aligned} \quad (1.7)$$

where the second line expresses a **selection effect** resulting from the agents *being different* conditional on their contract choice, while the third line expresses the **incentive effect** from having different contracts. Simple tests of asymmetric information will not distinguish between the two.

2 Chiappori and Salanié (2000)

Analysis of the French car insurance market. Two most common insurance contracts: Low (L, minimum coverage) and High (H, covers more damages).

The authors propose two tests of asymmetric information: a parametric test relying on a pair of probits (or, more efficiently, a bivariate probit); and a non-parametric test relying on a χ^2 test of independence.

2.1 Pair of Probits / Bivariate Probit

For each individual $i = 1, \dots, N$, we have exogenous variables X_i and two endogenous dichotomous variables: **(i)** whether has purchased H insurance, and **(ii)** whether has caused any accident. The two endogenous variables have the following probit specification

$$H_i = \mathbb{1}\{X_i\beta + \epsilon > 0\} \quad \quad \quad acc_i = \mathbb{1}\{X_i\gamma + \eta > 0\} \quad (2.1)$$

where by assumption the unobserved components

$$\begin{bmatrix} \epsilon \\ \eta \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

For the $\rho = 0$ case, the equations can be separately estimated and generalised residuals obtained as

$$\hat{\epsilon}_i = \frac{\phi(X_i\hat{\beta})}{\Phi(X_i\hat{\beta})}y_i - (1 - y_i)\frac{\phi(X_i\hat{\beta})}{\Phi(-X_i\hat{\beta})} \quad \quad \quad \hat{\eta}_i = \frac{\phi(X_i\hat{\gamma})}{\Phi(X_i\hat{\gamma})}y_i - (1 - y_i)\frac{\phi(X_i\hat{\gamma})}{\Phi(-X_i\hat{\gamma})} \quad (2.2)$$

A test statistic can then be formed, under the null of conditional independence $cov(\epsilon_i, \eta_i) = 0$, as

$$\Omega = \frac{(\sum_i \hat{\epsilon}_i \hat{\eta}_i)^2}{\sum_i \hat{\epsilon}_i^2 \hat{\eta}_i^2} \sim \chi^2(1) \quad (2.3)$$

If $\rho = 0$ is not assumed, it is more efficient to estimate the bivariate probit system: a Wald test of the $\rho = 0$ null is then a valid test of independence given the assumed normality of error terms.

2.2 Non-Parametric Test

This test is based on the popular χ^2 test of independence. First, transform X_i in m dummy variables; this effectively creates 2^m cells, in each of which all individuals share all characteristics. For each cell, create a two-by-two table classifying agents based on their values of the two dichotomous endogenous

variables; then, defining $Z_{H,acc}$ as the number of agents within each cell with $H = \{0, 1\}$ and $acc = \{0, 1\}$, and Z as the total number of agents in the cell, define the test statistic

$$T = \sum_{H, acc = \{0, 1\}} \frac{[Z_{H,acc} - (Z_H Z_{acc} / Z)]^2}{Z_{H,acc}} \quad (2.4)$$

which, under within-cell (i.e. conditional) independence of H and acc , is χ^2 -distributed. There are 2^m such statistics: their sum is distributed, under conditional independence, as $\chi^2(2^m)$.¹

2.3 Empirical Results

None of the aforementioned tests can, somewhat surprisingly (?), reject conditional independence of the endogenous variables. At least for some parts of the driver population there seems to be no evidence of abused asymmetric information. It should be noted that all of these tests are tests of *presence* of asymmetric information, binary in their nature:² they say nothing about the welfare implications of the market failure. For that, we need to impose further structure on the problem.

3 Einav, Finkelstein and Cullen (2010)

3.1 Setting Sketch

Assume a competitive insurance market in which all firms offer two contracts $\phi = \{H, L\}$ at prices $p_L = 0$; $p_H = p$. Consumer utility from the L contract is normalised to zero; as a result, H is chosen whenever $v(H, \zeta_i) - p > 0$. Aggregate demand is

$$Q(p) = \int \mathbb{1}\{v(H, \zeta) - p > 0\} dG(\zeta) \quad (3.1)$$

and inverse demand is $P(Q) \equiv Q^{-1}(Q)$. Average cost then is

$$\begin{aligned} AC(p) &= \frac{1}{Q(p)} \int c(H, \zeta_i) \mathbb{1}\{v(H, \zeta) - p > 0\} dG(\zeta) = \\ &= \mathbb{E}_{\zeta}[c(H, \zeta) | v(H, \zeta) > p] \end{aligned} \quad (3.2)$$

and marginal cost is

$$MC(p) = \mathbb{E}_{\zeta}[c(H, \zeta) | v(H, \zeta) = p] \quad (3.3)$$

As for welfare, (1.5) can be split in consumer surplus

$$CS(p) = \int \mathbb{1}\{v(H, \zeta) - p > 0\} [v(H, \zeta) - p] dG(\zeta) \quad (3.4)$$

and producer surplus

$$PS(p) = \int \mathbb{1}\{v(H, \zeta) - p > 0\} [p - c(H, \zeta)] dG(\zeta) \quad (3.5)$$

implying that it is socially efficient to ensure agents for whom $v(H, \zeta_i) > c(H, \zeta_i)$.

¹ Alternatives include the Kolmogorov-Smirnov test for the empirical CDF of test statistics and a binomial distribution test that relies on a sharp arbitrary threshold.

² Also, they cannot distinguish between adverse selection and moral hazard.

Note that under the perfect competition assumption firms set equal prices for the same contracts so each company gets a random draw from the pool of insurance buyers. Moreover, the ‘insurance production function’ is the same ensuring average costs are the same. Importantly, the equilibrium will be one where $\mathbf{p} = \mathbf{AC}(\mathbf{p})$: any higher price would imply 0 demand, any lower price would induce a loss.³

Furthermore note that, in this competitive market characterised by adverse selection, **demand, MC and AC curves are all downward sloping** in the Quantity-Price space. This is because the first units sold go to those who have the largest valuation, which implies the largest cost to the insurer. As quantity increases, insurance is sold to less ‘expensive’ agents, and thus firms are going to **lower** prices (which are kept at no-profit levels by the perfect competition assumption): hence the downward curves. Clearly, at 0 quantity $AC=MC$; as quantity increases, **MC slopes downwards faster than AC** because the latter, being an algebraic mean, is weighed by the larger cost of the early units, while the former is not – in econspeak, AC includes the burden of cross-subsidising agents due to adverse selection while MC does not, representing the perfect information case. Asymmetric information implies a deadweight loss which can be depicted as a cute triangle (see below).

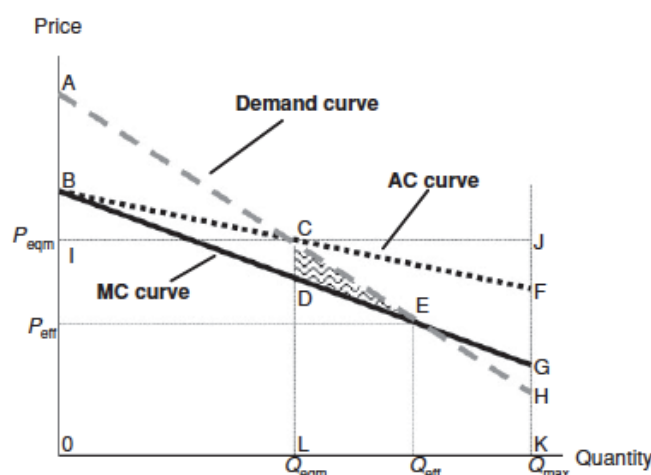


FIGURE I
Efficiency Cost of Adverse Selection

This figure represents the theoretical efficiency cost of adverse selection. It depicts a situation of adverse selection because the marginal cost curve is downward-sloping (i.e., increasing in price, decreasing in quantity), indicating that the people who have the highest willingness to pay also have the highest expected cost to the insurer. Competitive equilibrium is given by point C (where the demand curves intersect the average cost curve), whereas the efficient allocation is given by point E (where the demand curve intersects the marginal cost curve). The (shaded) triangle CDE represents the welfare cost from underinsurance due to adverse selection.

However, note that **advantageous selection** is just as sensible: if risk averse agents see obtaining insurance and taking precautionary actions as complements, those who demand more insurance might actually be the less expensive. In that case, MC and AC curves would be **upward** sloping; the market inefficiency would now result from agents buying **too much** insurance – as low-risk high-valuation agents will subsidise provision of underpriced contracts to high-risk low-valuation types. There’s a cute picture for this too, in the next page.

The idea of Einav, Finkelstein and Cullen is that these **demand and cost curves are sufficient statistics for welfare analysis** in the pure *Marschak maxim* sense; that is, the curves mediate the

³Intuitively, there will be no marginal cost pricing as a result of different agent types: the seller targets an ‘average’ agent to cross-subsidise losses from adversely selected agents with gains from less risky agents.

information content of the primitives, thus making knowledge of the primitives superfluous once the curves have been estimated. The authors then “estimate the demand and cost curves but remain agnostic about the underlying primitives that give rise to them” (p. 890), as “the precise *source* of selection is not germane for analysing the efficiency consequences of the resultant selection, or the welfare consequences of public policies that change the equilibrium price” (*ibidem*). Of course this requires **insurance contracts to be predetermined** – only their prices are allowed to change.

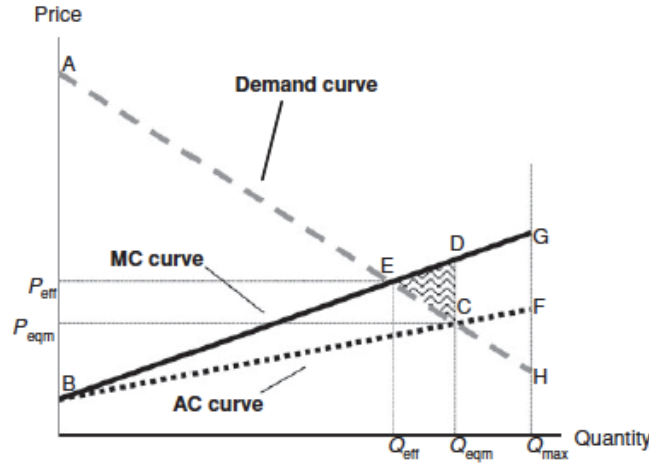


FIGURE II
Efficiency Cost of Advantageous Selection

This figure represents the theoretical efficiency cost of advantageous selection. It depicts a situation of advantageous selection because the marginal cost curve is upward-sloping, indicating that the people who have the highest willingness to pay have the lowest expected cost to the insurer. Competitive equilibrium is given by point C (where the demand curve intersects the average cost curve), whereas the efficient allocation is given by point E (where the demand curve intersects the marginal cost curve). The (shaded) triangle CDE represents the welfare cost from overinsurance due to advantageous selection.

3.2 Empirical Analysis

Data on health insurance purchases by 3779 Alcoa workers. Two levels of coverage, H and L, detailed demographics (including health conditions, which allows to calculate costs to the insurer). The setting is particular because, within the firm, employees doing comparable jobs in different organisational units face different insurance costs that are quasi-random.

Given price is assumed to be exogenous, the world smiles at the authors: they estimate a linear probability model for Q_i , whether i demands high insurance or not, and a linear regression for costs to the insurer

$$Q_i = \alpha + \beta p_i + \epsilon_i \quad AC_i = \gamma + \delta p_i + u_i \quad (3.6)$$

note the information from these two curves is sufficient to back out the marginal cost curve: in fact, given aggregate average costs and quantities for each contract,

$$TC(p) = AC(p)Q(p) \quad MC(p) = \frac{\partial TC(p)}{\partial Q(p)} = \frac{\partial AC(p)Q(p)/\partial p}{\partial Q(p)/\partial p} = \frac{AC(p)\beta + Q(p)\delta}{\beta} \quad (3.7)$$

Estimation by OLS suggests that demand is **decreasing** in price of H, and that the average cost of agents who select into more costly contracts is higher – indicating adverse selection. The welfare loss is about 3% of total estimated surplus; however, this welfare loss is lower than the loss that would be incurred under a mandated high contract. Estimation results are below.

TABLE III
ESTIMATION RESULTS

| Dependent variable (sample) | 1 if chose High (both High and Low) (1) | Incremental cost (only High) (2) |
|--|---|--|
| Panel A: Estimation results | | |
| Relative price of High (US\$) | −0.00070 (0.00032) [.034] | 0.15524 (0.06388) [.021] |
| Constant | 0.940 (0.123) [.000] | 391.690 (26.789) [.000] |
| Mean dependent variable | 0.652 | 455.341 |
| Number of observations | 3,779 | 2,465 |
| R^2 | .008 | .005 |
| Panel B: Implied quantities of interest | | |
| Competitive outcome (point C in Figure I) | | $Q = 0.617, P = 463.5$ |
| Efficient outcome (point E in Figure I) | | $Q = 0.756, P = 263.9$ |
| Efficiency cost from selection (triangle CDE) | | 9.55 |
| Total surplus from efficient allocation (triangle ABE) | | 283.39 |
| Efficiency cost from mandating contract H (triangle EGH) | | 29.46 |

4 Einav, Finkelstein and Schripf (2010)

UK market for annuities: at retirement, people are forced to annuitise accumulated funds in retirement savings accounts. Individuals are allowed a choice of *guarantee period* between 0, 5, and 10 years, which ensures an annuitised stream **irrespective of death** for the chosen period. The longer the guarantee period, however, the smaller the stream returned while alive: no reason to select guarantee shorter than expected life span.

Under perfect competition and complete information, each plan would have a net expected return of 0 for retirees conditional on their type $\zeta_i = (\mathbf{x}_i, \nu_i)$. Under imperfect information, there is scope for **self-selection on likelihood of premature death** as riskier agents select longer guarantees to leave to their dependents. The authors set to estimate the welfare losses from this selection.

Data on age, guarantee selected, year of death and contract information for 9000+ retirees.

4.1 Model

For each consumer, $\mathbf{x}_i = \{sex_i, age_i\}$, $\nu_i = \{\alpha_i, \beta_i\}$ – where α_i is heterogeneity in the propensity to die and β_i is heterogeneity in taste for bequest. Agents with large draws of α and β will be more likely to select into longer guarantees.

The probability of dying is given by hazard rate $\kappa_t(\alpha_i, \lambda)$, with $\kappa'_t(\alpha_i) > 0$ and λ being a shape parameter. Utility for period t is defined as

$$v(w_t, c_t) = [1 - \kappa_t(\alpha_i, \lambda)]u(c_t) + \kappa_t(\alpha_i, \lambda)\beta_i u(w_t + Z_t^\phi) \quad (4.1)$$

where c_t is consumption, w_t is the value of assets, and Z_t^ϕ is the present value of post-death guarantee payments ($\phi \in \{0, 5, 10\}$).

The optimal consumption plan then obeys

$$\begin{aligned} V^\phi(w_t, \zeta_i) &= \max_{c_t \geq 0} [1 - \kappa_t(\alpha_i, \lambda)] \{u(c_t) + \delta V^\phi(w_{t+1}, \zeta_i)\} + \kappa_t(\alpha_i, \lambda) \beta_i u(w_t + Z_t^\phi) \\ \text{s.t. } w_{t+1} &= (1 + r)(w_t + Z_t^\phi - c_t) \geq 0 \end{aligned} \quad (4.2)$$

and the choice of guarantee ϕ obeys $\max_\phi V^\phi(w_t, \zeta_i)$.

The model is estimated with some useful parametric assumptions: first, on the form of the mortality hazard rate; second, imposing CRRA utility; third, assuming that $\{\alpha, \beta\}$ is bivariate lognormal (with unspecified mean, variance and covariance to be estimated). Furthermore, the discount factor δ , the CRRA parameter and the interest rate are calibrated.

The log-likelihood for the sample is then relatively easy to derive: the probability of i dying at age s after having chosen guarantee ϕ is

$$Pr(s, \phi | \alpha, \lambda) = Pr(s | \phi, \alpha, \lambda) \times Pr(\phi | \alpha, \lambda) \quad (4.3)$$

denote by μ and Σ the mean vector and variance-covariance matrix for the joint distribution of α and β ; the likelihood then is

$$\mathcal{L}_i(s, \phi | \mu, \Sigma, \lambda) = \int_{\alpha} Pr(s | \phi, \alpha, \lambda) \times Pr(\phi | \alpha, \lambda) dF(\alpha; \mu, \Sigma) \quad (4.4)$$

and the maximum likelihood estimator is defined as, for $\theta = \{\mu, \Sigma, \lambda\}$,

$$\hat{\theta} \in \arg \max_{\theta} \sum_i \ln \mathcal{L}_i(s, \phi | \mu, \Sigma, \lambda) \quad (4.5)$$

Incidentally, the estimation procedure is actually *significantly* more complicated: in the paper, λ is first estimated by maximum likelihood; conditional on this estimate, a second maximum likelihood estimator yields distributional parameters of the heterogeneity. However, the second maximum likelihood estimator requires in principle re-evaluating the decision rule (4.2) for the continuum of α values. The computation is simplified by the monotonicity of optimal guarantee choice in the heterogeneity, which allows to (i) define cutoff rules for guarantee choice; (ii) calculate cutoffs *once* for a grid of α ; (iii) interpolate likelihood. The two-step MLE is necessary for computational feasibility as the second MLE would otherwise require iterating over the λ space as well.

4.2 Results

The authors find significant heterogeneity in both propensity to die α and taste for bequest β ; furthermore, they find a **large positive correlation** between the two components.

TABLE III
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 | |

^aThese estimates are for the baseline specification described in the text. Standard errors are in parentheses. Since the value of λ is estimated separately, in a first stage, we bootstrap the data to compute standard errors using 100 bootstrap samples.

4.3 Welfare

To analyse the implications of the model for welfare, the authors use its estimates to compute the ‘wealth equivalent’ for each individual, i.e. the amount of wealth that would make her indifferent with choosing her actual wealth and annuity. Given an actuarially fair annuity rate and a risk averse individual, the wealth equivalent will be **higher** than actual wealth plus guarantee, as the guarantee removes the risk from dying.⁴ However, for an annuity rate sufficiently below the perfect information equilibrium, the wealth equivalent can be **lower**. In general, the difference gives an estimate of the consumer gain from the availability of the guarantee: this allows authors to estimate the value of available contracts, as well as counterfactuals – e.g. with a compulsory guarantee of 10 years, without guarantee, etc.

The first finding is that, as expected, the **main beneficiaries from the pooling equilibrium are low- α individuals**: they enjoy higher annuity rates than required as a result of the effort by companies to deter high- α from selecting into longer guarantees.⁵ Fixing asymmetric information allows firms to price optimally and transfers the surplus away from these agents, benefitting discriminated-against higher- α agents.

A second finding is that the effect of a **mandated guarantee** largely depends on the length of the mandated contract, and that the optimal choice critically depends on the joint distribution of risk and preferences. The mandated guarantee length that improves over the status quo the most is a **mandated 10-year guarantee**; however, by the compulsory nature of the UK system, it is not possible to account for the welfare response of those who **do not** want to annualise their retirement savings.

⁴If one selects the wealth equivalent, one can only consume it, while the guarantee can be passed to dependents and generate further utility upon death; hence, the wealth equivalent must be larger to induce consumers to accept it.

⁵Note also that **low- β individuals** benefit from the annuity market as they care less about not having wealth at death (without the annuity market, they would have to keep their entire wealth non-annuitised at a risk).