

# The Impact of Price Discrimination in Markets with Adverse Selection

André Veiga\*

University of Oxford

This version: November 25, 2016

[\[Please click here to download the latest version\]](#)

## Abstract

Would consumer surplus increase if annuity rates were age-neutral? This paper characterizes the socially optimal contractibility of a given signal in markets with adverse selection. A signal (e.g., age) partitions consumers into subsets (e.g., young and old). A regulator restricts price-discrimination on the basis of the signal if the consumer subsets where the level of cost is higher are also the subsets where there is greater deadweight loss due to adverse selection. Such signals are empirically common. To illustrate the welfare benefit of price discrimination policy, I use a structural model to estimate its impact on the UK annuities market. The model is estimated using proprietary data that include the annuity seller's estimate of each individual's longevity. I find that restricting price discrimination can increase consumer surplus by the equivalent of about £6.5 million per year.

**Keywords:** Adverse Selection, Price Discrimination, Structural Estimation

**JEL Classification Codes:** D82, L52, D41

---

\*I thank Simon Cowan, Ian Crawford, Andrew Dilnot, James Duffy, Liran Einav, Basile Grassi, Sukriti Issar, Ian Jewitt, Michael Keane, Paul Klemperer, Jonathan Levin, Margaret Meyer, Simon Quinn, Howard Smith, Kirsten Smith, John Vickers, and E. Glen Weyl. Segye Shin provided outstanding research assistance. Department of Economics and Somerville College, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK; [andre.veiga@economics.ox.ac.uk](mailto:andre.veiga@economics.ox.ac.uk) or [andreveiga@gmail.com](mailto:andreveiga@gmail.com).

# 1 Introduction

Would consumer surplus increase if annuity rates were not allowed to vary with age? Should these rates vary with gender? There is little guidance as to which individual characteristics (e.g., gender, age) should be contractible in a given setting. The issue is especially relevant in markets like annuities and insurance, since characteristics like age and gender are often correlated with the private information responsible for adverse selection.

The disparate set of contractibility regulations across countries and markets suggests that their effect on welfare is not well understood. For instance, a controversial 2012 EU ruling mandated that all insurance prices be gender-neutral but age, place of residence and other characteristics remain fully contractible. Similarly, the United States Affordable Care Act mandates that pre-existing health conditions are not contractible by insurers, but prices can vary with the individual's zip code and smoking status. Understanding the welfare effect of such policies is particularly important because these policies are politically expedient to change and have a low implementation cost. Conversely, policies like subsidies and mandates require substantial public funds or they can be perceived as restricting consumer choice.

This paper characterizes the socially optimal contractibility of a given signal. A signal (e.g., age) partitions consumers into subsets (e.g., young and old). Restricting the contractibility of age shrinks the difference in the prices charged to young and old individuals. I show how the optimal contractibility of a signal is determined by the characteristics of the consumer partition induced by the signal. I then illustrate empirically the potential welfare gain of such a policy in the context of UK annuities. The results suggest that optimal price discrimination policy can achieve welfare gains equivalent to 0.09% of total annuitized wealth, or £6.5 million per year.

I consider an industry affected by selection, where firms compete in prices (as in [Akerlof \(1970\)](#), but unlike [Rothschild and Stiglitz \(1976\)](#)). A product is adversely selected if its infra-marginal buyers (those with the highest relative valuations for that product) are costlier than its marginal buyers. If all individuals must be charged the same price (for information or regulatory reasons), firms will cover the high cost incurred on infra-marginal buyers by charging a high price to all individuals. Then, low-cost marginal individuals will face a price higher than they would face if information were symmetric. Therefore, an adversely selected product exhibits a price that is inefficiently high.

Often, firms have access to a signal (e.g., age) that is imperfectly correlated with individual costs and valuations. Realizations of the signal partition consumers into multiple subsets (e.g., young and old). A regulator can, for instance, mandate that the price difference between the young and the old cannot exceed a given amount. A 'full community rating' or 'full CR' policy would mandate the price difference be zero. A 'full price discrimination' or 'full PD' policy would impose no constraint.

The key tension is that CR raises prices for some individuals and lowers it for others. If the industry is competitive, firms break even in any scenario. Relative to PD, CR never results in a Pareto improvement, but it can increase overall consumer surplus. This paper characterizes the policy that maximizes total consumer surplus.

If each consumer subset is homogeneous in cost, full PD is the welfare maximizing CR policy. In this case, there is no private information within each consumer subset. Under perfect competition, full PD eliminates all deadweight loss, as in [Pauly \(1970\)](#). CR introduces adverse selection where it previously did not exist.

However, there often remains some private information conditional on the signal realization. Suppose young consumers possess significant amounts of private information. For instance, a young annuity buyer might be purchasing an annuity at an early age because a health condition has forced her to retire, or because she is wealthy enough to retire early. This private information can create adverse selection within the subset of young buyers. Suppose there is a single annuity contract for sale. Under full PD, adverse selection causes the annuity price among the young to be excessively high, resulting in a deadweight loss. If old annuity buyers are homogeneous, the price charged to the old is efficient under full PD. Now suppose a regulation marginally shrinks the price difference across these two consumer subsets. Prices charged to the old would rise, but this would have no first-order effect on welfare, because the old were charged the efficient price. The price charged to the young would fall, increasing welfare within that consumer subset. Thus, overall welfare would increase.

In this (extreme but illustrative) example, CR is beneficial because the high-cost consumer subset (the young) is also the subset experiencing the greater adverse selection distortion. This pattern is empirically common in a wide range of markets, because individuals tend to have private information about costly outcomes ([Hendren \(2013\)](#), [Brown et al. \(2014\)](#)). Therefore, some amount of CR is often desirable. Moreover, the intuition applies to a wide range of markets from annuities to health insurance, life insurance and credit markets.

The intuition described above generalizes in several ways. The model can accommodate signals that partition consumers into more than two subsets. I also extend the model to a setting where firms offer multiple products, but consumers must purchase one of them, as is the case in my empirical application to UK annuities, described below. In both extensions, the qualitative results and their intuition are similar to those described above.

I proceed by illustrating empirically the potential benefit of CR policy in the context of the £12bn/year UK annuities market. Annuities provide a stream of payments conditional on the annuitant being alive. In this setting, annuity purchase is mandatory but individuals can choose their guarantee period. For instance, a five-year guarantee implies the buyer's estate receives the annuity payments for five years even if the individual dies within that period. Contracts with shorter guarantees have larger monthly payments. All else being equal, short guarantees are appealing to individuals with high longevity since they are unlikely to be affected by guarantees. Because long-lived individuals are costlier buyers of annuities, there can be adverse selection into contracts with short guarantees. This adverse selection raises the cost of short-guarantee contracts, lowering the monthly payments in such contracts. In equilibrium, adverse selection results in too few individuals choosing contracts with short guarantees (but large monthly payments).

I begin by estimating the joint distribution of longevity and bequest motives among UK mandatory annuitants. I use a structural model of annuity contract choice that makes the following assumptions. First, individual mortality follows a proportional hazard Gompertz process. Second, period utilities from consumption and bequests both exhibit constant relative risk aversion with the same curvature parameter. Third, retirement timing is exogenous to the choice in contract.

The model is estimated using a proprietary dataset of individual level annuity purchases from a large UK insurer over a span of two years. The data include all information used in pricing and several other individual-level covariates. The data also include the firm's estimate of each consumer's life expectancy, using the firm's proprietary algorithm. This is, to my knowledge, the only dataset to include the firm's perceived cost in an annuities setting. This information allows me to take individual longevity as observed (to the econometrician), although it is not contractible by the firm.

The model is identified by variation in individual choices and variation in the rates offered to different individuals. Rates vary with the size of pension funds be-

cause firms are allowed to price discriminate on this characteristic. However, fund size does not directly affect guarantee choices under the CRRA assumption, so this variation is exogenous given the model's assumptions. To make the estimation maximally flexible, I estimate the model independently in four consumer subsets: men and women, purchasing at ages 60 and 65.

I use the estimation results to compute the competitive equilibrium at a continuum of policies between full PD and full CR. At each policy, I measure consumer welfare as the increase in non-annuitized wealth that would deliver the same utility as the individual's preferred annuity contract. For concreteness, I express this amount as a share of total annuitized wealth (£12bn in 2013). I find that optimally restricting the contractibility of gender increases consumer surplus among 65-year-olds (men and women) by the equivalent of 0.01% of annuitized wealth, and by 0.09% among 60-year-olds. Optimally restricting the contractibility of age increases welfare by 0.11% among women (60- and 65-year-olds). However, full PD on the basis of age is the optimal policy for men.

A key theoretical innovation of this article is to characterize, in a tractable way, the optimal contractibility of an arbitrary signal. [Levin \(2001\)](#) finds that, in markets with adverse selection, revealing private information always increases the probability of trade. That article does not consider welfare directly. Moreover, it focuses on binary signals that partition individuals into one subset where everyone has a higher valuation than anyone in the other subset. That result does not hold when more general signals are considered.

I focus on a novel benefit of CR: reducing *static* deadweight loss from adverse selection. [Arrow \(1963\)](#); [Handel, Hendel and Whinston \(2015\)](#); [Koch \(2014\)](#) emphasize that age-based CR can increase welfare because it allows individuals to obtain insurance against changes in health status over one's lifetime (i.e., reclassification risk). My results show that CR can be beneficial even in settings where reclassification risk is absent, as is the case in annuity markets.

The innovations of my theoretical model include allowing for residual asymmetric information conditional on signal realizations and considering the continuum of policies between full PD and full CR. By contrast, this kind of heterogeneity has not been considered by the existing literature on third-degree price discrimination (e.g., [Aguirre, Cowan and Vickers \(2010\)](#), [Schmalensee \(1981\)](#)). [Chen and Schwartz \(2015\)](#) consider a monopoly selling to two subsets of individuals, each with a different level of cost, but each consumer subset is homogenous. Moreover, this literature tends

to focus on comparing the two extreme policies of full PD and full CR.

My focus is on CR policies which are implementable using only knowledge of aggregate quantities such as the joint distribution of cost and willingness to pay. By contrast, [Bergemann, Brooks and Morris \(2015\)](#) shows that any feasible split of consumer and producer surplus is achievable under an appropriate information structure. However, implementing these allocations would require a regulator to know each individual's valuation.

[Crocker and Snow \(1986\)](#), [Hoy \(1982\)](#) and others find that costless consumer categorization (i.e., PD) can expand the utilities possibility frontier and even lead to a Pareto improvement. In this context, [Finkelstein, Poterba and Rothschild \(2009\)](#) consider specifically the annuities market. Those articles consider settings with endogenous quality (like [Rothschild and Stiglitz \(1976\)](#)) whereas I consider firms that compete only in prices. Competition in prices with otherwise fixed contract characteristics is common in many significantly regulated markets, like annuities and health insurance.

My empirical framework is similar to that of other studies of annuity choice such as [Kotlikoff and Spivak \(1981\)](#), [Mitchell et al. \(1999\)](#), [Davidoff, Brown and Diamond \(2005\)](#). The most closely related article is [Einav, Finkelstein and Schrimpf \(2010\)](#), henceforth EFS), who also study UK annuities. Their dataset does not include individual expected longevity or a number of other covariates I observe. Also, because it is less recent, their data does not include variation in rates. Moreover, their focus is on optimal mandates, whereas I focus on CR policy.

To my knowledge, this article is the first study of CR in the context of annuities markets. [Blumberg and Buettgens \(2013\)](#), [Orsini and Tebaldi \(2015\)](#), and [Ericson and Starc \(2015\)](#) empirically study the effect of age-based CR in the context of US health insurance. [Orsini and Tebaldi \(2015\)](#) find CR has little effect on enrollment but a significant effect on government expenditures. [Ericson and Starc \(2015\)](#) find that CR reduces profit and increases overall consumer surplus.

The paper is organized as follows. Section 2 contains the theoretical results. Section 3 describes the data and institutional context. Section 4 describes the structural model of annuity choice. Section 5 describes the estimation and results. Section 6 contains the counterfactuals. Section 7 concludes. The appendices following the main text contain all proofs and additional details regarding the data, estimation and additional calibrations.

## 2 Theory

### 21 Baseline Model

I begin by describing a baseline model where there is a single set of consumers and a single product being sold, as in EFC.<sup>1</sup> I consider a continuum of consumers with unit mass. A single contract is offered (e.g., an annuity or a health insurance plan). Consumer willingness to pay is  $u \in [0, \bar{u}]$ , with smooth PDF  $f(u)$ .<sup>2</sup>

All firms are identical, and I focus on allocations where all firms charge the same price  $p$ . Price competition is common in regulated markets such as annuities (Einav, Finkelstein and Cullen (2010)). A consumer buys the contract if  $u \geq p$ . Industry demand is

$$Q = Q(p) = \int_p^{\bar{u}} f(u) du$$

with slope  $Q' = Q'(p) < 0$ . The semi-elasticity of demand is

$$\sigma = \sigma(p) = -\frac{Q'}{Q} > 0.$$

Individuals with willingness to pay  $u$  have a cost  $c = c(u) \geq 0$ .<sup>3</sup> I will refer to  $c(p)$  as the “industry marginal cost” since it is the derivative of industry total cost with respect to quantity, at price  $p$ .<sup>4</sup> The industry average cost is

$$AC = AC(p) = \frac{1}{Q(p)} \int_p^{\bar{u}} c(u) f(u) du,$$

with slope  $AC' = \sigma(AC - c)$ . Industry profit is

$$\pi(p) = Q(p)(p - AC(p)).$$

with slope  $\pi' = \pi'(p) = Q'(p - c) + Q$ .

<sup>1</sup>Relative to EFC, I consider demand, cost, and so on as functions of prices, not quantities. This setup is more natural in the context of price discrimination and will result in simpler expressions for the results below. It is straightforward to express all functions in terms of quantities, as in Einav, Finkelstein and Cullen (2010); Mahoney and Weyl (2014).

<sup>2</sup>Appendix I explores examples of possible micro-foundations of  $u$ .

<sup>3</sup>That is,  $c(u)$  is the expected cost among all individuals with valuation  $u$ .

<sup>4</sup>Letting total cost be  $C = \int_p^{\bar{u}} c(u) f(u) du$ , then  $\frac{dC}{dQ} = \frac{dC}{dp} \frac{1}{dQ/dp} = \frac{cQ'}{Q} = c$ .

I use the following definition of consumer surplus:

$$U = U(p) = \int_p^{\bar{v}} (u - p) f(u) du,$$

with slope  $U' = U'(p) = -Q(p)$ . Therefore, welfare is  $W(p) = U(p) + \pi(p)$ , with slope  $W' = W'(p) = Q'(p - c)$ . Notice that, when a product is priced efficiently (i.e., at marginal cost, so  $p - c = 0$ ), then a marginal increase in the price has no first-order effect on welfare.

This definition of consumer surplus implies income effects are negligible, so redistribution has no intrinsic value. The definition also implies the absence of the [Hirshleifer \(1971\)](#) effect, in which ex-post redistribution provided a rationale for restricting the information available to insurers. Therefore, any benefit of CR in this setting must arise due to a decrease in static deadweight loss.<sup>5</sup>

I assume free entry of firms into this industry. Under the regularity conditions presented below, a unique pure-strategies Nash equilibrium price  $p^*$  exists that satisfies

$$\pi(p^*) = 0 \Rightarrow p^* = AC(p^*).$$

At the equilibrium price  $p^*$ , each firm breaks even and so does the industry as a whole.

However, any welfare-maximizing (interior) price  $p^{**}$  satisfies

$$W'(p^{**}) = 0 \Rightarrow p^{**} = c(p^{**}).$$

That is, the socially optimal price  $p^{**}$  is determined by the fixed point of  $c(p)$ , whereas the equilibrium price  $p^*$  is determined by the fixed point of  $AC(p)$ . Therefore, whenever  $c(p) \neq AC(p)$ , the competitive price will not be the welfare-maximizing price ( $p^* \neq p^{**}$ ).

As emphasized by EFC,  $c(p)$  might differ from  $AC(p)$  due to adverse or advantageous selection. The product “buy” is adversely selected if  $c' = c'(u) > 0$ , so those with higher willingness to pay also have higher cost. Adverse selection implies that, at a given  $p$ , infra-marginal buyers have a higher cost than marginal buyers. Therefore,

$$\mathbb{E}[c \mid u > p] > c \Rightarrow AC > c.$$

---

<sup>5</sup>This definition of consumer surplus and welfare is ubiquitous in the PD literature and common in the adverse selection literature.



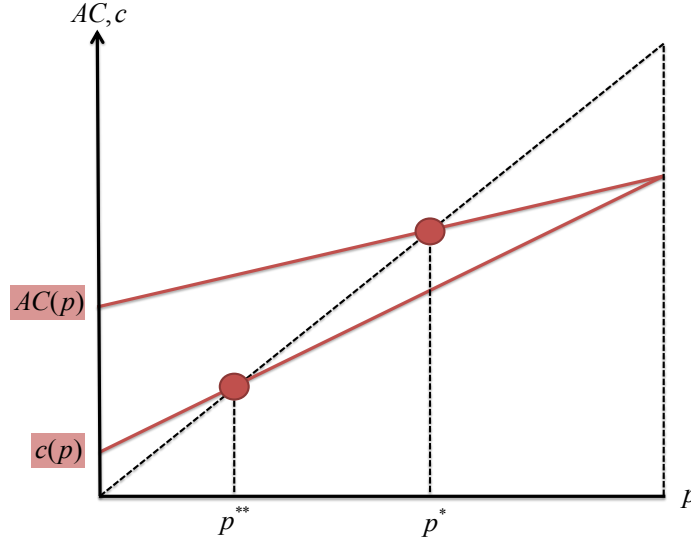


Figure 1: An example of equilibrium and socially optimal price in a market with adverse selection. The diagonal dashed line is the 45° line.

This wedge between  $c(p)$  and  $AC(p)$  generates the selection distortion emphasized by EFC. If the product is adversely selected, then  $p^* > p^{**}$ . That is, the product's equilibrium price will be higher than is socially optimal, which distorts individual choices. Relative to a setting with symmetric information, fewer individuals buy the product. Figure 1 illustrates the equilibrium and socially optimal price in a market where the product is adversely selected.

It is also possible that the product is advantageously selected, if  $c'(u) < 0$ , implying  $AC < c$  and  $AC' < 0$ . In this case, infra-marginal buyers are less costly than marginal buyers, which drives down the price of the product resulting in  $p^* < p^{**}$ . In this case, relative to a setting with symmetric information, more individuals choose to buy the product. I will assume  $c(u)$  is monotonic, so selection is globally signed.

Because  $AC' = \sigma(AC - c)$ , adverse selection also implies  $AC' > 0$ . The slope of  $AC$  is proportional to the wedge between  $AC$  and  $c$ , and therefore  $AC'$  captures the importance of adverse selection. For instance, without selection, the  $AC$  curve is flat ( $c' = 0 \Rightarrow AC' = 0$ ).<sup>6</sup> If  $AC'$  is large and negative, then advantageous selection is significant.

<sup>6</sup>Notice also that as  $p \rightarrow \bar{u}$ ,  $AC - c \rightarrow 0$ , and therefore  $AC' \rightarrow 0$ . Moreover,  $AC(\bar{u})$  is the average cost in the entire population.

I make four regularity assumptions. First, I assume  $Q$  is log-concave so  $\sigma'(p) > 0$ . This assumption implies welfare is responsive to prices when markets experience a large price distortion, but small welfare gains result from correcting prices in markets that are already close to their efficient price. Second, I assume  $c' < 1$ , which implies adverse selection cannot be too extreme. This assumption requires that, as valuations  $u$  increase, the surplus consumers obtain beyond simply imposing a cost on the insurer ( $u - c(u)$ ) is increasing. Third, I assume  $u - c(u) > 0$  over the relevant range of prices, which implies consumers derive some surplus beyond the cost they impose on the insurer. This condition is valid, for instance, if all consumer are sufficiently risk averse or if moral hazard is sufficiently small. Fourth, I assume the relevant range of prices is always below the monopoly price.<sup>7</sup> These assumptions imply the following:

**Lemma 1.**  $\pi'' < 0$ ,  $AC' < 1$  and  $p^* = AC(p^*)$  is the unique pure-strategies Nash equilibrium.

## 22 Price Discrimination

PD becomes relevant if firms observe a signal that partitions the set of consumers into multiple subsets. Suppose two such subsets exist, indexed by  $m \in \{A, B\}$  (say, young and old). Let the subscript  $m$  identify the demand, price, and so on in each set.<sup>8</sup> I assume, for simplicity, that the same (unit) mass of consumers exists in each market.<sup>9</sup> Notice that I consider an arbitrary partition  $\{A, B\}$ , in contrast to [Levin \(2001\)](#) who assumed that the subsets  $A, B$  are ranked by strong set order.

I continue to assume free entry of firms into the industry. However, if there are constraints on PD, firms cannot reject a willing buyer, because such rejection would amount to PD. Therefore, each firm must serve both consumer subsets and, in equilibrium, the industry will charge prices  $p_A, p_B$  such that firms break even across both consumer subsets:

$$\pi_A(p_A) + \pi_B(p_B) = 0. \quad (1)$$

Under full PD, prices are  $p_A = p_A^*$  and  $p_B = p_B^*$ , such that  $\pi_A(p_A^*) = \pi_B(p_B^*) = 0$ .

<sup>7</sup>The assumptions in [Appendix A](#) imply  $AC(p)$  is either decreasing or a contraction, and therefore has a unique fixed point. The condition is equivalent to the assumption in [Mahoney and Weyl \(2014\)](#) that average cost (as a function of quantity) is everywhere less steep than inverse demand.

<sup>8</sup>For instance,  $Q_A(p_A)$  is demand in subset  $A$ ;  $p_B^*$  is the equilibrium price in subset  $B$  under full PD.

<sup>9</sup>It would be straightforward to scale each market up by a different factor capturing its size.

Under full CR, firms charge  $\bar{p}$  to all consumers such that  $\pi_A(\bar{p}) + \pi_B(\bar{p}) = 0$ .<sup>10</sup> At full CR, the industry makes a loss on one subset of individuals and a profit on the other, breaking even overall.<sup>11</sup> Henceforth, I use the superscripts  $\star$ ,  $\star\star$ , and  $\bar{\phantom{x}}$  to denote functions evaluated at  $p^\star$ ,  $p^{\star\star}$ , and  $\bar{p}$ , respectively.<sup>12</sup>

WLOG, I define consumer subset  $A$  as the “high-cost” subset in the sense that, at  $\bar{p}$ , the industry makes a loss on  $A$  and a profit on  $B$ :

$$\pi_A(\bar{p}) < 0 < \pi_B(\bar{p}).$$

This condition is equivalent to  $p_A^\star > \bar{p} > p_B^\star$  or  $AC_A(\bar{p}) > \bar{p} > AC_B(\bar{p})$ , hence subset  $A$  being “high cost.” Consumer subset  $B$  is the “low-cost” set.

I consider a continuum of policies between full PD to full CR. Let  $\chi \in [0, 1]$  be the “CR policy.”<sup>13</sup> For each set  $m$ , let each price  $p_m$  be determined by the functions  $p_m(\chi)$  defined implicitly by

$$\pi_m(p_m(\chi)) = \chi \bar{\pi}_m. \quad (2)$$

At full PD ( $\chi = 0$ ), equation 2 becomes  $\pi_m(p_m) = 0 \Rightarrow p_m = p_m^\star$ . At full CR ( $\chi = 1$ ),  $\pi_m(p_m) = \bar{\pi}_m \Rightarrow p_m = \bar{p}$ . The break-even condition (equation 1) is satisfied at each  $\chi \in [0, 1]$ , because  $\sum \pi_m(p_m(\chi)) = \chi \sum \bar{\pi}_m = 0$ . Then,

$$\frac{dp_m}{d\chi} = \frac{\bar{\pi}_m}{\pi'_m}.$$

Since profit is concave and all prices remain below the relevant monopoly price, then  $\pi'_m(p_m) > 0$ . Because  $\bar{\pi}_A < 0 < \bar{\pi}_B$ , an increase in the CR policy  $\chi$  lowers the price of the high-cost market ( $\frac{dp_A}{d\chi} < 0$ ) and increases the price of the high-cost market ( $\frac{dp_B}{d\chi} > 0$ ).

<sup>10</sup>Full CR implies  $\bar{p} = \frac{1}{Q_A + Q_B} (Q_A AC_A + Q_B AC_B)$ . As was the case when there was a single set of consumers, the price equals the average cost in the entire market (the quantity-weighted average of  $AC_m$  in each subset).

<sup>11</sup>Notice that the industry makes a loss on some consumers and a profit on others even when a single consumer set is involved.

<sup>12</sup>For instance,  $Q_m^\star = Q_m(p_m^\star)$  and  $\bar{AC}_m = AC_m(\bar{p}_m)$ .

<sup>13</sup>I discuss below how a particular  $\chi$  can be implemented through standard policies such as a price ceiling.

## 23 Optimal CR policy

My goal is to characterize the policy  $\chi$  that maximizes total welfare,

$$W(\chi) = W_A(p_A(\chi)) + W_B(p_B(\chi)).$$

Henceforth, I will define welfare as a function of the policy  $\chi$ . My focus is not on policies (like subsidies, etc) that implement the socially optimal prices, which is the focus of other articles like [Geruso \(2016\)](#). Instead, I focus on maximizing the welfare that can be obtained, in equilibrium, through an appropriate CR policy  $\chi$ .

Notice that, since prices increase for one consumer subset and decrease for the other, CR policy will never result in a Pareto improvement. There will necessarily be a redistribution from subset  $A$  (where price falls) to subset  $B$ . However, this process can result in higher overall consumer surplus.

Let  $\bar{\pi} = \bar{\pi}_B = -\bar{\pi}_A$ . The effect of the CR policy  $\chi$  on total welfare has the following simple expression:

$$\frac{dW}{d\chi} = \bar{\pi} \left( \frac{Q_A}{\pi'_A} - \frac{Q_B}{\pi'_B} \right).$$

Moreover, under the regularity conditions assumed above, it can be shown that welfare is strictly concave in the policy  $\chi$ .

**Lemma 2.**  *$W(\chi)$  is strictly concave.*

The regularity assumptions described above ensure that, in each market, the deadweight loss is increasing and convex in price. As  $\chi$  increases,  $p_A$  falls and welfare in subset  $A$  increases but at a decreasing rate. Conversely, as  $p_B$  increases, the marginal effect on welfare in consumer subset  $B$  is increasingly severe. Because  $W(\chi)$  is concave, the welfare-maximizing CR policy (denoted  $\tilde{\chi}$ ) is full PD ( $\tilde{\chi} = 0$ ) if and only if  $W'(0) < 0$ . The welfare-maximizing policy is full CR ( $\tilde{\chi} = 1$ ) if and only if  $W'(1) > 0$ . Otherwise, a unique interior policy  $\tilde{\chi}$  exists that maximizes welfare. The remainder of the analysis presents conditions on primitives for each of these scenarios to be the case.

**Proposition 1** (Full PD). *Full PD maximizes welfare ( $\tilde{\chi} = 0$ ) if*

$$AC'_A - AC'_B < 0,$$

*where quantities are evaluated at the full PD prices  $p_m^*$ .*

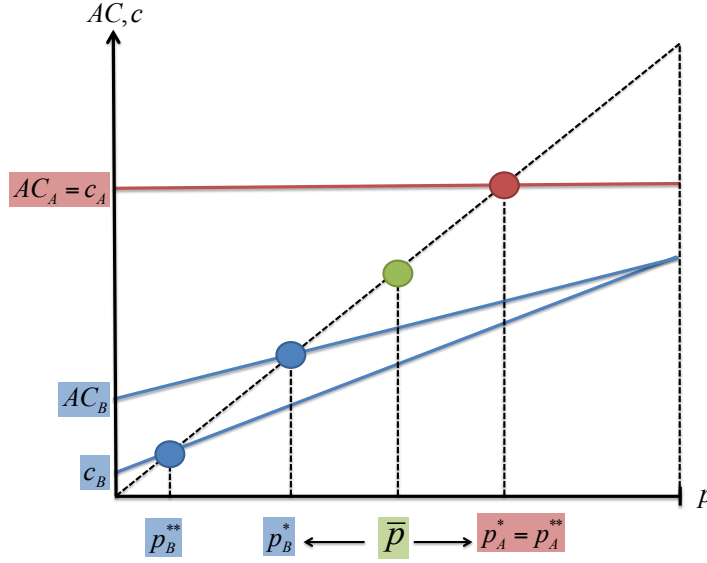


Figure 2: In this example, PD increases efficiency in both markets, relative to CR. The high-cost market A has no selection, while the low-cost market B has adverse selection.

Full PD is optimal when the low-cost consumer subset ( $B$ ) has more significant adverse selection ( $AC'_B$  positive and large) than the high-cost subset ( $A$ ), at the full PD prices ( $p_A^*, p_B^*$ ). To build intuition, suppose  $AC'_A = 0$  (no selection in the high-cost subset) and  $AC'_B > 0$  (adverse selection in the low-cost subset), as illustrated by Figure 2. Full PD allows  $p_A = p_A^*$ , resulting in full efficiency in set A, because this set has no selection. Similarly, PD induces  $p_B = p_B^* < \bar{p}$ , which increases welfare in set B since  $p_B^{**} < p_B^*$ . Therefore, in this example, PD brings both sets closer to their efficient outcomes.

Notice the environment illustrated in Figure 2 is much more extreme than the condition required by Proposition 1. In this case, full PD increases efficiency in both markets, whereas PD will typically increase welfare in one subset and reduce it in the other. Proposition 1 describes sufficient conditions for the net effect to be positive at full PD ( $\chi = 0$ ). Full PD can be the welfare-maximizing policy when both consumer subsets feature advantageous selection, or both feature adverse selection. For instance, when both sets are advantageously selected, full PD is optimal if  $0 > AC'_B(p_B^*) > AC'_A(p_A^*)$ .

However, Proposition 1 seems empirically unlikely to hold. As [Hendren \(2013\)](#)

emphasizes, “there is one way to be healthy, but many (unobservable) ways to be sick.” That is, more significant private information, and thus adverse selection, tends to exist among the high-cost individuals, which implies  $AC'_B < AC'_A$ . This pattern is likely to be present in markets such as life, health, and auto insurance, where risks are negative outcomes. A similar pattern is described in [Brown et al. \(2014\)](#).

When the high-cost market features a greater adverse selection distortion, the first unit of CR has a positive value. In that case, some amount of CR is likely to increase overall welfare,  $\tilde{\chi} > 0$ . To build intuition, suppose no selection exists in the low-cost consumer subset  $B$ , so  $p_B^* = c_B$ . An infinitesimal increase in  $\chi$  would increase  $p_B$  and lower  $p_A$ . This rise in  $p_B$  would not have a first-order effect on welfare in set  $B$ , because that subset is price at marginal cost. However, the drop in  $p_A$  would increase welfare in set  $A$  if it were adversely selected:  $\frac{dW_A}{dp_A} = Q'_A (p_A^* - c_A^*) = Q'_A (AC_A^* - c_A^*) < 0$ . Therefore, some amount of CR is likely to be optimal. The following result characterizes the interior optimal CR policy.

**Proposition 2 (Optimal CR).** *The interior welfare-maximizing CR policy  $0 < \tilde{\chi} < 1$  satisfies*

$$\sigma_A (p_A - c_A) = \sigma_B (p_B - c_B).$$

The intuition for the expression in Proposition 2 is the following. First, the effect of a change in price on welfare in each market is  $\frac{dW_m}{dp_m} = Q'_m (p_m - c_m)$ . Moreover, as  $\chi \rightarrow 1$ , (policy increases toward full CR), it is as if each consumer subset's average cost curve  $AC_m$  converges to the market-wide average cost curve.<sup>14</sup> This shift will be large if  $m$  is a small share of the overall market, and vice versa. For this reason, the change in  $p_m$  is proportional to  $\frac{1}{Q_m}$ . Then, the definition  $\sigma_m = -\frac{Q'_m}{Q_m}$  yields the result.

As the CR policy  $\chi$  increases,  $p_A$  falls and  $p_B$  rises. If both markets feature adverse selection, the selection distortion is reduced in  $A$  but is exacerbated in  $B$ . The optimal amount of PD occurs when the value of the marginal distortion of the two sets is equal.

As mentioned above, it is empirically common that  $AC'_A > AC'_B$ , so some amount of CR is typically welfare increasing. Constraining price differences across sets is worthwhile because the welfare loss is increasing and convex in the price distor-

<sup>14</sup>The market-wide average cost is  $\frac{Q_A AC_A + Q_B AC_B}{Q_A + Q_B}$ .

tion. If subset A has a larger price distortion, the reduction in  $p_A$  induced by CR will achieve a large welfare gain relative to the welfare loss caused by the accompanying rise in  $p_B$ . However, as market A approaches its efficient price, the marginal effect of an additional price correction becomes small. Simultaneously, the marginal welfare loss from the increase in  $p_B$  grows large. For this reason, an interior point  $\tilde{\chi}$  typically exists at which the market as a whole no longer benefits from additional CR.

Intuitively, an increase in the semi-elasticity of the high-cost market ( $\sigma_A$ ) leads to an increase in the optimal policy  $\tilde{\chi}$  because welfare becomes more sensitive to price in the market where price is falling (A). Conversely, an increase in  $\sigma_B$  implies welfare becomes more sensitive to price in the market where price is rising. Therefore, the cost of CR increases, so  $\tilde{\chi}$  decreases.<sup>15</sup>

It is possible that the benefit of reducing  $p_A$  is larger than the cost of increase  $p_B$ , even at  $\chi = 1$ . This can occur if the adverse selection distortion in subset A is much larger than in subset B. Alternatively, full CR can be optimal if the changes in prices induced by CR are relatively small, which occurs when the costs in the two markets are very similar. This intuition is formalized in the following result.

**Proposition 3 (Full CR).** *Full CR maximizes welfare ( $\tilde{\chi} = 1$ ) if*

$$0 < \frac{\sigma_B \frac{1}{Q_B} + \sigma_A \frac{1}{Q_A}}{\frac{1}{Q_B} + \frac{1}{Q_A}} (AC_A - AC_B) < AC'_A - AC'_B, \quad (3)$$

where all quantities are evaluated at  $\bar{p}$ .

Recall that  $\tilde{\chi} = 0$  when  $AC'_A - AC'_B < 0$  (at full PD prices). For  $\tilde{\chi} = 1$ , I require not only that  $0 < AC'_A - AC'_B$  (at  $\bar{p}$ ), but that this term must be sufficiently large. The term  $AC'_A - AC'_B$  captures the benefit of CR. When  $0 < AC'_A - AC'_B$ , lowering  $p_A$  implies a large welfare gain if subset A has a larger adverse selection distortion. CR simultaneously increases  $p_B$ , but this implies a small welfare loss if set B has little adverse selection.<sup>16</sup>

However, under the regularity conditions assumed above, the marginal benefit of CR is decreasing in subset A and increasing in subset B. If there are large differences in the level of cost between subsets A and B, then CR will induce large changes in the price. Then, it is more likely that the unit of CR ( $\chi = 1$ ) has a small

<sup>15</sup>Appendix A includes a proof of this result for the case where  $\sigma_m$  is fixed.

<sup>16</sup>Indeed, if  $AC'_B < 0$  (advantageous selection in set B), the rise in  $p_B$  would result in an increase in efficiency in set B.

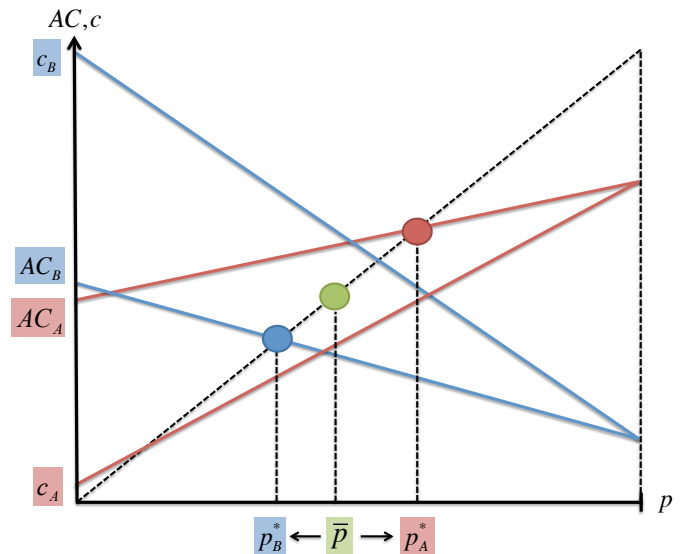


Figure 3: In this (extreme) example, CR increases efficiency in both markets, relative to PD. The high-cost market has adverse selection, while the low-cost market has advantageous selection

marginal gain in subset  $A$  and a large marginal loss in subset  $B$ . Therefore, when markets differ significantly in the level of their costs, full CR is only the welfare maximizing policy if  $AC'_A - AC'_B$  is sufficiently large. Moreover, when demands are elastic ( $\sigma_m$  large), welfare decreases rapidly in market  $B$  as  $p_B$  rises, so the last unit of CR is more likely to induce a large welfare loss in market  $B$  and a small gain in market  $A$ . In sum, if costs differ significantly across markets, then CR is a powerful tool so full CR is likely to excessively constrain prices but some amount of CR can still increase welfare relative to full PD.

The following corollary is illustrative. Suppose that  $\sigma_A(p) = \sigma_B(p) = \sigma(p)$  at every  $p$ , as assumed in [Chen and Schwartz \(2013\)](#). Then, full CR is optimal if  $\bar{c}_A < \bar{c}_B$ . Since  $\bar{AC}_A > \bar{AC}_B$ , full CR is only optimal if the marginal and average cost in each of the two markets are on opposite sides of the price  $\bar{p}$ . Figure 3 illustrates this scenario. In this case,  $AC'_A > 0 > AC'_B$  and, moreover, the level of cost is similar at  $\bar{p}$ . Such a scenario is extreme but simple to visualize, because CR increases efficiency in both sets relative to PD.

In general, CR increases welfare in one market and reduces it in the other. Full CR can be the welfare-maximizing policy when both consumer subsets are adversely



selected, both are advantageously selected, or any combination. That is, the condition described by Proposition 3 is much less extreme than the environment described in Figure 3.

In Appendix B presents an extension of the baseline model to the case where a signal partitions consumers into multiple subsets. All intuitions extend to that model in a straightforward way. In this setting, there are multiple “high-cost” consumers subsets (where prices falls under full CR), and similarly multiple “low-cost” subsets (where prices rise under full CR). The conditions for the optimality of full PD and full CR are more restrictive in this case. For instance, full PD is the welfare maximizing policy when, at full PD, all low-cost markets have a higher slope of their average cost than any high-cost market. The optimal interior CR policy equates the benefit of additional CR in all high-cost markets, to the cost of additional CR in all low-cost markets.

## 24 Implementation

To implemented a given value of  $\chi$ , the regulator could directly choose the price  $p_B$ , a floor on  $p_B, p_A$ , or a ceiling on  $p_A$ . The regulator could choose the maximum price difference between the two prices ( $p_A, p_B$ ) or impose a price per unit of the price difference  $p_A - p_B$ .

The CR policy  $\chi$  effectively parameterized a path between full PD and full CR. There are multiple possible parameterizations, other than the one described by Equation 2. In the baseline model (one product and two consumer sets), any one-dimensional parameterization that induces monotonic paths for  $p_B$  yields the same results. Because  $\pi'_m > 0$ , for each value of  $p_B$ , the value of  $\pi_B$  is uniquely determined. Then, by Equation 1, the values of  $\pi_A, p_A$  are also uniquely pinned down. Therefore, all such parameterizations are equivalent.

In general, a one-dimensional policy involves some loss of generality. If multiple sets exist, the path of price for one consumer set does not entirely pin down the paths of profit for the other sets.

## 25 Extension to Two products

I extend the analysis to a market with two products and mandatory purchase, thereby considering a setting more similar to that of UK annuities. The model setup is similar to that of [Handel, Hendel and Whinston \(2015\)](#). I present here the main intuition

for the results, and defer the details to Appendix C.

I begin by describing a given consumer set (thus omitting the subscript  $m$ , which will capture consumer subsets). In this setting, firms offer two products, indexed by  $k \in \{H, L\}$ .<sup>17</sup> Firms compete in prices. The price for  $H$  and  $L$  are  $p_H$  and  $p_L$ , respectively. The difference in prices between the two products is  $\Delta p = p_H - p_L$ .

Consumers must purchase one of the two products. Let  $u$  capture WTP for  $H$  over  $L$ , so a consumer buys  $H$  when  $u > \Delta p$ , and buys  $L$  otherwise. Let  $u$  have smooth PDF  $f(u)$ . The demands for the two products are linked. If demand for  $H$  is  $Q_H = Q(\Delta p) = \int_{\Delta p}^{\bar{u}} f(u) du$ , then demand for  $L$  is  $Q_L = 1 - Q(\Delta p)$ . I define the semi-elasticities  $\sigma_H = -\frac{Q'}{Q} > 0$  and  $\sigma_L = -\frac{Q'}{1-Q} > 0$ .

The cost of those with WTP  $u$  in contracts  $H$  and  $L$  is  $c_H(u)$  and  $c_L(u)$ , respectively. I assume  $H$  is more costly for any individual:  $c_H(v) > c_L(v), \forall v$ . That is, one can think of  $H$  as a “comprehensive product” and  $L$  as a “bare-bones product.” Let  $\Delta c(\Delta p) = c_H - c_L$ .

The average cost of all individuals that choose products  $H$  and  $L$  is  $AC_H(\Delta p) = \mathbb{E}[c_H | u > \Delta p]$  and  $AC_L = \mathbb{E}[c_L | u < \Delta p]$ , respectively. I define the difference in the average costs between the two products as  $\Delta AC = AC_H - AC_L$ .

Industry profit in contract  $k$  is  $\pi_k(p_H, p_L) = Q_k(\Delta p) [p_k - AC_k(\Delta p)]$ . Welfare is  $W(\Delta p) = \int_{\Delta p}^{\infty} (u - \Delta c(u)) f(u) du$ . The welfare maximizing  $\Delta p$  satisfies  $\Delta p^{**} = \Delta c(\Delta p^{**})$ .

I assume free entry of firms into each contract. Therefore, equilibrium requires that each contract  $k$  break even:

$$\pi_H(p_H^*, p_L^*) = \pi_L(p_H^*, p_L^*) = 0.$$

Equivalently,  $p_H^* = AC_H(\Delta p^*)$  and  $p_L^* = AC_L(\Delta p^*)$ , implying  $\Delta p^* = \Delta AC(\Delta p^*)$ .<sup>18</sup>

Since  $u$  captures WTP for product  $H$ , then product  $H$  is adversely selected if  $c'_k(u) > 0, \forall k$ . That is, those with higher WTP for  $H$  are particularly costly in either contract. From [Weyl and Veiga \(2016\)](#), adverse selection implies that  $\Delta AC(\Delta p) > \Delta c(\Delta p)$ . Therefore, if product  $H$  is adversely selected,  $\Delta p^* > \Delta p^{**}$ . Notice that adverse selection into  $H$  implies that marginal buyers of  $L$  are more costly than the average buyer of  $L$ . Therefore, adverse selection into  $H$  can equivalently be described as advantageous selection into  $L$ . Ultimately, the result is that “too

<sup>17</sup>For instance,  $L$  can be a baseline insurance plan and  $H$  a comprehensive plan.

<sup>18</sup>In this model, a pure-strategies Nash equilibrium does not necessarily exist. However, [Handel, Hendel and Whinston \(2015\)](#) show that a unique [Riley \(1979\)](#) equilibrium always exists.

few” individuals buy  $H$ , and “too many” buy  $L$ .

It is worth emphasizing the analogies of this model to the baseline model. In the baseline model, individuals faced a binary choice between two “products,” namely “buy” and “not buy.” The value, price and cost of “not buy” were fixed at zero. Product “buy” was adversely selected when its infra-marginal buyers were costlier than its marginal buyers. If “buy” was adversely selected, its price relative to “not buy” was higher than what it would be under symmetric information, so “too few” people chose “buy,” and too many people chose “not buy.”

In the setting with two products, individuals make a binary choice between  $H$  and  $L$ . The value, price and cost of  $L$  are not zero. Product  $H$  is adversely selected when its infra-marginal buyers are costlier than its marginal buyers. If  $H$  is adversely selected, its price relative to  $L$  is higher than under symmetric information, so “too few” people chose  $H$  and too many people chose  $L$ . In this case, the infra-marginal buyers of  $L$  are less costly than its marginal buyers.

I now turn to the matter of CR in this setting. Suppose that a given signals partitions consumers into the subsets  $m \in \{A, B\}$ . The full PD equilibrium requires that the four prices  $(p_{HA}^*, p_{LA}^*, p_{HB}^*, p_{LB}^*)$  satisfy

$$\pi_{HA}(p_{HA}^*, p_{LA}^*) = \pi_{LA}(p_{HA}^*, p_{LA}^*) = \pi_{HB}(p_{HB}^*, p_{LB}^*) = \pi_{LB}(p_{HB}^*, p_{LB}^*) = 0.$$

At the full CR equilibrium, each product  $(H, L)$  breaks even across the two consumer subsets  $A, B$ . The full CR equilibrium requires that the prices  $\bar{p}_H, \bar{p}_L$  satisfy

$$\pi_{HA}(\bar{p}_H, \bar{p}_L) + \pi_{HB}(\bar{p}_H, \bar{p}_L) = \pi_{LA}(\bar{p}_H, \bar{p}_L) + \pi_{LB}(\bar{p}_H, \bar{p}_L) = 0.$$

Let  $\Delta p_m^* = p_{Hm}^* - p_{Lm}^*$  and  $\bar{\Delta}p = \bar{p}_H - \bar{p}_L$ . I assume that  $A$  is the high-cost consumer subset in the sense that the relative price falls in market  $A$  under full CR, relative to full PD. That is,  $\Delta p_A^* > \bar{\Delta}p$ . Similarly, I assume that  $B$  is the low-cost subset:  $\Delta p_B^* < \bar{\Delta}p$ .

I consider a CR policy  $\chi \in [0, 1]$ . For each market  $m$  and good  $k$ , the path of both prices in market  $m$ ,  $p_{Hm}(\chi), p_{Lm}(\chi)$ , are defined by

$$\begin{bmatrix} \pi_{Hm}(p_{Hm}(\chi), p_{Lm}(\chi)) \\ \pi_{Lm}(p_{Hm}(\chi), p_{Lm}(\chi)) \end{bmatrix} = \chi \begin{bmatrix} \pi_{Hm}^* \\ \pi_{Lm}^* \end{bmatrix}, \quad (4)$$

where  $\pi_{Hm}^*, \pi_{Lm}^*$  are the levels of profit in market  $m$  under full CR. Notice that  $-\pi_{HA}^* =$

$$\pi_{HB} = \pi_H \text{ and } -\pi_{LA} = \pi_{LB} = \pi_L.$$

Appendix A shows that this system is globally invertible. That is, for any pair of profits  $(\chi\pi_{Hm}, \chi\pi_{Lm})$ , there is a unique pair of prices  $(p_{Hm}(\chi), p_{Lm}(\chi))$  such that Equation 4 is satisfied. The proof requires showing that the Jacobian of the left hand side of Equation 4 is everywhere non-vanishing. Invertibility then follows by the Hadamard-Caccioppoli Theorem. Assuming concavity of welfare in  $\chi$  yields the following results.

**Proposition 4 (Full PD (2)).** *With two products, full PD is optimal ( $\tilde{\chi} = 0$ ) if  $Q_B > Q_A$ , and*

$$\Delta AC'_A(\Delta p_A^*) - \Delta AC'_B(\Delta p_B^*) < 0,$$

where all quantities are evaluated at the full PD prices  $p_m^*$ .

This condition requires that adverse selection into product  $H$  is less significant among the high-cost subset of consumers ( $A$ ) than among  $B$ . In this setting, adverse selection of contract  $H$  relative to contract  $L$  in subset  $m$  is captured by  $\Delta AC' > 0$ .

However,  $\tilde{\chi} = 0$  requires the additional assumption that, greater share of individuals in subset  $B$  purchase product  $H$  ( $Q_B^* > Q_A^*$ ). Intuitively, this condition is satisfied when individuals in market  $B$  have derive a greater surplus from the comprehensive product  $H$  than individuals in  $A$ . The intuition for this condition is that CR would increase the price  $\Delta p_B$  reducing this surplus.

**Proposition 5 (Full CR (2)).** *With two products, the interior optimal CR policy  $\tilde{\chi}$  satisfies*

$$\frac{\sigma_{HB}(p_{HB} - c_{HB}) + \sigma_{LB}(p_{LB} - c_{LB}) - 1}{X_B} = \frac{\sigma_{HA}(p_{HA} - c_{HA}) + \sigma_{LA}(p_{LA} - c_{LA}) - 1}{X_A},$$

where

$$X_m = Q_m \left[ \frac{\pi_H}{Q_m} - \frac{\pi_L}{1 - Q_m} \right].$$

The optimal policy equates the marginal distortion on markets  $A$  and  $B$ . In each market, the condition considers the distortion in each product's price, weighted by the semi-elasticity of that product's demand, which reflects how sensitive welfare is to a change in that product's price. Moreover, the marginal distortion in each market is weighted by  $\frac{1}{X_m}$ , which is increasing in  $Q_m$ . That is, a market  $m$  with a greater surplus from purchasing  $H$  will have a larger  $Q_m$  and therefore the price

distortion in that market will be weighted more heavily. Notice that when  $\bar{\pi}_L = 0$ ,  $X_A = X_B = \bar{\pi}_H$  and  $\sigma_{LA} = \sigma_{LB} = 0$ , I recover Proposition 2. For brevity, I present and describe the conditions under which full CR is optimal ( $\tilde{\chi} = 1$ ) in Appendix C.

Appendix H presents a calibration of the welfare results of the model above, based on the obtained by Handel, Hendel and Whinston (2015) in the context of employer-provided health insurance. Like the empirical application below, the results suggest that CR policy can significantly reduce the deadweight loss caused by selection.

### 3 Context and Data

The theoretical predictions described above show that CR policy (i.e., restricting the contractibility of a given signal) can increase consumer surplus. However, the optimal price constraint and the welfare gain it induces are ultimately empirical questions. To illustrate the potential benefit of CR, I consider the effect of age- and gender-based CR policies in the context of UK annuities. I begin by describing the setting and the data. Then I will introduce a structural model of annuity choice. I describe my estimation procedure including the sources of identifying variation. Finally, I use the estimates to simulate the welfare effects of age-based and gender-based CR in this context.

#### 31 Annuities in the UK

Annuities guarantee the buyer an income while she lives, thereby providing insurance against poverty in old age. Until 2015, the UK had a semi-compulsory annuity market. About 20% of individuals were mandated to make mandatory tax-preferred contributions, throughout their working lives, from their wages into a pension fund.<sup>19</sup> I will denote the amount in this pension fund, at retirement, by  $\phi$ . Upon retirement, this fund had to be used to purchase an annuity from a menu of possible contracts, described below.<sup>20</sup> In 2013, 353K annuities were sold in the UK, worth a total of about £12bn.<sup>21</sup>

<sup>19</sup>The share of individuals making in “defined contribution” pension schemes in 2013 was about 20% (<http://tinyurl.com/z5bhbu>).

<sup>20</sup>Individuals can withdraw 25% of this pension fund without penalty, and virtually all individuals do so.

<sup>21</sup>ABI, The UK Annuity set: Facts and Figures, February 2014, <http://tinyurl.com/nto7bv>.

The compulsory annuitization scheme is known to individuals during their working life, although the choice of annuity contract occurs only once an individual retires. Annuity rates vary over time, so the rates offered to each individual are only observed at the time of retirement. During this period, the typical retirement age for men was 65, and for women was 60.

The market is quite competitive. In 2013, 14 providers offered annuities. UK regulation authorities have made promoting competition in the annuities market a priority, for instance, by creating online resources that facilitate the comparison between annuities (see Appendix F). Individuals are actively encouraged to compare annuity providers. Annuity rates are similar across different firms, again suggesting a competitive industry. Moreover, based on the life expectancies and rates observed in the data, firms seem close to breaking even. Using similar data (from an earlier period), EFS also find observed rates are close to breaking-even levels.

## 32 Annuity contracts

An annuity prescribes a path of income over time, with payments conditional on the buyer being alive.<sup>22</sup> Annuity contracts are characterized by a rate  $r \in [0, 1]$ . If an individual with pension fund  $\phi$  obtains an annuity with rate  $r$ , her yearly nominal income from the annuity is  $y = \phi r$ . Rates are typically in the region of  $r \in [0.05, 0.1]$ , that is, 5%-10%.<sup>23</sup>

In this setting, the main dimension of choice for individuals is the annuity's guarantee period, denoted  $g$ . If an individual chooses a guarantee of  $g$  years, payments accrue to the buyer of her estate for  $g$  years, even if the buyer dies during this period. Beyond the period of the guarantee, payments are conditional only on the survival of the annuitant. Longer guarantees  $g$  are associated with lower rates  $r$ , and thus lower payments while the individual is alive. Throughout the industry, annuity sellers offer guarantees of  $g \in \{0, 5, 10\}$  years.<sup>24</sup> I will refer to these contracts as  $g_0, g_5, g_{10}$ , and to their respective rates as  $r_0, r_5, r_{10}$ . The vector of all rates is  $r = [r_0, r_5, r_{10}]$ .

Other dimensions of choice exist beyond  $g$  but, in practice, very little variation

<sup>22</sup>Stochastic payments (e.g., linked to a portfolio of investments) and annuities explicitly linked to inflation are not sold during the period covered by the data. Notice that inflation is not an idiosyncratic risk and therefore would be particularly costly for insurers to cover.

<sup>23</sup>EFS use data from an earlier period, where rates are significantly higher.

<sup>24</sup>Some consumers choose intermediate values of guarantee, but they are an extremely small minority, which I exclude from my sample.

is present in choice along these dimensions, so I restrict my sample in a way that abstracts from these additional dimensions. For instance, individuals can choose income profiles that increase over time at a rate of 3% per year, but only about 4% of individuals chose such a contract. Appendix D describes other dimensions of contract choice that behave in a similar way. A number of additional institutional details about the UK annuities market can be found in Banks and Emmerson (1999), Murthi, Orszag and Orszag (2000), Finkelstein and Poterba (2002), Finkelstein and Poterba (2004) and Einav, Finkelstein and Schrimpf (2010).<sup>25</sup>

### 33 Choices and adverse selection

In the empirical model that I use, an individual's choice over guarantee lengths will depend on three factors: mortality, bequest motives and rates ( $r$ ). I will use a parameter  $\alpha$  (described in further detail below) to capture individual mortality. That is, individuals with high  $\alpha$  are less likely to be alive at any given future time. Since payments accrue while the annuitant is alive (or to the end of the guarantee period), long-lived individuals (low  $\alpha$ ) are the most costly buyers of any contract.

Moreover, I will use a parameter  $\beta$  to capture the significance on individual bequest motives. Individuals with high  $\beta$  derive significant utility from having large savings, at any given point in time, which they can pass onto their spouse or children in the event of their death. Since  $\beta$  merely captures individual preferences, it does not directly affect the firm's cost.

Individuals with high mortality (high  $\alpha$ ) and high bequest values (high  $\beta$ ) prefer longer guarantees like  $g_{10}$ . These individuals are likely to die relatively young so a long guarantee significantly affects the payments received by their descendants. Also, they place a high value on leaving bequests to their descendants. Therefore, the benefit of a long guarantee is likely to outweigh the smaller rates associated with contracts such as  $g_{10}$ .

Individuals with high longevity (low  $\alpha$ ) and low bequest motives (low  $\beta$ ) are likely to prefer short guarantees like  $g_0$ . These individuals do not care significantly about bequests, and their high longevity means they are unlikely to be affected by guarantees. Therefore, they are attracted by the higher rates of contracts such as  $g_0$ .

First, suppose that  $\alpha$  and  $\beta$  are distributed independently. Since, long-lived individuals (low  $\alpha$ ) are likely to choose  $g_0$ , one would expect this contract to be adversely selected. The high-cost infra-marginal buyers of  $g_0$  raises the cost of  $g_0$

---

<sup>25</sup>For a comparison with the annuities set in the United States, see Mitchell et al. (1999).

thereby lowering  $r_0$ . Simultaneously, high-mortality individuals (high  $\alpha$ ) prefer long guarantees like  $g_{10}$ . This lowers the cost of  $g_{10}$  and raises  $r_{10}$ . Therefore,  $g_0$  is likely to be adversely selected relative to  $g_{10}$ , while  $g_{10}$  is likely to be advantageously selected relative to  $g_0$ . If this is the case, relative to a setting with symmetric information, individual choices are skewed towards  $g_{10}$  and away from  $g_0$ .

However, suppose that there was a strong negative correlation between  $\alpha$  and  $\beta$ . In this case, individuals choosing  $g_{10}$  because of high  $\beta$  will tend to be low-mortality (low  $\alpha$ ). This raises the cost of  $g_{10}$  and lowers  $r_{10}$ . Conversely, individuals choosing  $g_0$  because of their low  $\beta$  will tend to be high-mortality (high  $\alpha$ ). These individuals will lower the cost of  $g_0$  and increase  $r_0$ . In this case,  $g_{10}$  is adversely selected, while  $g_0$  is advantageously selected. My empirical results suggest that, indeed,  $\alpha$  and  $\beta$  are negatively correlated, so equilibrium choices are skewed towards  $g_0$  and away from  $g_{10}$ , relative to a setting with symmetric information.

### 34 Observable Covariates

I use a proprietary dataset obtained from a large UK annuities seller.<sup>26</sup> Regarding the contract, the dataset includes the date of origination, the monthly payment, and the level of the guarantee. Otherwise, contracts are standardized in the industry.

Regarding individual-specific characteristics, the dataset includes age, gender, and the size of one's pension pot ( $\phi$ ). I observe whether the individual used a financial advisor to purchase the annuity. I also observe if the individual is "internal" (pension pot was held at the annuity seller prior to the annuity purchase) or "external." The dataset also includes whether the individual lived (at the time of purchase) in a post code with high, medium, or low longevity.

Importantly, the data include the firm's estimate of each consumer's life expectancy, which I will therefore treat as observable. The firm computes this expectation using all the data it collects, including contractible and non-contractible information. However, firms cannot price discriminate directly on the estimated life expectancy. I discuss in detail the variables on which firms were allowed to price discriminate, in Section 39 below.<sup>27</sup>

---

<sup>26</sup>The dataset is at the contract, not individual, level. However, individuals obtain better rates when they annuitize larger pots (larger  $\phi$ ), so no incentive exists to split one's pension fund into multiple contracts.

<sup>27</sup>I also observe whether each individual has died or not by March 2015. However, I do not use this information directly, and instead use the firm's estimate of individual life expectancy.



### 35 Sample restrictions

I focus on contracts originated between July 2006 and June 2008. Restricting the dataset to this short window limits the impact of changing macro-economics variables such as interest rates and inflation expectations.<sup>28</sup> The period of the sample is one of relative stability in terms of long-term interest rates, which are the major macro-economic determinants of annuity rates (Appendix D). This restriction also allows me to focus on a period during which there were no regulatory changes to the industry. Moreover, the restriction mitigates concerns regarding changing demographics over time.<sup>29</sup> After June 2008, the firm changed its pricing formula in a way that makes the empirical analysis below unfeasible.<sup>30</sup>

I focus on individuals purchasing annuities at ages 60 and 65. This subset includes the vast majority of individuals, because female and male retirement ages are 60 and 65, respectively. In particular, I will consider four distinct subsets of individuals: Men 65, Women 65, Men 60, and Women 60. My empirical analysis will consider each of these subsets independently.

### 36 Summary Statistics

Below are summary statistics for each of the four age-gender subsets I consider.

The entire sample consist of just over 11,000 individuals. The two largest sets are Men-65 and Women-60, given the ages of retirement in the UK. Across all subsets, approximately 65% of individuals choose a 5-year guarantee ( $g_5$ ), with the remaining consumers fairly evenly split between  $g_0$  and  $g_{10}$ . This pattern seems to suggest an increase in preference heterogeneity relative to the data used in EFS, where only about 3% of individuals chose  $g_{10}$ . The average size of the pension pot ( $\phi$ ) is about £15K. Life expectancy is predictably higher for women and for 60-year-olds. Appendix D includes additional descriptions of the data.

<sup>28</sup>Annuities are backed largely by UK government bonds, and therefore the yields on these bonds capture the cost for the insurer. When government bond yields rise, annuity payments also tend to rise.

<sup>29</sup>For instance, a modest spike occurs in the number of retirees in 2011, corresponding to the baby boomers born in 1946-47 following World War II.

<sup>30</sup>Effectively, the firm began price discriminating against consumers based on several other covariates that are not perfectly captured in the data. Therefore, after this time period, imputing non-observed prices (Section 39) becomes impossible.

Table 1: Summary Statistics by Group

|                   | Men 65 | Women 65 | Men 60 | Women 60 |
|-------------------|--------|----------|--------|----------|
| Garantee 10-yrs   | 0.181  | 0.184    | 0.198  | 0.224    |
| Garantee 5-yrs    | 0.703  | 0.614    | 0.680  | 0.623    |
| Internal          | 0.824  | 0.460    | 0.714  | 0.485    |
| Life Expectancy   | 23.58  | 26.14    | 28.64  | 31.14    |
| Fund (1000s)      | 14.31  | 19.73    | 17.04  | 19.87    |
| Financial Advisor | 0.382  | 0.712    | 0.451  | 0.641    |
| Postcode High     | 0.295  | 0.498    | 0.371  | 0.423    |
| Postcode Med      | 0.359  | 0.296    | 0.362  | 0.344    |
| Observations      | 3679   | 830      | 1733   | 4857     |

### 37 Representativeness

The population that purchases annuities is not representative of the UK overall. Annuitants are richer and live longer than non-annuitants ([Banks and Emmerson \(1999\)](#)). For instance, life expectancy for 65-year-old males in the UK is 18.6 years, but 23 years for annuity buyers.

However, individuals in the the data seem representative of the population of UK workers for whom annuity purchase is mandatory. For instance, [Banks and Emmerson \(1999\)](#) find that among compulsory annuity buyers, the median pension pot was about £20,000, which is is approximately true in my data as well. [Einav, Finkelstein and Schrimpf \(2010\)](#) find the average age of annuitization is 62, and about 5% of individuals choose a contract with payments increasing over time, which is also true in the data I use. My sample also seems similar to the individuals considered in other studies of the UK annuities market, namely, [Murthi, Orszag and Orszag \(1999\)](#); [Finkelstein and Poterba \(2002\)](#); [Einav, Finkelstein and Schrimpf \(2010\)](#).

Because I use data from a single firm, my analysis assumes that buyers of this firm are representative of the market overall. That is, I continue to assume a symmetric equilibrium, as in the theoretical model of Section 2.

The firm is also representative of the industry. For instance, all firms offer the same menu of guarantees (0, 5, and 10 years). Moreover, the pricing formula the

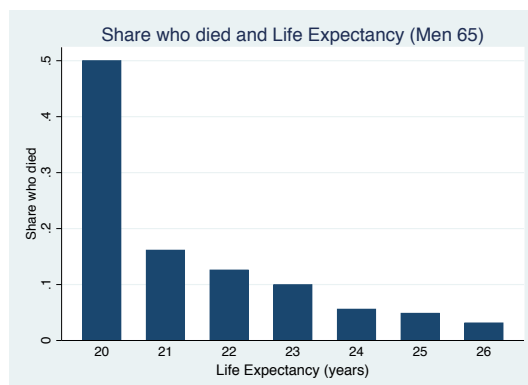


Figure 4: Among Men 65, the share of individuals who died before 2015 is monotonically decreasing in the firm's prediction of life expectancy.

firm uses (which I describe in detail below) is typical of the industry overall in the relevant time period.

### 38 Mortality

A crucial element of my analysis is the firm's estimate of each individual's life expectancy.<sup>31</sup> Figure 4 shows this measure is a good predictor of mortality rates in the data (which I observe but do not use in estimation). For instance, among Men 65, the share of individuals who died before 2015 is monotonically decreasing in the firm's prediction of life expectancy, as expected.<sup>32</sup>

### 39 Annuity Rates

During the period covered in the data, annuity rates depend on the individual's gender, age, pension fund size  $\phi$ , and guarantee  $g$ . Rates do not depend on any other individual-level characteristics, as illustrated in Figure 16. Moreover, annuity rates do vary over time (with interest rates, etc).

I observe the rate each buyer obtains on the contract she chooses, but not on

<sup>31</sup>The dataset also includes each individual's date of death. In principle, I can use the data to estimate the mortality parameters of individuals in the population, but in practice, the dataset is too recent to obtain reasonable estimates, because a small share of the sample has died.

<sup>32</sup>It is not possible to directly compare predicted longevity with true longevity, since the only observed death times are those that (stochastically) occurred particularly early. Life expectancy is a significant predictor of mortality hazard (and has the expected effect) in a Cox semi-parametric duration model (results available upon request).

the other contracts she could have chosen, so I impute these missing data. The firm follows a deterministic (but unobserved to the econometrician) rule for the rates offered to each consumer. This rule is of the type  $r_g = r_g(\phi, \tau)$ , where  $g$  indexes guarantee levels ( $g = \{0, 5, 10\}$ ),  $\phi$  is the size of the individual's pension pot, and  $\tau$  indexes time periods (months).<sup>33</sup>

I estimate the rate offered for contract  $g$  to individual  $i$  with fund  $\phi_i$  in month  $\tau$  as

$$r_{gi\tau} = r_g^\phi(\phi_i) + FE_\tau + \epsilon_{gi\tau},$$

where  $r_g^\phi(\phi)$  is an arbitrary continuous function of the fund size  $\phi$  and  $FE_\tau$  is a month  $\tau$  fixed effect. That is, I assume a baseline relationship  $r_g^\phi(\phi)$  exists that has a fixed shape, but the level of rates can change over time through the time fixed effects. Formally, this assumption corresponds to  $\frac{\partial r_{gi\tau}}{\partial \phi_i \partial \tau} = 0$ . I estimate  $r_g^\phi(\cdot)$  non-parametrically. Figure 17 shows the results of this exercise for Men 65. For additional details, see Appendix D3.

## 4 Contract Choice Model

### 4.1 Value of guarantee $g$

In this section, I model the value of a given annuity contract. I will consider an individual that is rational, forward looking, risk-averse and perfectly informed of her preferences and mortality.

Let time periods (in years) be indexed by  $t \in \{1, 2, \dots, T\}$  with  $T = 65$ .<sup>34</sup> Suppose that the individual retires at  $t = 1$  and lives, at most, until  $t = T$ . For a given individual  $i$ , the value of an annuity with guarantee  $g$  is  $V_{gi}$  defined by

$$V_{gi} = \max_{c_t, w_t} \sum_{t=1}^T \delta^t S_{ti} u_i(c_t) + \sum_{t=1}^{T+1} \delta^t H_{ti} v_i(w_t + G_t^g)$$

subject to :  $w_{t+1} = R(w_t + y_t - c_t)$ .

<sup>33</sup>Rates are quite stable over time, and I assume they are constant within each period of observation (one month). Therefore, I ignore the issues of selection due to missing prices emphasized by Erdem, Keane and Sun (1998) and others. If rates varied significantly within a month, individuals choosing a given contract might have been doing so because they retired at a point when the rates offered for that contract were relatively high.

<sup>34</sup>The results are robust to the choice of  $T$ . This is not surprising since, at  $t = T$ , every individual's survival probability is extremely small.

The discount rate is  $\delta \in (0, 1)$  and the interest rate is  $R > 1$ . In period  $t$ , consumption is  $c_t$  and remaining wealth (before the annuity payment) is  $w_t$ . The annuity payment in this period is  $y_t = \phi r$  as described above. The value  $V_g$  is determined assuming the path of  $c_t, w_t$  is chosen optimally to maximize lifetime utility. For individual  $i$ , the probability of surviving up to period  $t$  is  $S_{ti}$ , and the probability of dying during period  $t$  (the hazard rate) is  $H_{ti}$ , described further below.

If the individual is alive in period  $t$ , she obtains utility from consumption  $u_i(c_t)$ . If she dies during period  $t$ , she obtains utility from leaving a bequest in the amount  $w_t + G_t^g$ , where

$$G_t^g = \sum_{\tau=t}^{\tau=g} \frac{y_\tau}{R^{\tau-t}}$$

is the present value of the annuity payments up to the length of the guarantee  $g$ . The utility from bequests is determined by the function  $v_i(w_t + G_t^g)$ . Both utilities can depend on individual-level preferences, hence the subscript  $i$ .

This specification involves a number of assumptions. First, I assume that the individual's timing of retirement, as well as her labor supply and savings at retirement ( $w_1$ ), are exogenous to the choice of annuity contract. I also assume that  $\phi$  is exogenous to the guarantee choice, which seems plausible since  $\phi$  is the result of a lifetime of contributions.

I assume that individuals do not adjust their wealth portfolios in response to their choice of guarantee. For instance, individuals could plausibly hedge the risk of choosing  $g_0$  by investing a share of their remaining wealth into life insurance. These effects should be mild if the life insurance industry features the same adverse selection distortions as the annuities industry, or if there are significant transaction costs.

Third, I assume no moral hazard. That is, the survival process  $S_t, H_t$  are independent of the chosen path of  $c_t$  and of the contract  $g$ . The assumption of no moral hazard is common in annuity settings, where it is arguably less heroic than in contexts like health insurance.<sup>35</sup>

---

<sup>35</sup>However, Philipson and Becker (1998) argues that the presence of an annuity does tend to increase one's life expectancy.

## 42 Parameterization of mortality

I follow [Yaari \(1965\)](#) and the vast majority of the literature in modeling mortality risk as a non-stochastic discounting of each period.<sup>36</sup> I assume that  $S_{ti}$  follows a proportional hazard process (namely Gompertz) at the individual level.<sup>37</sup> From the perspective of  $t = 1$ , an individual of mortality type  $\alpha_i$  has a probability of being alive at time  $t \leq T$  of

$$S_{ti} = S(t | \alpha_i) = \exp \left[ \frac{\alpha_i}{\lambda} (1 - e^{\lambda t}) \right].$$

At  $t = 1$ , each individual's survival probability is close to 1, but it decreases at a rate which depends on  $\alpha$  and  $\lambda$ . The parameter  $\alpha_i$  captures mortality, with higher  $\alpha$  lowering  $S(t | \alpha)$  at every  $t \leq T$ . The parameter  $\lambda$  affects how quickly mortality hazard increases over time. All individuals necessarily die at  $t = T$ , so  $S_{T+1} = 0$ .

I will assume that  $\alpha_i$  is observable. Each consumer's life expectancy (LE) is observed in the data. LE at  $t = 1$  is  $LE = \int_1^\infty S(x | \alpha) dx$ . Assuming knowledge of  $\lambda$ , it is straightforward to compute the corresponding value of  $\alpha_i$ .<sup>38</sup> Then, the hazard probabilities of dying in each period  $t$  ( $H_{ti}$ ) can be computed as  $H_{ti} = \frac{1}{S_{ti}} (S_{ti} - S_{t+1,i})$ . Figure 5 illustrates the effect of  $\alpha_i$  on the Gompertz survival function  $S_{ti}$  and hazard function  $H_{ti}$ .

## 43 Parameterization of utility

I assume utility from consumption is CRRA:

$$u_i(c) = u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where  $c$  is consumption in a given time period(year), and  $1/\gamma$  is the inter-temporal substitution of consumption.<sup>39</sup> Moreover, I assume that utility from bequests is also

<sup>36</sup>As [Bommier \(2006\)](#) emphasizes, this effectively assumes individuals are risk averse with respect to their death times.

<sup>37</sup>Typically, the Gompertz survival probabilities are assumed to hold at the population level. The Gompertz process is widely used to describe adult mortality in developed countries ([Tableau, van den Berg Jeths and Heathcote \(2001\)](#); [Horiuchi and Coale \(1982\)](#)).

<sup>38</sup>The ability to take these costs as observable means that my estimation procedures focuses on preferences, as in [Handel \(2013\)](#) and HHW.

<sup>39</sup>There are no stochastic shocks in this model and therefore individuals do not have a precautionary savings motive. Therefore, the role of  $\gamma$  is to capture inter-temporal elasticity of substitution, rather than risk aversion. The distinction between the two (emphasized by [Epstein and Zin \(1989\)](#)) does not play a role here.

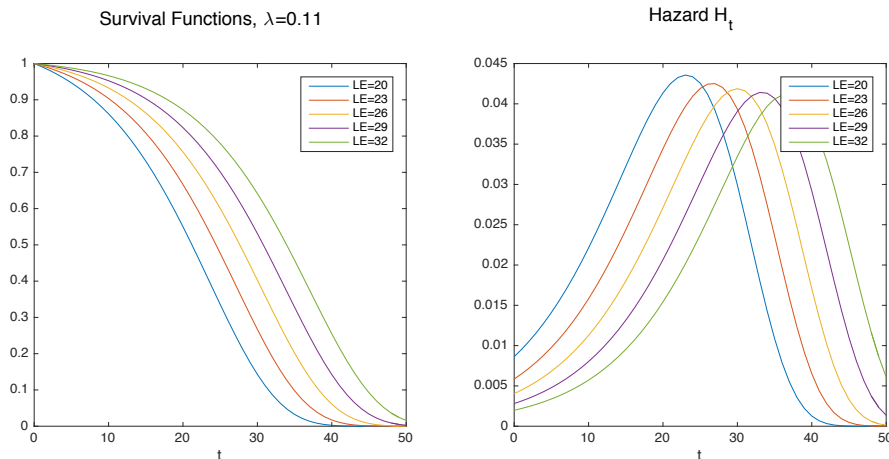


Figure 5: On the left, Gompertz survival curves for different values of life expectancy (LE), for fixed  $\lambda$ . On the right, an illustration of Hazard probabilities  $H_t$  for different values of  $\alpha$ .

CRRA:

$$v_i(w) = \beta_i \frac{w^{1-\gamma}}{1-\gamma},$$

where  $\beta_i$  measures the intensity of individual  $i$ 's bequest motives.

There are several reasons why individuals can value bequests. For instance, simple inter-generational altruism, “warm glow” giving ([Andreoni \(1989\)](#)) or regrets ([Braun and Muermann \(2004\)](#)). I remain agnostic as to the precise origin of these bequest motives.

The assumption that  $v(\cdot)$  and  $u(\cdot)$  share the same risk aversion parameter  $\gamma$  is not without loss of generality. However, it is common in the literature ([Einav, Finkelstein and Schrimpf \(2010\)](#); [Kotlikoff and Spivak \(1981\)](#); [Mitchell et al. \(1999\)](#); [Davidoff, Brown and Diamond \(2005\)](#); [Lockwood \(2012\)](#)) because it simplifies the estimation in two important ways.<sup>40</sup> First, it allows the individual's optimal savings problem to be solved numerically in a fast and precise way ([Appendix E](#)). Second, this assumption implies that preferences are homothetic so the choice over  $g$  is invariant to initial wealth (which is not observed in the data).

The fact that individual choices depends only on  $(r, \alpha, \beta)$  necessarily assumes

<sup>40</sup>An exception is [Kopczuk and Lupton \(2007\)](#) where bequest motives are linear rather than CRRA. This induces significantly different consumption paths for consumers, where consumption remains everywhere below a “satiation” point.

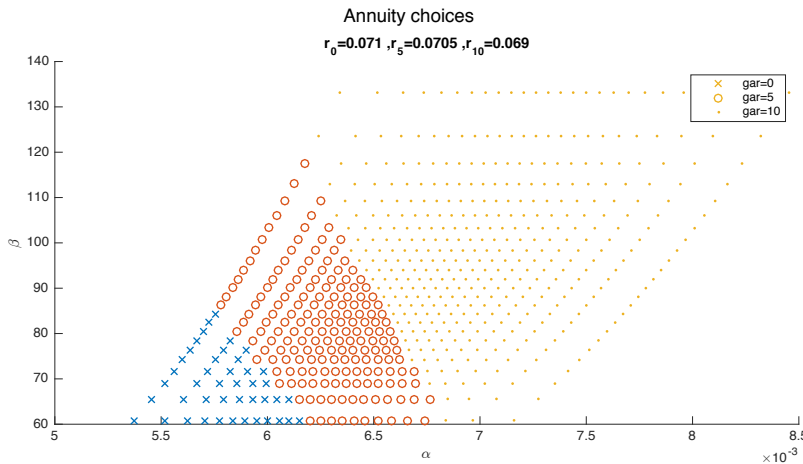


Figure 6: Decisions of consumers, as a function of  $\alpha, \beta$ , for a fixed vector  $r$  corresponding to the average rates among Men 65. Individual with higher mortality  $\alpha$  or higher bequest motives  $\beta$  chose higher guarantee lengths.

away a number of possible alternative explanations. An alternative specification could, for instance, calibrate a homogeneous value of  $\beta$  and estimate the distribution of the CRRA parameter  $\gamma_i$ . However, it seems natural to link the choice of the guarantee  $g$  to bequest motives. Moreover, to the extent that both  $\gamma$  and  $\beta$  are heterogeneous, it seems reasonable to calibrate  $\gamma$  since there is a larger literature estimating this parameter. Alternatively, it is possible that the choice over  $g$  is driven primarily by the composition of each individual's savings portfolio. However, this heterogeneity is not observable in my data.

Figure 6 illustrates the decisions of consumers over  $g$ , as a function of  $\alpha, \beta$ , for a fixed vector  $r$ .

#### 44 Calibrated parameters

There exists a long literature estimating the CRRA parameter  $\gamma$ . The estimated value varies between about 3 and 1. A value of  $\gamma = 3$  is frequently used in simulations of lifetime consumption paths (e.g., [Davis, Kubler and Willen \(2006\)](#); [Scholz, Seshadri and Khitatrakun \(2006\)](#)). However, other articles have estimated  $\gamma$  closer to one ([Laibson et al. \(1998\)](#)) especially among the elderly ([Hurd \(1989\)](#)). Since I focus on an older population, I assume  $\gamma = 2$  in my main specification and report below a sensitivity analysis on this parameter.



I assume that total non-annuitized wealth  $w_1$  is proportional to annuitized wealth. In particular, I follow the survey by Banks et al. (2005) in assuming  $w_1 = 4\phi$ .<sup>41</sup> That is, beyond the annuitized amount  $\phi$ , individuals have additional savings in the amount  $4\phi$ , so that individuals annuitize approximately 20% of their total wealth.

I follow Einav, Finkelstein and Schrimpf (2010) in calibrating  $R$  using the average, over the sample period, of the inflation-indexed zero-coupon 10-year Bank of England bond (approximately 3,13%).<sup>42</sup> Therefore, I calibrate  $R = 1.0313$  and  $\delta = \frac{1}{R}$ .

I use present values throughout to account for the effect of inflation. During the period of the sample, inflation was on average 2,31% per year. It is assumed that individuals acts as if this level of inflation was going to remain constant into the future. Inflation rates have been quite stable in the UK in recent years, and the Bank of England has an explicit mandate to target a level of inflation of 2% per year.

I calibrate  $\lambda = 0.11$  for all consumers. A large literature, summarized in Levy and Levin (2014) has found this value to be approximately  $\lambda = 0.1$ .<sup>43</sup> I use  $\lambda = 0.11$  since this is the value estimated by Einav, Finkelstein and Schrimpf (2010) in the context of UK annuities.

## 45 Heterogeneity

Each individual  $i$  is characterized by the the observable vector

$$\theta_i = (\phi_i, \alpha_i, FA_i, INT_i, Pcode_H, Pcode_M).$$

$\phi_i$  is the size of the individual's pension fund.  $\alpha_i$  is the individual's mortality, calculated using her life expectancy (LE).  $FA_i$  is an indicator for whether the purchased was mediated through a financial advisor.  $INT_i$  ("internal") is an indicator for whether the individual had her pension fund with the insurer prior to purchasing the annuity.  $Pcode_H$  is an indicator for whether the individual belongs to a post-code in the top third of average longevities, and  $Pcode_M$  is the analogous indicator for the middle third (the excluded category are post-codes in the bottom third of longevities).

<sup>41</sup>Banks et al. (2005) report that, for individuals with compulsory annuitization, approximately 20% of income comes from the compulsory annuity.

<sup>42</sup>Information on UK government bond rates is available at <http://www.bankofengland.co.uk/boeapps/iadb/>.

<sup>43</sup>The authors state "[T]he estimate of the Gompertz parameter in most analyses of human mortality data using the pure Gompertz model is about 0.10" (Section 5.1).

The only heterogeneity unobserved to the econometrician is  $\beta_i$ . I assume that  $\beta_i$  follows a log-normal distribution within each gender-age set. I allow the mean of this distribution to depend on observable characteristics, including  $\alpha_i$ . Formally, I assume  $\log(\beta) \sim \mathcal{N}(\bar{\beta}(\theta), \sigma_{\beta}^2)$  with

$$\begin{aligned}\bar{\beta}(\theta) = & \beta_0 + \beta_{\phi} \log(\phi) + \beta_{\alpha} \log(\alpha) + \beta_{FA} FA + \dots \\ & \dots + \beta_{INT} INT + \beta_{PcodeH} Pcode_H + \beta_{PcodeM} Pcode_M.\end{aligned}$$

The assumption of log-normality imposes that the bequest motives  $\beta$  are weakly positive, as is intuitive in this setting. The assumption is common in the literature (EFS).

The parameters being estimated are  $(\sigma_{\beta}^2, \beta_0, \beta_{\alpha}, \beta_{\phi}, \beta_{FA}, \beta_{INT}, \beta_{PcodeH}, \beta_{PcodeM})$ . To make the model maximally flexible, I estimate these parameters for each gender-age subset independently.

## 5 Estimation and Results

### 5.1 Likelihood

For individual  $i$  with type  $\theta_i$ , the bequest motive  $\beta$  is drawn from the log-normal distribution with PDF  $f_{\beta}(\beta | \theta_i, \Theta)$ . The probability of consumer  $i$  choosing contract  $g$  is

$$P_{ig} = \int_{\beta} I\{V_g \geq V_{g'}, \forall g' \neq g\} f_{\beta}(\beta | \theta_i, \Theta) d\beta.$$

I approximate the integral over  $\beta$  using Gaussian quadrature. I use a Logit-smoothed Accept-Reject estimator, as described in [Train \(2009, p.121\)](#). Therefore, I approximate  $P_{ig}$  using

$$P_{ig} \approx \int_{\beta} \frac{\exp(\varsigma V_g)}{\sum_j \exp(\varsigma V_j)} f_{\beta}(\beta | \theta_i, \Theta) d\beta.$$

As  $\varsigma \rightarrow 0$ , the integrand approaches  $\frac{1}{3}$  (each product  $g$  is chosen with equal probability). As  $\varsigma \rightarrow \infty$ , the integrand approaches the indicator function  $I\{V_g \geq V_{g'}, \forall g' \neq g\}$ .

The typical value of  $V_k$  is of the order of  $10^{-3}$ , and I take  $\varsigma = 10^6$ .<sup>44</sup> The contribution of individual  $i$  choosing contract  $g$  to the log-likelihood (LL) is  $\log(P_{ig})$ .

To maximize the log-likelihood, I follow a three-step procedure that attempts to make the maximization robust to multiple starting points and locally maxima. First, I evaluate the LL on a grid of 20,000 drawn from a uniform distribution on a large hyper-rectangle of the parameter space. Second, I use the resulting 200 points with the largest value of the LL as the starting values for a gradient-based maximization of the LL, but restrict the procedure to 150 evaluations of the LL. Then, I use the resulting 50 points with the highest value of the LL as the starting values for a full gradient-based maximization of the LL.

## 52 Identification

The model is identified by the variation in individual choices over the three contract options, and how these correlate with observables such as  $\alpha$ ,  $\phi$ , and so on. Intuitively, similar market shares in all three contracts would indicate large values of  $\sigma_\beta$ . On the other hand, a higher market share in  $g_5$  and  $g_{10}$  would indicate a high value of  $\bar{\beta}$ .

For identification, I also rely on the variation in rates offered to individuals of different fund sizes  $\phi$ . Fund size does not affect the guarantee choice under the assumption of CRRA instantaneous utility, and therefore this variation is exogenous given the models' assumptions. That is, under CRRA, two individuals who differ in  $\phi$  but have the same  $(\alpha, \beta, r)$  would choose the  $g$ . However, individuals do obtain different rates based on  $\phi$ , so the variation in rates is exogenous given the model's assumption. Although this variation is useful, it relies heavily on the model's structure.<sup>45</sup>

---

<sup>44</sup> All resulting choice probabilities are close to zero or unity. The results are robust to different values of  $\varsigma$ .

<sup>45</sup> A better source of price variation would be an exogenous change in contracts over time, assuming the population of individuals retiring in a given month is drawn from the same distribution. Although such variation is present in the data, it cannot plausibly be considered exogenous, because it is likely caused by changes in interest rates, which can also affect demand. Several studies use variation over time in the prices a given individual faces, if prices depend on insurance events (Israel (2004); Abbring et al. (2003)). Instead, I am using variation across consumers, within a cross section.

|                      | Men-65  |         | Women-65 |         | Men-60  |         | Women-60 |         |
|----------------------|---------|---------|----------|---------|---------|---------|----------|---------|
| $\log(\sigma_\beta)$ | -1.111  | (0.006) | -0.569   | (0.011) | -0.655  | (0.007) | -0.857   | (0.019) |
| $\beta_0$            | -10.996 | (0.130) | -6.882   | (0.316) | -17.271 | (0.210) | -6.190   | (0.123) |
| $\beta_\phi$         | 0.702   | (0.005) | 0.021    | (0.015) | 0.706   | (0.007) | 0.181    | (0.018) |
| $\beta_\alpha$       | -2.046  | (0.052) | -2.229   | (0.126) | -3.171  | (0.028) | -1.777   | (0.035) |
| $\beta_{FA}$         | 0.019   | (0.008) | 0.135    | (0.037) | 0.023   | (0.012) | 0.107    | (0.016) |
| $\beta_{INT}$        | 0.031   | (0.011) | 0.591    | (0.038) | 0.011   | (0.013) | 0.063    | (0.014) |
| $\beta_{PcodeM}$     | 0.063   | (0.012) | 0.020    | (0.018) | -0.067  | (0.014) | 0.045    | (0.016) |
| $\beta_{PcodeH}$     | 0.020   | (0.012) | -0.003   | (0.027) | -0.050  | (0.013) | 0.041    | (0.015) |

Table 2: Point estimates and standard errors for the preferred specification, for each age-gender set.

### 53 Estimates

Table 2 contains the estimates corresponding to the preferred specification. The results suggest a significant dispersion exists in the value of  $\beta$  within each subset of consumers. Moreover, the variance of  $\beta$  differs significantly across consumer subsets, which is in accordance with much of the recent literature (Lockwood (2014); Kopczuk and Lupton (2007)). The findings also suggest private information about both costs and preferences are important determinants of choice, as highlighted by Finkelstein and McGarry (2006); Fang, Keane and Silverman (2008); Cohen and Einav (2007).

For all subsets of consumers,  $\beta_\alpha < 0$ , suggesting a negative relationship between bequest motives  $\beta$  and mortality  $\alpha$ . As discussed above, this pattern makes it likely that, relative to a setting with symmetric information, equilibrium choices are skewed toward  $g_0$  and away from  $g_{10}$ . This pattern is the opposite of the one found in EFS, where  $\beta$  and  $\alpha$  were found to have a positive correlation. Figure 7 present a histogram and several moments of the estimated joint distribution of  $(\alpha, \beta)$  for Men 65.

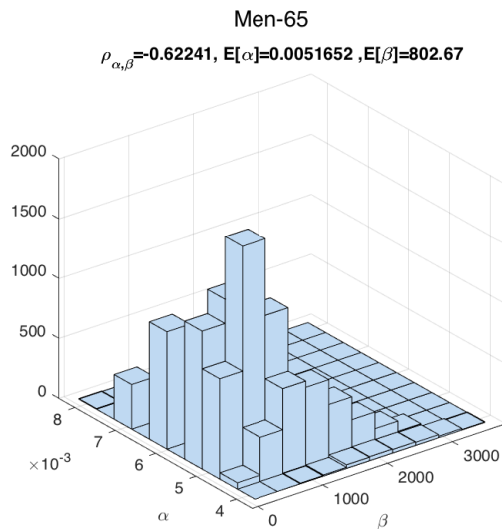


Figure 7: A histogram, and several moments, of the estimated joint distribution of  $(\alpha, \beta)$  for Men 65.  $\rho_{\alpha\beta}$  captures the correlation between  $\alpha$  and  $\beta$ .

## 6 Counterfactuals

In the annuities context, the welfare-maximizing outcome is not obvious.<sup>46</sup> EFS find, empirically, that under symmetric information all individuals would choose  $g_{10}$ , but I find this not to be the case in my setting.

My measure of welfare is the amount that each individual would be willing to pay for her preferred annuity contract, as in the theoretical model of Section 2. I denote this amount by the “wealth equivalent”  $WE_i(r)$ .<sup>47</sup> That is, individual  $i$  is indifferent between obtaining her preferred annuity at the given equilibrium rates  $r$  or obtaining no annuity and increasing her non-annuitized initial wealth  $w_1$  by  $WE_i(r)$ , and then consuming optimally throughout her life. For concreteness I will present changes in  $WE$  as a fraction of annuitized wealth ( $\phi$ ), since the total annuitized wealth in this market is known to be about £12bn (in 2013).

Since the annuities market features mandatory purchase, it is possible that  $WE$  is less than the annuitized wealth  $\phi$ . When this is the case, the individual would be

<sup>46</sup>For instance, in a standard insurance setting without moral hazard, full insurance for all individuals is the welfare-maximizing allocation.

<sup>47</sup>This is also the measure of welfare in EFS.

better off if she was not forced to purchase an annuity. The estimates suggest that this is the case for some individuals, especially those with low longevity (high  $\alpha$ ). Annuities force these individuals to defer consumption into the future more than is optimal.

In all counterfactuals, I assume that retirement age remains exogenous. For instance, men who were retiring at age 60 prior to any CR policy, continue to retire at that age. This assumption seems fairly innocuous. First, CR policy has only a small effect on annuity rates in this setting. Second, it is not always obvious how CR policy would change the optimal retirement age. For instance, if gender-neutral rates were mandated for both 65- and 60-year-olds, women would obtain higher rates whenever they retire. The issue is more of a concern for the discussion of age-based CR.

For simplicity, I begin by considering a simple setting where there is no PD on the basis of fund size  $\phi$ . That is, I discuss CR on the basis of gender and age only, assuming that rates do not also vary with  $\phi$ . I do this only for clarity of exposition. After discussing the effect of CR in this simpler setting, I also estimate the welfare effect of CR when insurers can charge different rates depending on  $\phi$ .

## 61 Computing Equilibria

I begin by computing the competitive equilibrium in each consumer subset (i.e., assuming full PD). A pure-strategies Nash equilibrium is not guaranteed to exist in this setting ([Handel, Hendel and Whinston \(2015\)](#)), so I focus on zero-profit allocations, which [Azevedo and Gottlieb \(2015\)](#) have shown always exists.

I simulate consumer subsets with the joint distributions of  $\alpha, \beta$  estimated above. The profit obtained from individual  $i$  in contract  $g$ , when that contract offers rate  $r_g$  is

$$\pi_{gi}(r_g) = \phi_i - \sum_{t=1}^{t=g} \frac{1}{R^t} \phi_i r_g - \sum_{t=g+1}^{t=T} \frac{1}{R^t} S_{ti} \phi_i r_g.$$

The revenue is  $\phi_i$ . The firm is certain to make payments  $\phi_i r_g$  until the end of the guarantee period ( $t = g$ ). Beyond this time, payments occur if the individual is alive, which happens with probability  $S_{ti}$ .

Let  $\pi_g(r)$ , where  $r = [r_0, r_5, r_{10}]$  be the average profit in contract  $g$ , taking into account that this contract is chosen by those individuals who prefer  $g$  over the other

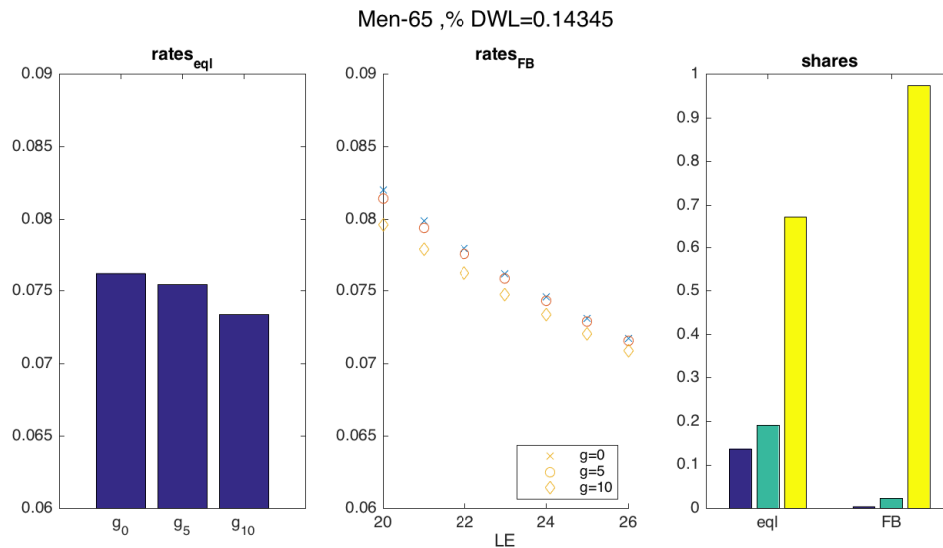


Figure 8: Rates and shares in each contract, at equilibrium and first best, for 65-year-old Men. The equilibrium shows a shift towards  $g_0$  and away from  $g_{10}$ .

options (therefore  $\pi_g(r)$  depends on the entire vector of rates offered in all contracts). The full PD equilibrium rates  $r^* = [r_0^*, r_5^*, r_{10}^*]$  are the rates at which each of the three contracts breaks even ( $\pi_g(r^*) = 0, \forall g$ ). For additional details, see Appendix G.

Figure 8 shows the equilibrium rates, as well as the rates each life expectancy (LE) type would obtain under symmetric information. Individuals with higher LE (lower mortality  $\alpha$ ) obtain lower rates. I also compute, for each individual, the “first-best” (or symmetric information) rates at which the firm would break even given that individual’s mortality  $\alpha$ . The figure shows choices at the equilibrium and at the first-best rates. As discussed above, Men-65 exhibit a strong negative correlation between  $\alpha$  and  $\beta$ . This implies that those individuals choosing  $g_0$  because of their low bequest values  $\beta$  tend to have high mortality  $\alpha$  and therefore are low-cost individuals. These choices result in higher rates  $r_0$  and therefore an equilibrium where choices are distorted towards  $g_0$  and away from  $g_{10}$ , relative to the choices individuals would make if offered first-best rates. This pattern is present also in the other consumer subsets (Appendix G).

The computed equilibrium shows a moderate loss of welfare due to asymmetric information. The equilibrium welfare for 65-year-old men is about 0.14% lower than

the welfare under symmetric information. EFS find this figure to be closer to 2%. For the other consumer subsets, the deadweight loss as a percentage of first-best welfare is 0.15% (Women-65), 0.41% (Men-60) and 0.25% (Women-60). The results suggest that 60-year-olds experience significantly greater share of welfare lost due to asymmetric information, relative to 65-year-olds.

## 62 Computing partial CR policies

For computational simplicity, I consider here a different parameterization of the continuum of policies between full PD and full CR than the one considered in Section 2. I let the vector of rates  $r = [r_0, r_5, r_{10}]$  for one consumer subset (say,  $B$ ) change linearly from its level under full PD ( $r^* = [r_0^*, r_5^*, r_{10}^*]$ ) to its level under full CR ( $\bar{r} = [\bar{r}_0, \bar{r}_5, \bar{r}_{10}]$ ). Then, at several points of this linear path, I compute the rates that must be offered to the other subset ( $A$ ) so that each contract breaks even across both subsets. For additional details, see Appendix G.

## 63 Gender-neutral rates

I begin by considering gender-neutral pricing. That is, I consider a policy that mandates that men and women of a given age must obtain the same rates. Recall that, for now, I am not allowing firms to offer different rates on the basis of  $\phi$ .

The results are shown in Figures 9 and 10. Among 65-year-olds (men and women), a small constraint on prices increases welfare by the equivalent of 0.01% of annuitized wealth. Full CR would result in a significant welfare loss, but this moderate constraint on prices increase welfare.

Such a policy also entails significant redistribution. Under CR, rates decrease for men and increase for women. Welfare (measured by average  $WE$ ) rises significantly for 65-year-old women, but this group is relatively small. 65-year-old men experience a smaller decrease in welfare, although this group is significantly larger.

Among all 60-year-olds, the optimal CR policy involves significantly constraining rates and results in a welfare increase equivalent to 0.09% of annuitized wealth. In this case, welfare increases mildly for the larger subset of women and decreases significantly for the smaller subset of 60-year-old men.



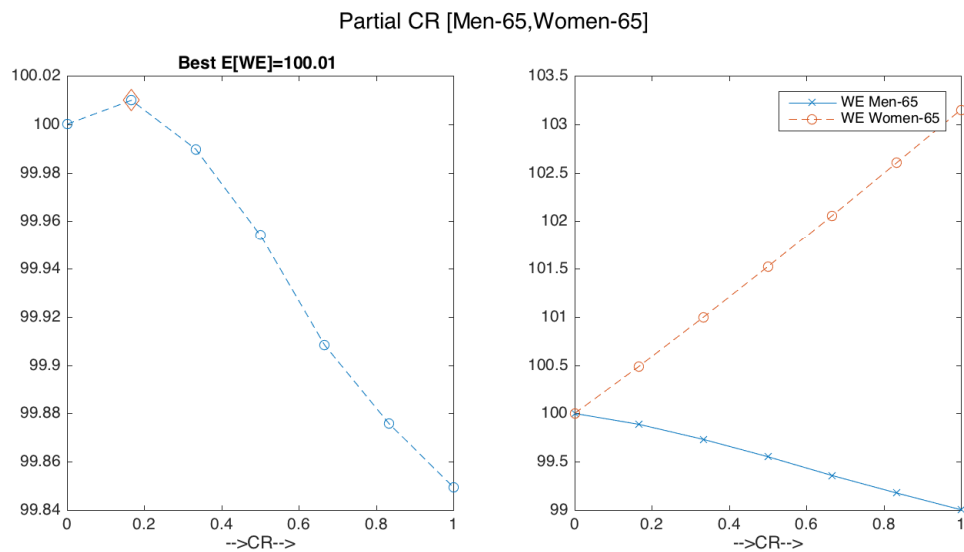


Figure 9: Optimal CR on the basis of gender, among 65-year-olds.

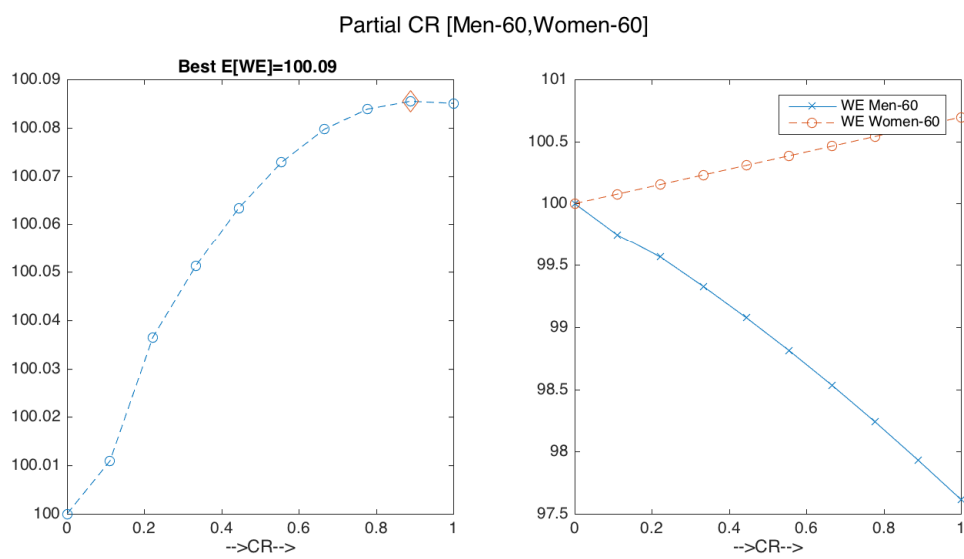


Figure 10: Optimal CR on the basis of gender, among 60-year-olds.

## 64 Age-neutral rates

I now consider a policy that mandates age-neutral pricing. Such a policy would restrict the difference in rates offered to Women-65 and Women-60. In this setting, I continue to assume that firms cannot price discriminate on the basis of  $\phi$ .

Figure 11 captures the effect of CR policy among women. Full CR achieves an increase in welfare corresponding to 0.11% of annuitized wealth. Nonetheless, there is significant redistribution from older women (retiring at age 65) to younger women.

Figure 12 captures the results of CR policy among men. In this case, the optimal policy is full PD. CR causes rates to rise for 60-year-olds and to fall for 65-year-olds, with corresponding changes in welfare and accompanying redistribution.

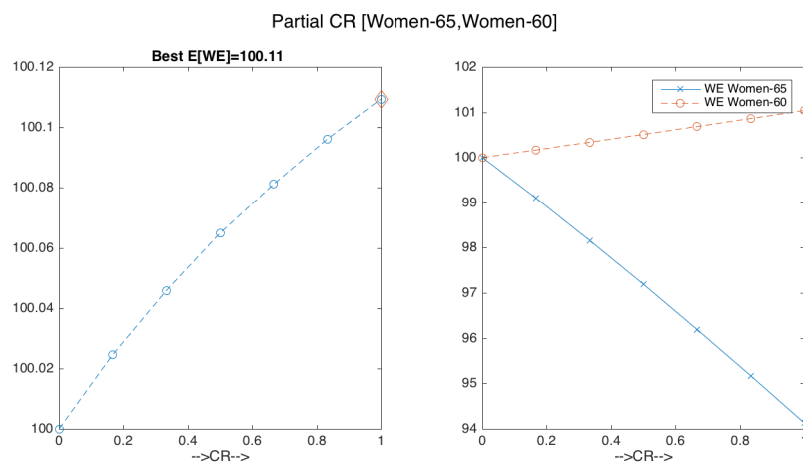


Figure 11: The effect of gender-based CR among Women.

In this setting, the assumption of an exogenous retirement age is more significant. CR on the basis of age always increases rates for younger individuals, which will increase the incentive to retire earlier.

## 65 Allowing PD on fund-size

Using the methodology described above, it is straightforward to consider the effect of CR when firms are allowed to price discriminate on the basis of age and fund size ( $\phi$ ). In this section, I will consider the effect of gender-neutral rates in such a setting. This policy is a more realistic approximation to the policy mandated by the EU in 2012 for gender-neutral insurance prices.

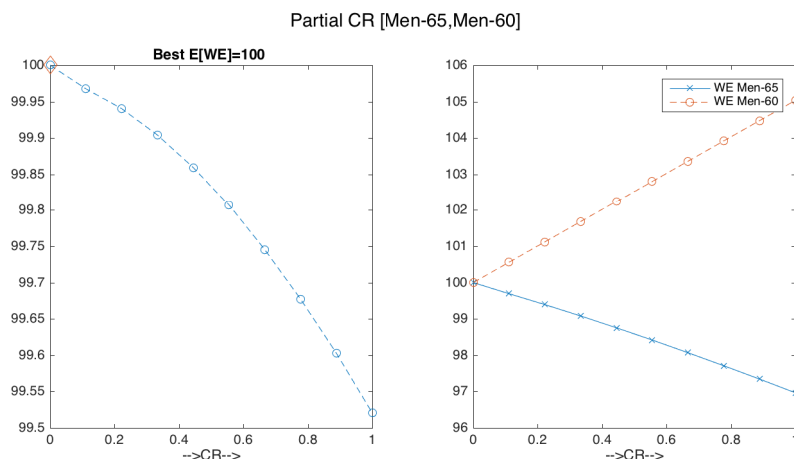


Figure 12: Optimal CR on the basis of age, among men.

I consider, within each age-gender subset, a further partition of consumers into 4 tiers according to fund size  $\phi$ . These partitions were chosen so that each includes approximately 25% of the mass of all individuals. The thresholds of  $\phi$  used to defined these tiers are £9,000, £14,000, £21,000 and £40,000.

Since computing the equilibrium rates is computationally costly, I do not compute the optimal intermediate CR policy in this setting. Instead, my goal is to estimate the welfare effect of a shift from full PD to full CR (gender-neutral prices), as an approximation to the welfare effect of the EU 2012 policy.

Among 60-year-olds (Figure 13), the estimates suggest that full CR increases welfare by the equivalent of about 0.09% of annuitized wealth. The left hand side panel shows that, at each income tier, consumer surplus increases among women and decreases among men. However, the right hand side panel shows that total consumer surplus increases at every income tier among 60-year-olds. This results mirrors that obtained above where, when PD on the basis of  $\phi$  was not allowed. In that case, CR on the basis of gender significantly increased consumer surplus among 60-year-olds. Notice that welfare among 60-year-old men decreased significantly, while welfare among women increased by a smaller amount, but women constitute a larger portion of 60-year-olds

60-year-olds are approximately 60% of my sample. As a rough approximation, assuming that these welfare gains would apply to 60% of the annuities market as a whole (£12bn), the total welfare gain is approximately £6.5 million.

Among 65-year-olds (Figure 14), the effect of full CR on welfare is negligible. Overall, the estimates suggest that welfare increased by less than the equivalent of 0.01% of annuitized wealth. Moreover, among 65-year-olds, the incidence of the policy was more heterogeneous. Individuals with lower pension funds were harmed by the move from full PD to full CR, while wealthier individuals benefited from the policy.

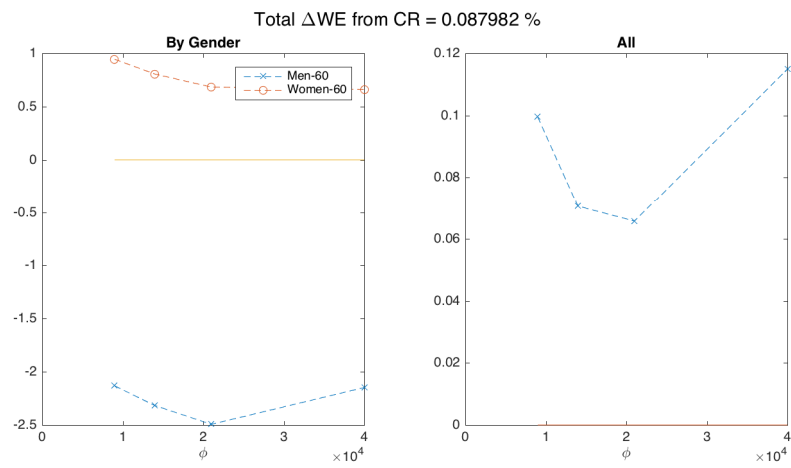


Figure 13: The effect of gender-based CR among 60-year olds, when insurers can price-discriminate by tiers of  $\phi$ . Each graph shows the change in average  $WE$  resulting from the change from full PD to full CR, for each tier of fund size  $\phi$ .

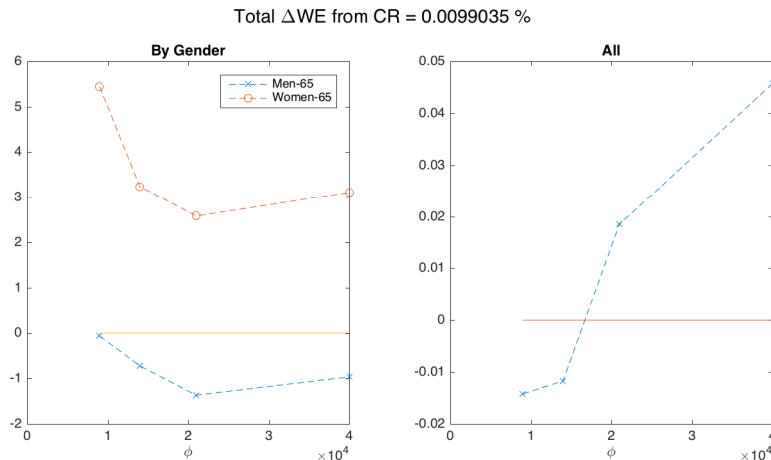


Figure 14: The effect of gender-based CR among 65-year olds, when insurers can price-discriminate by tiers of  $\phi$ . Each graph shows the change in average  $WE$  resulting from the change from full PD to full CR, for each tier of fund size  $\phi$ .

## 66 Calibration to US health insurance

Appendix [H](#) contains a calibration of the benefit of CR in the context of US health insurance. In that simpler setting, the assumptions of Constant Absolute risk aversion and Gaussian wealth shocks imply linear certainty equivalents, so the theoretical model of Section [2](#) applies literally (utility is quasi-linear). The model is calibrated to fit moments of the distribution of risk and risk aversion estimation by [Handel, Hendel and Whinston \(2015\)](#). The calibration suggests that optimal CR policy can achieve significant reductions in the deadweight loss due to adverse selection, on the order of 2%-7%.

## 7 Conclusion

This paper has investigated the effect on welfare of third-degree price discrimination in competitive markets with adverse or advantageous selection. Perfect (first-degree) price discrimination achieves the first-best outcome. However, away from this limit, third-degree price discrimination can increase or decrease overall welfare. I have characterize the optimal degree of price discrimination on the basis of a given signal. I have shown that when high-cost consumer subsets are also those

who experience the greater adverse selection distortions, community rating can increase overall welfare. I have illustrated the potential benefit of CR policy in the context of UK annuities. I have found that optimal CR policy can achieve increases in welfare equivalent of 0.09% of total annuitized wealth or £6.5 million per year.

CR policy is one of many mechanisms that can reduce the deadweight loss of adverse selection. This mechanism is already used in practice even though its welfare implications are poorly understood. However, CR is not a panacea. In fact, CR need not be the best way to reduce adverse selection distortions in a given context. However, it seems politically palatable and easy to implement since it does not require additional tax revenue, unlike other policies like subsidies. CR can therefore easily be combined with subsidies or taxes to achieve a greater reduction in DWL. For instance, if the cost of public funds is convex in the amount of funds raised, optimal CR policy can achieve a reduction in DWL at a significantly lower cost than the use of subsidies alone.

Several important avenues remain for future work. First, I have considered only competition in prices. It would be interesting to explore the effect of CR when firms use menus of products to screen buyers, as in [Rothschild and Stiglitz \(1976\)](#). Second, it would be important to determine the extent to which market power affects interacts with CR policy. Third, it would be interesting to extend the model to accommodate the presence of multiple signals. Such a model would show how the optimal contractibility of one signal (e.g., age) affects the optimal contractibility of another (e.g., gender). Fourth, I have assumed away in my analysis the firm's incentives to collect additional information about consumers if CR is mandated. Analyzing that kind of response by firms would allow for a better understanding of how firms might react to CR.

## References

- Abbring, Jaap H, James J Heckman, Pierre-André Chiappori, and Jean Pinquet.** 2003. "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?" *Journal of the European Economic Association*, 1(2-3): 512–521.
- Aguirre, Inaki, Simon Cowan, and John Vickers.** 2010. "Monopoly price discrimination and demand curvature." *The American Economic Review*, 100(4): 1601–1615.
- Akerlof, George A.** 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3): 488–500.
- Andreoni, James.** 1989. "Giving with impure altruism: Applications to charity and Ricardian equivalence." *The Journal of Political Economy*, 1447–1458.
- Arrow, Kenneth Joseph.** 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.
- Azevedo, Eduardo M, and Daniel Gottlieb.** 2015. "Perfect competition in markets with adverse selection." *The Wharton School Research Paper*, , (70).
- Bagnoli, Mark, and Ted Bergstrom.** 2005. "Log-concave probability and its applications." *Economic theory*, 26(2): 445–469.
- Banks, James, and Carl Emmerson.** 1999. "UK ANNUITANTS." *Institute for Fiscal Studies*.
- Banks, James, Carl Emmerson, Zoë Oldfield, and Gemma Tetlow.** 2005. "Prepared for retirement? The adequacy and distribution of retirement resources in England."
- Bergemann, Dirk, Benjamin Brooks, and Stephen Morris.** 2015. "The Limits of Price Discrimination." *The American Economic Review*, 105(3): 921–957.
- Blumberg, Linda J, and Matthew Buettgens.** 2013. "Why the ACA's Limits on Age-Rating Will Not Cause "Rate Shock": Distributional Implications of Limited Age Bands in Nongroup Health Insurance." *Urban Institute Policy Brief* <http://www.urban.org/UploadedPDF/412757-Why-the-ACAs-Limits-on-Age-Rating-Will-Not-Cause-Rate-Shock.pdf>. Published March, 4.

- Bommier, Antoine.** 2006. "Uncertain lifetime and intertemporal choice: Risk aversion as a rationale for time discounting\*." *International Economic Review*, 47(4): 1223–1246.
- Braun, Michael, and Alexander Muermann.** 2004. "The impact of regret on the demand for insurance." *Journal of Risk and Insurance*, 71(4): 737–767.
- Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston.** 2014. "How does risk selection respond to risk adjustment? New evidence from the Medicare Advantage Program." *The American Economic Review*, 104(10): 3335–3364.
- Chen, Yongmin, and Marius Schwartz.** 2013. "Differential Pricing When Costs Differ: A Welfare Analysis."
- Chen, Yongmin, and Marius Schwartz.** 2015. "Differential pricing when costs differ: a welfare analysis." *The RAND Journal of Economics*, 46(2): 442–460.
- Cohen, Alma, and Liran Einav.** 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review*, 97(3): 745–788.
- Crocker, Keith J, and Arthur Snow.** 1986. "The Efficiency Effects of Categorical Discrimination in the Insurance Industry." *The Journal of Political Economy*, 321–344.
- Davidoff, Thomas, Jeffrey R Brown, and Peter A Diamond.** 2005. "Annuities and Individual Welfare." *American Economic Review*, 95(5): 1573–1590.
- Davis, Steven J, Felix Kubler, and Paul Willen.** 2006. "Borrowing costs and the demand for equity over the life cycle." *The Review of Economics and Statistics*, 88(2): 348–362.
- Einav, Liran, Amy Finkelstein, and Mark R. Cullen.** 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *Quarterly Journal of Economics*, 125(3): 877–921.
- Einav, Liran, Amy Finkelstein, and Paul Schrimpf.** 2010. "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the uk annuity market." *Econometrica*, 78(3): 1031–1092.



- Epstein, Larry G, and Stanley E Zin.** 1989. "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica: Journal of the Econometric Society*, 937–969.
- Erdem, Tülin, Michael P Keane, and Baohong Sun.** 1998. "Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters." *Journal of Econometrics*, 89(1): 177–196.
- Ericson, Keith M Marzilli, and Amanda Starc.** 2015. "Pricing regulation and imperfect competition on the massachusetts health insurance exchange." *Review of Economics and Statistics*, 97(3): 667–682.
- Fang, Hanming, Michael P Keane, and Dan Silverman.** 2008. "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy*, 116(2).
- Finkelstein, Amy, and James Poterba.** 2002. "Selection effects in the United Kingdom individual annuities market." *The Economic Journal*, 112(476): 28–50.
- Finkelstein, Amy, and James Poterba.** 2004. "Adverse selection in insurance markets: Policyholder evidence from the UK annuity market." *Journal of Political Economy*, 112(1): 183–208.
- Finkelstein, Amy, and Kathleen McGarry.** 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review*, 96(4): 938–958.
- Finkelstein, Amy, James Poterba, and Casey Rothschild.** 2009. "Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities." *Journal of financial economics*, 91(1): 38–58.
- Geruso, Michael.** 2016. "Demand Heterogeneity in Insurance Markets: Implications for Equity and Efficiency." National Bureau of Economic Research.
- Handel, Ben, Igal Hendel, and Michael D Whinston.** 2015. "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk." *Econometrica*, 83(4): 1261–1313.
- Handel, Benjamin R.** 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682.

- Hendren, Nathaniel.** 2013. "Private information and insurance rejections." *Econometrica*, 81(5): 1713–1762.
- Hirshleifer, Jack.** 1971. "The private and social value of information and the reward to inventive activity." *The American Economic Review*, 561–574.
- Horiuchi, Shiro, and Ansley J Coale.** 1982. "A simple equation for estimating the expectation of life at old ages." *Population Studies*, 36(2): 317–326.
- Hoy, Michael.** 1982. "Categorizing risks in the insurance industry." *The Quarterly Journal of Economics*, 321–336.
- Hurd, Michael D.** 1989. "Mortality risk and bequests." *Econometrica: Journal of the econometric society*, 779–813.
- Israel, Mark.** 2004. "Do we drive more safely when accidents are more expensive? Identifying moral hazard from experience rating schemes."
- Koch, Thomas G.** 2014. "One pool to insure them all? Age, risk and the price (s) of medical insurance." *International Journal of Industrial Organization*, 35: 1–11.
- Kopczuk, Wojciech, and Joseph P Lupton.** 2007. "To leave or not to leave: The distribution of bequest motives." *The Review of Economic Studies*, 74(1): 207–235.
- Kotlikoff, Laurence J, and Avia Spivak.** 1981. "The Family as an Incomplete Annuities Market." *The Journal of Political Economy*, 372–391.
- Laibson, David I, Andrea Repetto, Jeremy Tobacman, Robert E Hall, William G Gale, and George A Akerlof.** 1998. "Self-control and saving for retirement." *Brookings papers on economic activity*, 1998(1): 91–196.
- Levin, Jonathan.** 2001. "Information and the Market for Lemons." *RAND Journal of Economics*, 32(4): 657–666.
- Levy, Gilberto, and Bruce Levin.** 2014. *The Biostatistics of Aging: From Gompertzian Mortality to an Index of Aging-relatedness*. John Wiley & Sons.
- Lockwood, Lee M.** 2012. "Bequest motives and the annuity puzzle." *Review of economic dynamics*, 15(2): 226–243.
- Lockwood, Lee M.** 2014. "Incidental bequests: Bequest motives and the choice to self-insure late-life risks." National Bureau of Economic Research.

- Mahoney, Neale, and E Glen Weyl.** 2014. "Imperfect competition in selection markets." National Bureau of Economic Research.
- Mitchell, Olivia S, James M Poterba, Mark J Warshawsky, and Jeffrey R Brown.** 1999. "New Evidence on the Money's Worth of Individual Annuities." *The American Economic Review*, 89(5): 1299–1318.
- Murthi, Mamta, J Michael Orszag, and Peter R Orszag.** 1999. "Administrative costs under a decentralized approach to individual accounts: lessons from the United Kingdom." Citeseer.
- Murthi, Mamta, J Michael Orszag, and Peter R Orszag.** 2000. "The value for money of annuities in the UK: Theory, experience and policy." *Journal of Pensions Management*, May.
- Orsini, Joe, and Pietro Tebaldi.** 2015. "Regulated Age-Based Pricing in Health Insurance Exchanges." *Available at SSRN 2464725*.
- Pauly, Mark V.** 1970. "The welfare economics of community rating." *Journal of Risk and Insurance*, 407–418.
- Philipson, Tomas J, and Gary S Becker.** 1998. "Old-Age Longevity and Mortality-Contingent Claims." *Journal of Political Economy*, 106(3): 551–573.
- Riley, John G.** 1979. "Informational Equilibrium." *Econometrica*, 47(2): 331–359.
- Rothschild, Michael, and Joseph E. Stiglitz.** 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*, 90(4): 629–649.
- Schmalensee, Richard.** 1981. "Output and welfare implications of monopolistic third-degree price discrimination." *The American Economic Review*, 242–247.
- Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun.** 2006. "Are Americans Saving "Optimally" for Retirement?" *Journal of Political Economy*, 114(4).
- Tabeau, Ewa, Anneke van den Berg Jeths, and Christopher Heathcote.** 2001. *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective*. Vol. 9, Springer Science & Business Media.
- Train, Kenneth E.** 2009. *Discrete choice methods with simulation*. Cambridge university press.

- Veiga, André, and E Glen Weyl.** 2016. "Product Design in Selection Markets." *The Quarterly Journal of Economics*, 131(2): 1007–1056.
- Weyl, E, and André Veiga.** 2016. "Pricing Institutions and the Welfare Cost of Adverse Selection." *American Economic Journal: Microeconomics*.
- Yaari, Menahem E.** 1965. "Uncertain lifetime, life insurance, and the theory of the consumer." *The Review of Economic Studies*, 32(2): 137–150.

## Appendix

### A Proofs: Baseline Model

#### A1 Concavity of the profit function

Profit is  $\pi = \int_p^{\bar{u}} (p - c(u)) f(u) du$ . The slope of profit is  $\pi' = Q'(p - c) + Q$ . I require

$$\pi'' = Q''(p - c) + 2Q' - Q'c' = Q' \left[ \frac{Q''}{Q'}(p - c) + 2 - c' \right] < 0.$$

Sufficient conditions are concave direct demand ( $Q'' < 0$ ) and advantageous or mild adverse selection ( $2 - c' > 0$ ). [Aguirre, Cowan and Vickers \(2010\)](#) contains an additional discussion of the concavity of industry profit when marginal cost is constant.

#### A2 Uniqueness of Equilibrium

To show uniqueness of equilibrium, I show that there is a unique fixed point of  $AC(p)$ . I assume that  $c$  is monotonic,  $c' < 1$  and that the distribution of  $c$  is log-concave. If  $c' < 0$ , then  $AC' < 0$ . Since  $AC(0) > 0$ , the function  $AC(p)$  has a unique fixed point.

Now suppose  $c' > 0$ , but  $c' < 1$  and that the distribution of  $c$  is log-concave. Let  $c_p = c(p)$ . If  $G(c)$  log-concave, then  $AC - c = \mathbb{E}[c | c > c_p] - c_p$  is decreasing in  $c_p$ , from [Bagnoli and Bergstrom \(2005\)](#). Therefore,  $\frac{dAC}{dc_p} = \frac{dAC}{dp} \frac{1}{dc_p/dp} < 1 = \frac{dc_p}{dc_p}$  and, since  $\frac{dc_p}{dp} = c' < 1$ ,  $AC(\cdot)$  is contraction and therefore has a unique fixed point.<sup>48</sup>

#### A3 Concavity of welfare

Recall  $\frac{dp_A}{d\chi} < 0$  while  $\frac{dp_B}{d\chi} > 0$ . From the main text,

$$\frac{dW}{d\chi} = \bar{\pi} \left( \frac{Q_A}{\pi'_A} - \frac{Q_B}{\pi'_B} \right).$$

Further differentiation yields

$$\frac{d^2W}{d\chi^2} = \bar{\pi} \left( \frac{Q_A}{\pi'_A} \right)' \underbrace{\frac{dp_A}{d\chi}}_{-} - \bar{\pi} \left( \frac{Q_B}{\pi'_B} \right)' \underbrace{\frac{dp_B}{d\chi}}_{+}.$$

<sup>48</sup>The condition  $A' < 1$  is similar to the condition given for stability of equilibrium in [Mahoney and Weyl \(2014\)](#). In that paper, inverse demand ( $P(q)$ ) and average cost ( $AC(q)$ ) are expressed as function of quantity  $q$ , and stability requires  $P'(q) < AC'(q)$ . In my setting  $\frac{1}{dQ/dp} < \frac{A'}{dQ/dp}$ .

Therefore, a sufficient condition for concavity is  $\frac{d}{dp_m} \left( \frac{Q_m}{\pi'_m} \right) > 0$  or, equivalently,  $\frac{\pi'_m}{Q_m}$  decreasing in  $p_m$ . Omitting subscripts, I require

$$\frac{\pi'}{Q} = -\sigma(p - c) + 1 \Rightarrow \left[ \frac{\pi'}{Q} \right]' = -\sigma'(p - c) - \sigma(1 - c') < 0.$$

If  $Q$  is log-concave, then  $\sigma' > 0$ . Moreover,  $\sigma > 0$ . If  $p - c > 0$  and  $c' < 1$ , then  $\left[ \frac{\pi'}{Q} \right]' < 0$  as required.

**Proof of Proposition 1.** Full PD

I have established  $\frac{dW}{d\chi} = \bar{\pi} \left( \frac{Q_A}{\pi'_A} - \frac{Q_B}{\pi'_B} \right)$ . I require  $W'(0) < 0 \Leftrightarrow \frac{Q_A}{\pi'_A} < \frac{Q_B}{\pi'_B}$ . At  $\chi = 0$ , prices are determined by  $p_m^* = AC_m^*$ . At such a point, the slope of profit is  $\pi'(p) = Q'(p - AC) + Q(1 - AC') = Q(1 - AC')$ . Therefore, I require

$$\frac{Q_A}{Q_A(1 - AC'_A)} < \frac{Q_B}{Q_B(1 - AC'_B)} \Leftrightarrow AC'_A < AC'_B$$

where all quantities are evaluated at the full PD prices,  $p_m^*$ .

□

**Proof of Proposition 3.** Full CR

I require  $W'(1) > 0 \Leftrightarrow \frac{Q_A}{\pi'_A} > \frac{Q_B}{\pi'_B}$ , where all quantities are evaluated at the full CR price  $\bar{p}$ . Omitting subscripts, notice

$$\frac{\pi'}{Q} = \frac{Q'(p - AC) + Q(1 - AC')}{Q} = -\sigma(p - AC) - AC' + 1 > 0$$

Therefore, the required condition is

$$-\sigma_A(\bar{p} - AC_A) - AC'_A + 1 < -\sigma_B(\bar{p} - AC_B) - AC'_B + 1$$

Then, using  $\bar{p} = \frac{Q_A AC_A + Q_B AC_B}{Q_A + Q_B}$  yields

$$0 < \frac{\sigma_B \frac{1}{Q_B} + \sigma_A \frac{1}{Q_A}}{\frac{1}{Q_B} + \frac{1}{Q_A}} (AC_A - AC_B) < AC'_A - AC'_B.$$

□

**Proof of Proposition 2.** Optimal CR

Recall (omitting subscripts) that

$$\frac{\pi'}{Q} = \frac{Q'(p - c) + Q}{Q} = -\sigma(p - c) + 1.$$

Therefore, the optimal interior CR policy satisfies

$$\frac{dW}{d\chi} = \bar{\pi} \left( \frac{Q_A}{\pi'_A} - \frac{Q_B}{\pi'_B} \right) = 0 \Rightarrow \sigma_A (p_A - c_A) = \sigma_B (p_B - c_B).$$

This is the unique interior optimal policy since  $W$  is strictly concave in  $\chi$ .

□

#### A4 Effect of semi-elasticities on optimal CR policy

Suppose that demands have constant semi-elasticities  $\sigma_A$  and  $\sigma_B$ . The optimal policy  $\tilde{\chi}$  satisfies

$$\sigma_A (p_A (\tilde{\chi}) - c_A (p_A (\tilde{\chi}))) - \sigma_B (p_B (\tilde{\chi}) - c_B (p_B (\tilde{\chi}))) = 0.$$

This yields

$$\begin{aligned} \frac{d\tilde{\chi}}{d\sigma_A} &= - \frac{p_A - c_A}{\sigma_A (1 - c'_A) \frac{dp_A}{d\chi} - \sigma_B (1 - c'_B) \frac{dp_B}{d\chi}} > 0. \\ \frac{d\tilde{\chi}}{d\sigma_B} &= - \frac{-(p_B - c_B)}{\sigma_A (1 - c'_A) \frac{dp_A}{d\chi} - \sigma_B (1 - c'_B) \frac{dp_B}{d\chi}} < 0. \end{aligned}$$

I have assumed  $1 - c'_m > 0$  for all markets. The denominator is negative since  $\frac{dp_A}{d\chi} < 0$  and  $\frac{dp_B}{d\chi} > 0$ . Therefore  $\frac{d\tilde{\chi}}{d\sigma_A}$  has the sign of  $p_A - c_A > 0$ .

When demand becomes more elastic in the high-cost market (where price decreases), the optimal CR policy increases. If demand becomes more elastic in the low-cost market (where price increases), then a smaller amount of CR becomes optimal.

## B Proofs: Multiple Consumer sets

### B1 Exposition

Suppose multiple consumers sets  $m \in \{1, 2, 3, \dots, M\}$  exist. Full PD prices are  $p_m^*$  such that  $\pi_m(p_m^*) = 0$ . The full CR price  $\bar{p}$  satisfies  $\sum \pi_m(\bar{p}) = 0$ . Let the set  $\mathcal{A}$  be the subset of high-cost markets, so that  $m \in \mathcal{A} \Rightarrow \pi_m^- < 0$ . Let  $\mathcal{B}$  be the subset of low-cost markets ( $m \in \mathcal{B} \Rightarrow \pi_m^- > 0$ ).

Consider the CR policy  $\chi \in [0, 1]$  such that for each market,  $\pi_m(p_m(\chi)) = \chi \pi_m^-$ .<sup>49</sup> Prices  $p_m$  follow paths prescribed by  $\frac{dp_m}{d\chi} = \frac{\pi_m(\bar{p})}{\pi'_m(p_m)}$ . The denominator  $\pi'_m(p_m) > 0$  by assumption. The numerator is positive for low-cost markets ( $m \in \mathcal{B}$ ) and negative for high-cost markets ( $m \in \mathcal{A}$ ).

Welfare as a function of the PD policy  $\chi$  is  $W(\chi) = \sum_m W_m(p_m(\chi))$ , with slope

<sup>49</sup>As above, full PD corresponds to  $\chi = 0$ , full CR corresponds to  $\chi = 1$ , and  $\sum \pi_m(p_m(\chi)) = 0$  for all  $\chi$ .

$$\frac{dW}{d\chi} = \sum_m -Q_m \frac{\pi_m(\bar{p})}{\pi'_m(p_m)} = \sum_m \frac{\pi_m(\bar{p})}{\sigma_m(p_m - AC_m) - AC'_m + 1}.$$

Welfare is concave if  $\chi$  if the assumptions made above apply to all markets  $m$ . I can therefore obtain the following results.

**Proposition 6 (Full PD (Multiple Markets) ).** *For  $M > 2$ , full PD is the welfare-maximizing policy ( $\tilde{\chi} = 0$ ) if*

$$\sum_{m \in \mathcal{B}} \frac{\pi_m(\bar{p})}{AC'_m - 1} < \sum_{m \in \mathcal{A}} \frac{-\pi_m(\bar{p})}{AC'_m - 1}.$$

*A sufficient condition is*

$$\min_{m \in \mathcal{B}} \{AC'_m(p_m^*)\} > \max_{m \in \mathcal{A}} \{AC'_m(p_m^*)\}.$$

First, notice that when  $M = 2$ ,  $\bar{\pi}_A = -\bar{\pi}_B$ , so I recover Proposition 1. Full PD is optimal if all low-cost markets ( $m \in \mathcal{B}$ ) have more significant adverse selection than any high-cost market ( $m \in \mathcal{A}$ ) at the full PD prices. The condition is more restrictive than the one described by Proposition 1 because in a setting with multiple consumer sets, at least one set is likely to exhibit a pattern of selection such that some amount of CR can be beneficial.

**Proposition 7 (Full CR (Multiple Markets) ).** *For  $M > 2$ , full CR is the welfare maximizing policy ( $\tilde{\chi} = 1$ ) if*

$$\sum_{m \in \mathcal{B}} \frac{\pi_m(\bar{p})}{AC'_m + \sigma_m \underbrace{\frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n}}_{+} - 1} > \sum_{m \in \mathcal{A}} \frac{-\pi_m(\bar{p})}{AC'_m + \sigma_m \underbrace{\frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n}}_{-} - 1}.$$

*A sufficient condition is*

$$\max_{m \in \mathcal{B}} \{AC'_m(p_m^*)\} < \min_{m \in \mathcal{A}} \{AC'_m(p_m^*)\}$$

*and that all  $\bar{AC}_m$  are sufficiently similar.*

Full CR is optimal when every high-cost market has higher adverse selection ( $AC'_m$ ) than any low-cost market. Moreover, all costs should be relatively similar at  $\bar{p}$ . Among  $m \in \mathcal{B}$ , the term  $\frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n} > 0$  is positive; therefore, the inequality is most likely to be satisfied when, for  $m \in \mathcal{B}$ ,  $AC_n - AC_m \approx 0$ . Similarly, the term is negative for  $m \in \mathcal{A}$ . Therefore, the condition is most likely to be satisfied when each of these terms is close to zero, that is, when the levels of cost in all markets are sufficiently similar.



Again, the intuition is that if the levels of cost differ greatly, full CR will dramatically change prices. Then the final unit of CR is likely to impose a large welfare loss on the low-cost markets  $\mathcal{B}$  and a small welfare gain on the high-cost markets  $\mathcal{A}$ .

**Proposition 8 (Optimal CR (Multiple Markets) ).** *For  $M > 2$ , the optimal CR policy  $\tilde{\chi}$  satisfies*

$$\frac{dW}{d\chi} = 0 \Rightarrow \sum_m \frac{\pi_m(\bar{p})}{\sigma_m(p_m - c_m) - 1} = 0.$$

Again, the intuition for the result is that the optimal level of CR equates the marginal welfare gains in high-cost markets  $\mathcal{A}$  to the marginal welfare losses in low-cost markets  $\mathcal{B}$ .

## B2 Slope of welfare

Omitting subscripts,

$$\pi' = Q'(p - c) + Q = -Q[\sigma(p - c) - 1]$$

$$\pi' = Q'(p - AC) + Q(1 - AC') = -Q[\sigma(p - AC) - 1 + AC'].$$

The slope of welfare is

$$\frac{dW}{d\chi} = \sum_m \pi_m(\bar{p}) \frac{-Q_m}{\pi'_m(p_m)} = \sum_m \frac{\pi_m(\bar{p})}{\sigma_m(p_m - AC_m) - 1 + AC'_m}.$$

## B3 Proofs

**Proof of Proposition 6.** Full PD (Multiple Markets)

At  $\chi = 0$  (full PD),  $p_m - AC_m = 0$ . I require that, at full PD,

$$\frac{dW}{d\chi} = \sum_m \frac{\pi_m(\bar{p})}{AC'_m - 1} < 0.$$

$$\sum_{m \in \mathcal{B}} \frac{\pi_m(\bar{p})}{AC'_m - 1} < \sum_{m \in \mathcal{A}} \frac{-\pi_m(\bar{p})}{AC'_m - 1}.$$

Notice that all numerators are positive, since  $m \in \mathcal{A} \Rightarrow \pi_m(\bar{p}) < 0$ , while  $m \in \mathcal{B} \Rightarrow \pi_m(\bar{p}) > 0$ .

Moreover, the sum of the coefficients on both sides of the inequality sign are equal since  $\sum \pi_m(\bar{p}) = 0$ .

Therefore, a sufficient condition for the inequality to be satisfied is

$$\min_{m \in \mathcal{B}} \{AC'_m(p_m^*)\} > \max_{m \in \mathcal{A}} \{AC'_m(p_m^*)\}.$$

□

**Proof of Proposition 7.** Full CR (Multiple Markets)

We require that, at  $\chi = 1$  (full CR)

$$\frac{dW}{d\chi} = \sum_m \frac{\pi_m(\bar{p})}{\sigma_m(p_m - AC_m) - 1 + AC'_m} > 0.$$

At full CR  $\bar{p} = \frac{\sum_n Q_n AC_n}{\sum_n Q_n}$ . Therefore, the term  $p_m - AC_m$  in the denominator becomes

$$\bar{p} - AC_m = \frac{\sum_n Q_n (AC_n - AC_m)}{\sum_n Q_n} = \frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n}.$$

By definition, high cost markets ( $\mathcal{A}$ ) have  $\pi_m(\bar{p}) < 0$  and therefore have  $\bar{p} - AC_m < 0$ . Conversely, low-cost markets ( $\mathcal{B}$ ) have  $\bar{p} - AC_m > 0$ .

Therefore the condition becomes

$$\sum_{m \in \mathcal{B}} \frac{\pi_m(\bar{p})}{\underbrace{AC'_m + \sigma_m \frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n}}_{+} - 1} > \sum_{m \in \mathcal{A}} \frac{-\pi_m(\bar{p})}{\underbrace{AC'_m + \sigma_m \frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n}}_{-} - 1}.$$

Notice that all numerators are positive, since  $m \in \mathcal{A} \Rightarrow \pi_m(\bar{p}) < 0$ , while  $m \in \mathcal{B} \Rightarrow \pi_m(\bar{p}) > 0$ . The denominator is always positive since it has the sign of  $\pi'_m$ .

A sufficient condition is

$$\max_{m \in \mathcal{B}} \left\{ AC'_m + \sigma_m \frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n} \right\} < \min_{m \in \mathcal{A}} \left\{ AC'_m + \sigma_m \frac{\sum_{n \neq m} Q_n (AC_n - AC_m)}{\sum_n Q_n} \right\}.$$

□

**Proof of Proposition 8.** Optimal CR (Multiple Markets)

Recall that (omitting subscripts)  $\frac{\pi'}{-Q} = \sigma(p - c) - 1$ . The optimal CR policy prescribes

$$\frac{dW}{d\chi} = 0 \Rightarrow \sum_m \frac{\pi_m(\bar{p})}{\sigma_m(p_m - c_m) - 1} = 0$$

□

## C Proofs: Two Products

**Proposition 9** (Full CR (2)). *With two products, full CR is optimal ( $\tilde{\chi} = 1$ ) if  $Q_B < Q_A$  and*

$$0 < (AC_{HA} - AC_{HB}) \frac{\sigma_{HB} \frac{1}{Q_{HB}} + \sigma_{HA} \frac{1}{Q_{HA}}}{\frac{1}{Q_{HB}} + \frac{1}{Q_{HA}}} + (AC_{LA} - AC_{LB}) \frac{\sigma_{LB} \frac{1}{Q_{LB}} + \sigma_{LA} \frac{1}{Q_{LA}}}{\frac{1}{Q_{LA}} + \frac{1}{Q_{LB}}} < \Delta AC'_A - \Delta AC'_B.$$

For full CR to be the welfare-maximizing policy, the high-cost set must experience greater adverse selection than the low-cost set, as measured by  $\Delta AC'_A - \Delta AC'_B$ . Moreover, the levels of prices cannot be too different across markets in both products ( $AC_{HA} - AC_{HB} > 0$  small and  $AC_{LA} - AC_{LB} > 0$  small). As before, when price differences are large, full CR will result in large price changes relative to PD. When demands are elastic, the last unit of CR is likely to result in a large welfare loss in market  $B$  and a small welfare gain in market  $A$ , so the price difference is weighted by the demand elasticities for the relevant product.

Moreover, Proposition 9 requires that the high-cost set has higher demand for  $H$  at the CR prices  $\bar{p}_H, \bar{p}_L$  ( $Q_A > Q_B$ ). CR will increase prices in market  $B$ . However, if individuals in market  $B$  have little demand for  $H$ , this increase in prices will not significantly reduce welfare, and therefore the cost of implementing CR is lower.

The following corollary uses the fact that  $\Delta AC'(\Delta p) = \sigma_H (AC_H - c_H) + \sigma_L (AC_L - c_L)$ .

**Corollary 1.** *Suppose demands are proportional, as in [Chen and Schwartz \(2013\)](#), ( $Q_A(\Delta p) = zQ_B(\Delta p)$  for  $z > 0$ ). Then  $\tilde{\chi} = 1$  if*

$$\sigma_H c_{HB} + \sigma_L c_{LB} > \sigma_H c_{HA} + \sigma_L c_{LA}.$$

The example emphasizes that full CR is unlikely to be the optimal policy. When demands are proportional, the condition requires that marginal costs (weighted by the semi-elasticities  $\sigma$ ) are higher in the low-cost market  $B$  than in the high-cost market  $A$ , at the common full CR prices, even though average costs are higher in market  $A$  by assumption. In a single-product setting, I have  $\sigma_L = 0$ , so I recover the condition  $c_B > c_A$ .

### C1 Derivatives of profit

To make the notation tractable, let  $x = p_H$  and  $y = p_L$ . Moreover,  $f = f(x, y) = \pi_H(p_H, p_L)$  and  $g = g(x, y) = \pi_L(p_H, p_L)$ . Finally, let  $F = \bar{\pi}_H$  and  $G = \bar{\pi}_L$ . Let subscripts  $x, y$  denote partial derivatives (for instance,  $f_x = \frac{\partial f}{\partial x}$ ).

First, consider  $f = \pi_H = Q(p_H - AC_H)$ , so

$$f_x = \frac{d\pi_H}{dp_H} = Q'(p_H - AC_H) + Q(1 - AC'_H)$$

$$f_y = \frac{d\pi_H}{dp_L} = -Q' (p_H - AC_H) + QAC'_H - Q + Q = -f_x + Q$$

Second, consider  $g = \pi_L = (1 - Q) (p_L - AC_L)$ . Therefore,

$$g_y = \frac{d\pi_L}{dp_L} = Q' (p_L - AC_L) + (1 - Q) (1 + AC'_L)$$

$$g_x = \frac{d\pi_L}{dp_H} = -Q' (p_L - AC_L) + (1 - Q) [-AC'_L] - (1 - Q) + (1 - Q) = -g_y + (1 - Q)$$

## C2 Path of prices

Now consider the system

$$f(x(\chi), y(\chi)) = \chi F \Rightarrow f_x x' + f_y y' = F$$

$$g(x(\chi), y(\chi)) = \chi G \Rightarrow g_x x' + g_y y' = G$$

This can be written

$$\begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} F \\ G \end{bmatrix} \Rightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} = J^{-1} \begin{bmatrix} F \\ G \end{bmatrix}$$

assuming (for now) the invertibility of the matrix  $J$  at every point. Let  $\det(J) \neq 0$  be the determinant of  $J$ .

Explicitly computing  $J^{-1}$  yields

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{\det(J)} \begin{bmatrix} g_y & -f_y \\ -g_x & f_x \end{bmatrix} \begin{bmatrix} F \\ G \end{bmatrix} = \frac{1}{\det(J)} \begin{bmatrix} g_y F - f_y G \\ -g_x F + f_x G \end{bmatrix}.$$

Recall that  $f_y = -f_x + Q$ . Also,  $g_x = -g_y + (1 - Q)$ . Therefore,  $f_y + f_x = Q$  and  $g_y + g_x = 1 - Q$ . Therefore,

$$\det(J) = f_x g_y - f_y g_x = f_x (1 - Q) + Q g_y - Q (1 - Q) > 0.$$

Since I can write  $\det(J) = (f_x - Q)(1 - Q) + Q g_y > 0$  and  $f_x = \frac{d\pi_H}{dp_H} = Q' (p_H - c_H) + Q > Q$ .

Recall

$$\frac{f_x}{-Q} = \sigma_H (p_H - AC_H) + AC'_H - 1$$

$$\frac{g_y}{-(1-Q)} = \sigma_L (p_L - AC_L) - AC'_L - 1$$

where  $\sigma_H = -\frac{1}{Q} \frac{dQ_H}{dp_H} = -\frac{Q'}{Q} > 0$  and  $\sigma_L = -\frac{1}{1-Q} \frac{dQ_L}{dp_L} = -\frac{Q'}{1-Q} > 0$ .

The derivative of welfare will have terms of the form  $-Q \frac{d\Delta p}{d\chi} = -Q \left( \frac{dp_H}{d\chi} - \frac{dp_L}{d\chi} \right)$ . Therefore, I consider

$$-Q \frac{d\Delta p}{d\chi} = \frac{\bar{\pi}_H - \bar{\pi}_L \frac{Q}{1-Q}}{\sigma_H (p_H - AC_H) + \Delta AC'_H + \sigma_L (p_L - AC_L) - 1}.$$

### C3 Slope of welfare

Now using the fact that  $-\pi_{HA} = \pi_{HB} = \bar{\pi}_H$ , and similarly for  $L$ , the derivative of welfare  $\frac{dW}{d\chi} = -Q_A \frac{d\Delta p_A}{d\chi} - Q_B \frac{d\Delta p_B}{d\chi}$  becomes:

$$\begin{aligned} \frac{dW}{d\chi} = & \frac{Q_B \left[ \frac{\bar{\pi}_H}{Q_B} - \frac{\bar{\pi}_L}{1-Q_B} \right]}{\sigma_{HB} (p_{HB} - AC_{HB}) + \Delta AC'_B + \sigma_{LB} (p_{LB} - AC_{LB}) - 1} \\ & - \frac{Q_A \left[ \frac{\bar{\pi}_H}{Q_A} - \frac{\bar{\pi}_L}{1-Q_A} \right]}{\sigma_{HA} (p_{HA} - AC_{HA}) + \Delta AC'_A + \sigma_{LA} (p_{LA} - AC_{LA}) - 1} \end{aligned}$$

When the second product does not exist  $\pi_L = \sigma_L = 0$  and  $\Delta AC' = AC'$ , so I recover the expression for welfare of the baseline model.

Now notice

$$X_A = Q_A \left[ \frac{\bar{\pi}_H}{Q_A} - \frac{\bar{\pi}_L}{1-Q_A} \right] = Q_A \left[ Q_{HB} \frac{AC_{HA} - AC_{HB}}{Q_{HA} + Q_{HB}} - Q_{LB} \frac{AC_{LA} - AC_{LB}}{Q_{LA} + Q_{LB}} \right] > 0$$

It is reasonable to assume that this term is positive if there is a significant number of people purchasing  $H$  in set  $B$  ( $Q_{HB} > Q_{LB}$ ), since the difference in cost is likely to be particularly significant for the comprehensive product  $H$ , hence it is likely that  $AC_{HA} - AC_{HB} > AC_{LA} - AC_{LB}$ . For instance, in the baseline model,  $Q_{LB} = 0$ , so  $X_A > 0$

Similarly,

$$X_B = Q_B \left[ \frac{\bar{\pi}_H}{Q_B} - \frac{\bar{\pi}_L}{1-Q_B} \right] = Q_B \left[ Q_{HA} \frac{AC_{HA} - AC_{HB}}{Q_{HA} + Q_{HB}} - Q_{LA} \left( \frac{AC_{LA} - AC_{LB}}{Q_{LA} + Q_{LB}} \right) \right] > 0.$$

Therefore, I obtain

$$\frac{dW}{d\chi} = \frac{\frac{X_B}{\sigma_{HB}(p_{HB} - AC_{HB}) + \Delta AC'_B + \sigma_{LB}(p_{LB} - AC_{LB}) - 1}}{\frac{X_A}{\sigma_{HA}(p_{HA} - AC_{HA}) + \Delta AC'_A + \sigma_{LA}(p_{LA} - AC_{LA}) - 1}}.$$

Now notice that

$$X_A > X_B \Leftrightarrow Q_B > Q_A$$

## C4 Proofs

### Proof of Proposition 4. Full PD (2)

Full PD maximizes welfare if  $W'(0) < 0$ . I require that, at full PD,

$$\frac{X_B}{\sigma_{HB}(p_{HB} - AC_{HB}) + \Delta AC'_B + \sigma_{LB}(p_{LB} - AC_{LB}) - 1} < \frac{X_A}{\sigma_{HA}(p_{HA} - AC_{HA}) + \Delta AC'_A + \sigma_{LA}(p_{LA} - AC_{LA}) - 1}.$$

I assume  $X_A > X_B$ , implying that  $Q_B > Q_A$ .

Then, a sufficient condition is

$$\sigma_{HB}(p_{HB} - AC_{HB}) + \Delta AC'_B + \sigma_{LB}(p_{LB} - AC_{LB}) > \sigma_{HA}(p_{HA} - AC_{HA}) + \Delta AC'_A + \sigma_{LA}(p_{LA} - AC_{LA}).$$

At  $\chi = 0$ ,  $p_{Hm} = AC_{Hm}$  and  $p_{Lm} = AC_{Lm}$ , so a sufficient condition is

$$\Delta AC'_B > \Delta AC'_A.$$

□

### Proof of Proposition 9. Full CR (2)

Full CR maximizes welfare if  $W'(1) > 0$ . I require that, at full CR,

$$\frac{X_B}{\sigma_{HB}(p_{HB} - AC_{HB}) + \Delta AC'_B + \sigma_{LB}(p_{LB} - AC_{LB}) - 1} > \frac{X_A}{\sigma_{HA}(p_{HA} - AC_{HA}) + \Delta AC'_A + \sigma_{LA}(p_{LA} - AC_{LA}) - 1}.$$

I assume  $X_A < X_B$ , implying that  $Q_B < Q_A$ .

Then, a sufficient condition becomes

$$\sigma_{HB}(\bar{p}_H - AC_{HB}) + \Delta AC'_B + \sigma_{LB}(\bar{p}_L - AC_{LB}) < \sigma_{HA}(\bar{p}_H - AC_{HA}) + \Delta AC'_A + \sigma_{LA}(\bar{p}_L - AC_{LA}).$$

At full CR,  $\bar{p}_H = \frac{Q_{HA}AC_{HA} + Q_{HB}AC_{HB}}{Q_{HA} + Q_{HB}}$  and  $\bar{p}_L = \frac{Q_{LA}AC_{LA} + Q_{LB}AC_{LB}}{Q_{LA} + Q_{LB}}$ .

Therefore,

$$\bar{p}_H - AC_{HA} = \frac{Q_A AC_{HA} + Q_B AC_{HB} - AC_{HA} Q_A - AC_{HA} Q_B}{Q_A + Q_B} = -Q_B \frac{AC_{HA} - AC_{HB}}{Q_A + Q_B} < 0$$

Similarly

$$\bar{p}_H - AC_{HB} = Q_A \frac{AC_{HA} - AC_{HB}}{Q_A + Q_B} > 0$$

$$\bar{p}_L - AC_{LA} = -Q_{LB} \frac{AC_{LA} - AC_{LB}}{Q_{LA} + Q_{LB}} < 0$$

$$\bar{p}_L - AC_{LB} = Q_{LA} \frac{AC_{LA} - AC_{LB}}{Q_{LA} + Q_{LB}} > 0$$

Therefore, the condition becomes

$$(AC_{HA} - AC_{HB}) \frac{\sigma_{HA} \frac{1}{Q_{HA}} + \sigma_{HB} \frac{1}{Q_{HB}}}{\frac{1}{Q_{HA}} + \frac{1}{Q_{HB}}} + (AC_{LA} - AC_{LB}) \frac{\sigma_{LA} \frac{1}{Q_{LA}} + \sigma_{LB} \frac{1}{Q_{LB}}}{\frac{1}{Q_{LA}} + \frac{1}{Q_{LB}}} < \Delta AC'_A - \Delta AC'_B.$$

□

**Proof of Proposition 5.** Optimal CR (2 products)

Finally, I consider the optimal level of  $\tilde{\chi}$ . First, notice

$$\frac{d\pi_{Hm}}{dp_{Hm}} \frac{1}{Q} = \frac{Q'_H (p_H - c_H) + Q_H}{Q_H} = -\sigma_H (p_H - c_H) + 1$$

$$\frac{d\pi_{Lm}}{dp_{Lm}} \frac{1}{1-Q} = \frac{-Q'_L (p_L - c_L) + Q_L}{Q_L} = -\sigma_L (p_L - c_L) + 1$$

Therefore, I can express  $\frac{dW}{d\chi} = 0$  as

$$\frac{\sigma_{HB} (p_{HB} - c_{HB}) + \sigma_{LB} (p_{LB} - c_{LB}) - 1}{X_B} = \frac{(\sigma_{HA} (p_{HA} - c_{HA}) + \sigma_{LA} (p_{LA} - c_{LA}) - 1)}{X_A}$$

□

## D Additional Data Details

### D1 Other contract dimensions

Individuals can choose whether their income increases in a deterministic manner or stays nominally flat over time, but the majority chooses a nominally flat profile. Individuals can choose whether payments occur monthly, quarterly or yearly, but

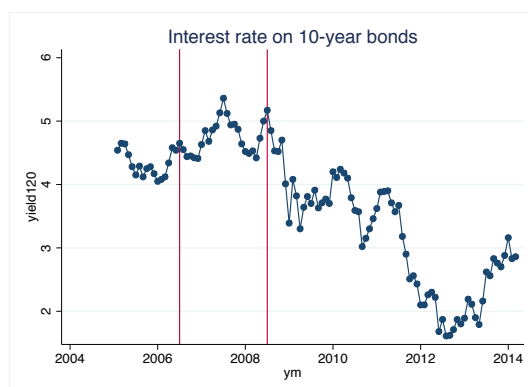


Figure 15: Timeline of long-term interest rates. The red lines illustrate the period describe by the data.

the vast majority choose monthly payments.

Beyond this, there exist other annuity products which have their own prices are therefore constitute entirely separate sets. This is the case for “joint” annuities and “enhanced annuities.” I do not consider these sets, instead restricting attention to “single life” and standard” annuities.<sup>50</sup>

I consider only annuities that are not part of an employer pension scheme. I consider only nominally flat, single-life and non-enhanced annuities.

## D2 Timeline of Interest Rates

Figure 15 shows the interest rate on 10-year UK government bonds from 2005 to 2014. In late 2008, the Bank of England began a policy of “quantitative easing” which induced a significant drop in interest rates. However, this occurs only after the period considered in the data.

## D3 Details of Price Imputation

Figure 16 illustrates that rates do not depend on whether the individual used of a financial advisor or on her internal/external status.

Figure 17 shows the results of the imputation for Men 65.

<sup>50</sup>If a joint annuity is purchase, a second individual (usually the spouse of the main annuitant) will continue to receive payments after the first annuity dies, until the second annuitant dies. I consider only single life annuities. Individuals who are significantly ill can purchase an “enhanced annuity” which yields higher rates.



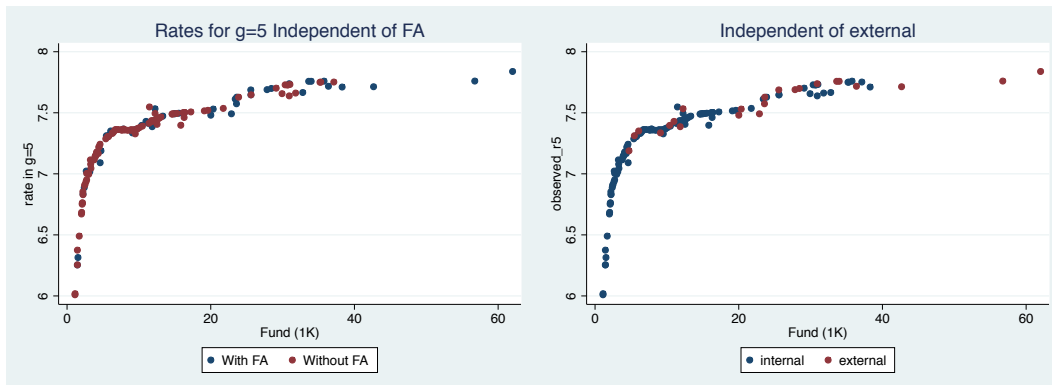


Figure 16: Rates offered for  $g_5$ , to Men 65, in a given month. The graphs illustrates that the rule used does not depend on whether the individual used of a financial advisor or on her internal/external status.

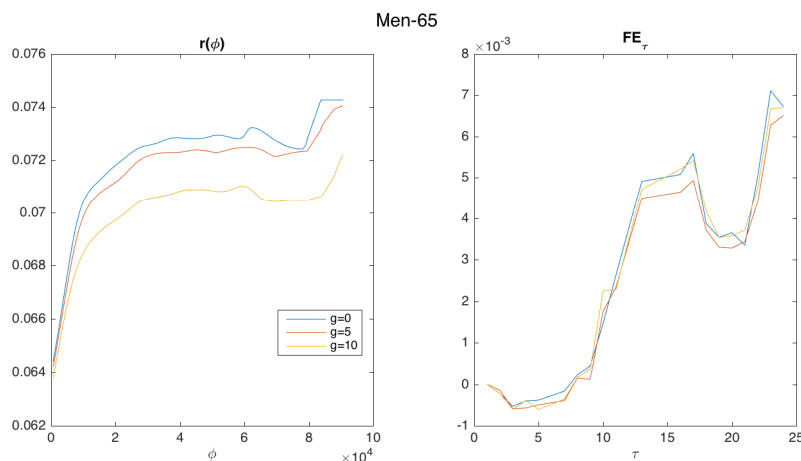


Figure 17: Semi-parametric imputation of rates for for Men 65.

I use the entire dataset to impute the missing rates. However, the semi-parametric estimator I use performs poorly on the boundaries of the data range. Therefore, in estimation, I use only individuals with  $\phi$  between £3,000 and £40,000. This implies only a marginal loss of data, as this range contains about 95% of the observations. However, this additional sample selection implies one must be careful to extrapolate the results to consumers with very low or very high levels of wealth. For consistency, in estimating demand, I will use only the imputed rates. That is, I do not utilize observed rates for the contract that was chosen.

I take the level of annuity rates to be those corresponding to the average value

of  $FE_\tau$  for each consumer subset. Doing so implies the rates I consider do not vary over time with interest rates, which would introduce concerns about endogeneity.<sup>51</sup> Notice, however, that rates nonetheless vary with fund size  $\phi$ .<sup>52</sup>

## E Optimal Savings Path

The overall pay-off from choosing contract  $k$  is obtain by maximizing  $\sum_{t=1}^T \delta^t S_t u(c_t) + \sum_{t=1}^{T+1} \delta^t H_t v(w_t)$ , subject to the constraint  $w_{t+1} = R_t(w_t + y_t - c_t) \Leftrightarrow 0 = w_t + z_t - c_t - \frac{w_{t+1}}{R_t}$ . Notice that  $w_t$  is wealth at the beginning of the period, before any annuity payments have accrued, and  $w_1$  is exogenous. The Lagrangian for this problem is

$$\mathcal{L} = \sum_{t=1}^T \delta^t S_t u(c_t) + \sum_{t=1}^{T+1} \delta^t H_t v(w_t) + \sum_{t=1}^T \psi_t \left[ w_t + z_t - c_t - \frac{w_{t+1}}{R_t} \right],$$

where  $\psi_t$  is the multiplier associated to the constraint in period  $t$ .

The First Order Conditions are

$$S_t c_t^{-\gamma} - \psi_t = 0 \Rightarrow S_t c_t^{-\gamma} = \psi_t$$

$$H_t \beta w_t^{-\gamma} + \psi_t - \psi_{t-1} \frac{1}{R_{t-1}} = 0 \Rightarrow \psi_t = \frac{\psi_{t-1}}{R_{t-1}} - H_t \beta w_t^{-\gamma}$$

Typically, this problem is solved by backward induction. However, the structure of the problem in this case permits a procedure that is more accurate and robust, as highlighted in EFS. Given the optimal choice for  $c_1$ , called  $c_1^*$ , and the known value of  $w_1$ , it is possible to compute  $\psi_1$  from the FOC  $S_1 c_1^{-\gamma} = \psi_1$ . These elements can then be used to compute  $(w_2, \psi_2, c_2)$  and so forth. For instance, the second step of this loop (for  $t = 2$ ) is

1. Compute  $w_2 = R_1(w_1 + y_1 - c_1)$ ,
2. Compute  $\psi_2 = \frac{\psi_1}{R_1} - H_2 \beta w_2^{-\gamma}$ ,
3. Compute  $c_2 = \left( \frac{\psi_2}{S_2} \right)^{-\frac{1}{\gamma}}$ .

Therefore, the problem becomes one of finding the values of  $c_1$  which maximizes the program. This requires simply maximizing over the initial value of consumption  $c_1$ .

<sup>51</sup>For instance, a change in interest rates would affect contract choices through a change in the annuity rates, but would also have a potential impact on decisions due to its effect on each individual's optimization of her portfolio of other assets.

<sup>52</sup>The estimates are robust to allowing the annuity rates to vary over time, which is not surprising given the relatively short time window under consideration.

Notice that the solution scales up with  $\phi$  and therefore the optimal choice is independent of  $\phi$ . Suppose that initial wealth increases from  $w_0$  to  $\mathcal{I}w_0$ . Similarly, payments increase from  $y_t$  to  $\mathcal{I}y_t$ . If the optimal consumption path changes from  $c_t^*$  to  $\mathcal{I}c_t^*$  and the optimal path of wealth changes from  $w_t^*$  to  $\mathcal{I}w_t^*$ , then the new consumption and wealth paths satisfy the problem's first order conditions. Therefore, the value of the problem changes from  $V^*$  to  $\mathcal{I}V^*$ . Therefore, the choice of the consumer between the two income streams becomes unchanged.

## F Annuity Comparison Websites

Figure 18 illustrates one of several websites that annuity buyers can use to compare the annuities offered by different firms. Actively choosing between annuity providers was strongly encouraged by UK regulatory bodies during the period covered by the data. Moreover, the rates offered by different providers are quite similar, suggesting a symmetric competitive outcome in this market.

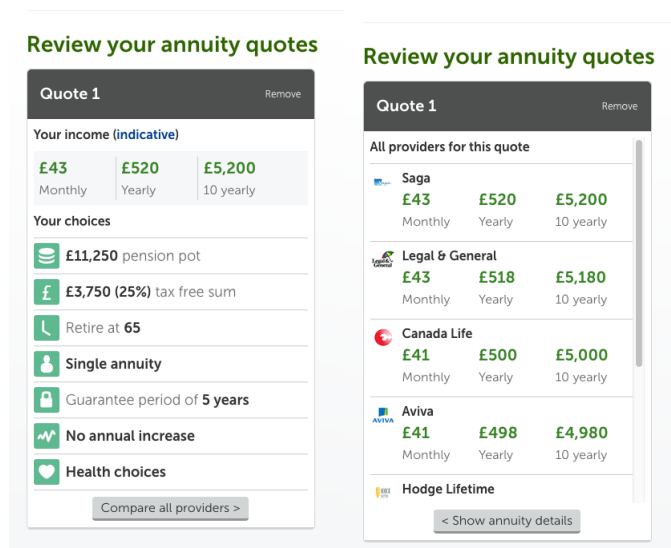


Figure 18: Screenshots of an online search engine for comparing annuities.

## G Equilibria and Counterfactuals

### G1 Equilibria

In practice, to compute the equilibrium-rates vector  $r^* = (r_0^*, r_5^*, r_{10}^*)$ , I find the vector  $r$  that minimizes the function  $\zeta(r) = \max(\|\pi_0(r)\|, \|\pi_5(r)\|, \|\pi_{10}(r)\|)$ . At the

numerical equilibrium points,  $\zeta(r^*)$  is of the order of  $10^{-3}$ , suggesting a close approximation to the true zero-profit vector of rates  $r^*$ .

The following figures describe the equilibrium in each gender-age subset of individuals.

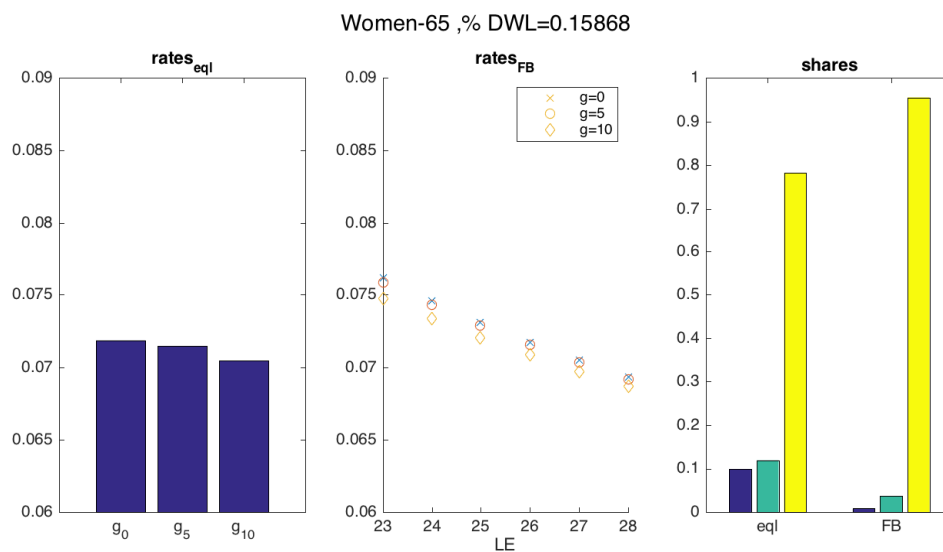


Figure 19: Equilibrium in the subset Women 65.

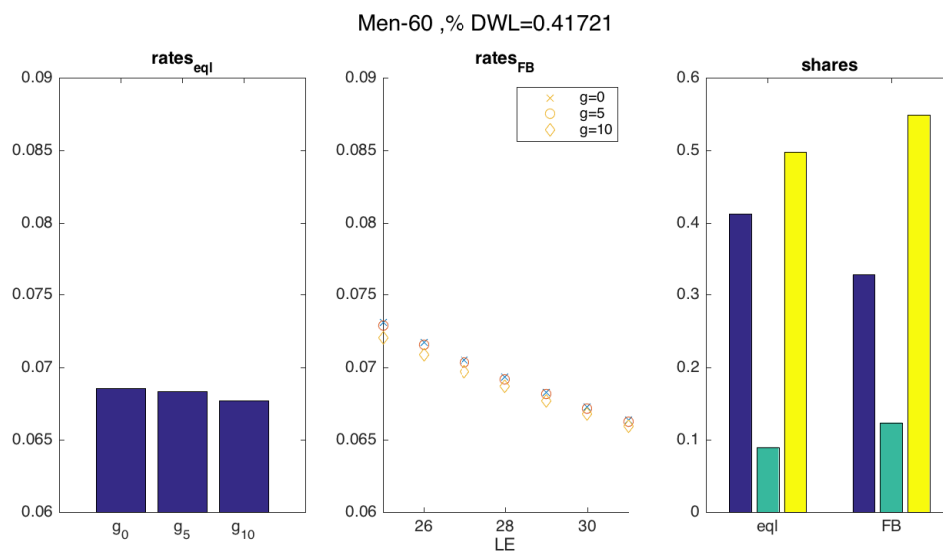


Figure 20: Equilibrium in the subset Men 60.

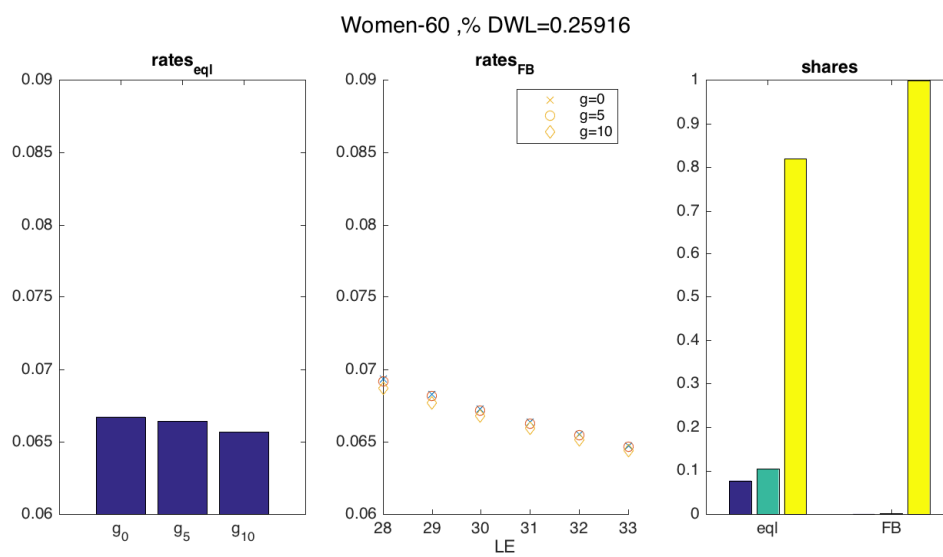


Figure 21: Equilibrium in the subset Women 60.

## G2 Computing partial CR policies

Formally, I begin by computing the vector of rates for each market under full PD,  $r_A^*, r_B^*$ . Then, I compute rates under full CR,  $\bar{r}$ . Then, for  $\chi \in (0, 1)$ , I impose  $r_B = (1 - \chi)r_B^* + \chi\bar{r}$ . Therefore,  $\chi = 0$  corresponds to full PD while  $\chi = 1$  corresponds to full CR.

For each value of  $r_B$ , I find the value of  $r_A$  at which all contracts break even across both consumer subsets. In practice, I find the value of  $r_A$  that minimizes

$$\zeta(r_A) = \max(\|\pi_{0A}(r_A) + \pi_{0B}(r_B)\|, \|\pi_{5A}(r_A) + \pi_{5B}(r_B)\|, \|\pi_{10A}(r_A) + \pi_{10B}(r_B)\|),$$

For each value of  $\chi \in (0, 1)$ , the profit per person in each contract remains of the order of  $10^{-3}$ , suggesting the obtained vector  $r_B$  is indeed the zero-profit point. Moreover, for different values of  $\chi$ , profit per person in each contract fluctuates randomly, suggesting numerical error near the true equilibrium points.

## H Calibration

I calibrate the welfare implications of the model in Section 2 using the estimates obtained by [Handel, Hendel and Whinston \(2015\)](#) (HHW) in a health insurance context.

### H1 Setup

I consider a continuum of consumers, who experience wealth shocks drawn from a Gaussian distribution  $\mathcal{N}(\mu, \sigma^2)$ . The mean ( $\mu$ ) and variance ( $\sigma^2$ ) differs across consumers. Each consumer choose an insurance contract (described below) to maximize their CARA utility  $\mathbb{E}[e^{-ac}]$ , where the expectation is taken over realizations of the wealth shock,  $c$  is consumption in each state of the world and  $a$  is each individual's (heterogeneous) risk aversion parameter. I define each individual's "insurance value" as  $v = a\sigma^2$ .

I assume contracts consist only of a price  $p$  and an actuarial rate  $x$ . That is, a contract will absorb a share  $x$  of the individual's wealth shock. There are two products in this market,  $(p_H, x_H)$  and  $(p_L, x_L)$ . Following [Handel, Hendel and Whinston \(2015\)](#), I set  $x_H = 0.9$ , and describe  $x_L$  below.

An insurance product with coverage  $x$  implies that the insurer absorbs a share  $x = 0.9$  of the buyer's wealth draw, leaving the buyer to pay a share  $1 - x$ . Consumer willingness to pay for this contract is

$$u = x\mu + \frac{1}{2} \left(1 - (1 - x)^2\right) v.$$

where  $\mu$  captures the individual's expected loss and  $v$  captures her risk aversion and

the variance of wealth shocks. The relevant types in this setting are  $(\mu, v)$ .

Competitive insurers offer insurance contracts characterized by a coverage-price pair  $(x, p)$ . The cost to the insurer is of covering an individual of type  $\mu$  is  $c = x\mu$ . No moral hazard exists, so the “first best” occurs when all consumers buy the contract.

## H2 Distribution of types

I assume that  $\mu, v$  have a joint log-normal distribution:

$$\begin{pmatrix} \log(\mu) \\ \log(v) \end{pmatrix} \sim \mathcal{N} \begin{pmatrix} \bar{\mu} & \sigma_{\mu}^2 & \rho\sigma_{\mu}\sigma_v \\ \bar{v} & \rho\sigma_{\mu}\sigma_v & \sigma_v^2 \end{pmatrix}.$$

The correlation parameter  $\rho$  captures the degree of adverse selection in this market.

I calibrate this distribution to the moments of the data in HHW data.<sup>53</sup> To two significant digits, those moments are:  $\mathbb{E}[\mu] \approx 6.6 \cdot 10^3$ ,  $\mathbb{E}[v] \approx 6.7 \cdot 10^4$ ,  $\mathbb{V}[\mu] = 5.0 \cdot 10^7$  and  $\text{Cov}[\mu, v] \approx 6.3 \cdot 10^7$ . I do not know  $\mathbb{V}[v]$ , but the variation in  $a$  seems to be quite small and only weakly correlated with that in  $\sigma^2$ , so I set  $\mathbb{V}[v] = \mathbb{E}[a]^2 \mathbb{V}[\sigma^2] \approx 9.8 \cdot 10^9$ .

Then  $\mathbb{E}[\mu] = e^{\bar{\mu} + \frac{1}{2}\sigma_{\mu}^2}$ ,  $\mathbb{E}[v] = e^{\bar{v} + \frac{1}{2}\sigma_v^2}$ ,  $\mathbb{V}[\mu] = \mathbb{E}[\mu]^2 (e^{\sigma_{\mu}^2} - 1)$ ,  $\mathbb{V}[v] = \mathbb{E}[v]^2 (e^{\sigma_v^2} - 1)$  and  $\text{Cov}[\mu, v] = \mathbb{E}[v]\mathbb{E}[\mu] (e^{\rho\sigma_{\mu}\sigma_v} - 1)$ . This system of equations of 5 equations with 5 unknowns can be solved analytically and uniquely to yield  $\sigma_{\mu}^2 = 8.4$ ,  $\bar{v} = 11$ ,  $\sigma_{\mu}^2 = .76$ ,  $\sigma_v^2 = 1.2$  and  $\rho\sigma_{\mu}\sigma_v = .63$ .

The log-correlation  $\rho$  is often used as a measure of adverse selection, because  $\rho$  induces a greater correlation between cost and WTP. For additional details about the data, see [Veiga and Weyl \(2016\)](#).

## H3 Equilibrium and Deadweight Loss

In what follows, I compute the equilibrium of each set as the point where  $\Delta P = \Delta AC$ , as in HHW. For the case of multiple products ( $x_L > 0$ ), this zero-profit point need not be a Nash equilibrium, and indeed such an equilibrium need not exist. However, [Azevedo and Gottlieb \(2015\)](#) show that a zero-profit point always exist, and that any such point is the limit of the Bertrand-Nash equilibria of a sequence of sets that become arbitrarily competitive (and firms have capacity constraints). In my setting, this point is always unique and I take it to be the equilibrium of the game.<sup>54</sup>

I compute the deadweight loss (DWL) as the difference between realized welfare and welfare under symmetric information (in which case it is clear that all individuals purchase the most generous insurance contract available).

<sup>53</sup>I think the authors for generously allowing me to use these results.

<sup>54</sup>For additional discussion regarding the existing of Riley and Wilson equilibria, see HHW.

#### H4 Differences across consumer sets

Since,  $\bar{\mu}$  captures the log-mean-cost in each consumer subset, it will determine which set is “high-cost” and “low-cost.” Let the log-mean of  $\mu$  in set  $m$  be  $\bar{\mu}_m$ , and let the log-correlation be  $\rho_m$ .

I consider two consumer sets,  $m \in \{A, B\}$ . The log-mean of  $\mu$  in set  $m$  is  $\bar{\mu}_m$ . set B will correspond to the HHW estimates, whereas set. set A is the “high-cost” market, and therefore  $\bar{\mu}_A = 1.04\bar{\mu}_B$ .

For computational simplicity, I impose that  $p_A$  follows a linear path from  $\bar{p}$  to  $p_A^*$ . That is, for  $\chi \in (0, 1)$ , I impose  $p_A = (1 - \chi)p_A^* + \chi\bar{p}$ . Therefore,  $\chi = 1$  corresponds to full CR, whereas  $\chi = 0$  corresponds to full PD. Then, given  $p_A$ , I compute the level of  $p_B$  at which the industry breaks even across both sets. I find the value of  $p_B$  that minimizes  $\zeta(p_B) = \|\pi_A(p_A) + \pi_B(p_B)\|$ . For all cases under study,  $\zeta(p_B)$  has a unique minimum. At each level of  $p_B$ , I compute the deadweight loss (DWL) as the difference between realized welfare and the first-best welfare.

#### H5 1 product ( $x_L = 0$ )

I begin by assuming that there is a single product, so  $x_L = 0$ . Figure 22 considers the case in which  $\rho_A = 1.3\rho_B$ , so the high-cost set has greater adverse selection. Therefore, CR is likely to outperform PD in this case. Full PD leads to an 8% increase in total DWL, relative to CR. However, full CR is not optimal: allowing a small amount of PD achieves a reduction in DWL of 0.4% relative to full CR.

Figure 23 considers the case in which  $\rho_A = 0.6\rho_B$ . The high-cost subset now has less significant adverse selection. In this case, full PD achieves a reduction in DWL of about 4.2% relative to full CR. However, full PD is not the optimal policy. A small constraint on price differences reduces DWL by 2.3% relative to full PD.

#### H6 2 products

Now, I assume  $x_L = 0.6$ , following HHW. I consider the same combinations of  $\bar{\mu}_m$  and  $\rho_m$  as above. I impose that the prices in set A change linearly from their CR level to its PD level. That is, for  $\chi \in (0, 1)$ , I impose  $p_{HA} = (1 - \chi)p_{HA}^* + \chi p_{\bar{H}A}$  and  $p_{LA} = (1 - \chi)p_{LA}^* + \chi p_{\bar{L}A}$ . Then, I compute the prices in set B, at which each contract ( $H$  and  $L$ ) breaks even across both sets. In practice, I choose the pair  $(p_{HB}, p_{LB})$  that minimize

$$\max [\|\pi_{HA}(p_{HA}, p_{LA}) + \pi_{HB}(p_{HB}, p_{LB})\|, \|\pi_{LA}(p_{HA}, p_{LA}) + \pi_{LB}(p_{HB}, p_{LB})\|].$$

Again, this function has a unique maximum in the region of prices that I considered. The results trace a smooth path between the PD and CR prices for set B, as shown below.

Figure 24 considers the case where the high-cost set (A) also has greater adverse selection ( $\rho_A > \rho_B$ ). In this case, it is expected that CR can be beneficial. Full PD



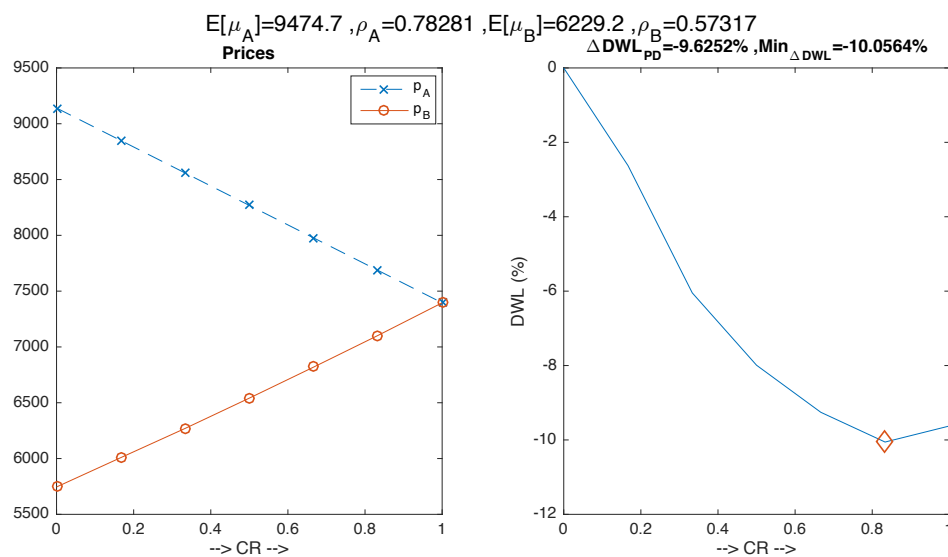


Figure 22: CR policy when the high-cost consumer subset has greater adverse selection.

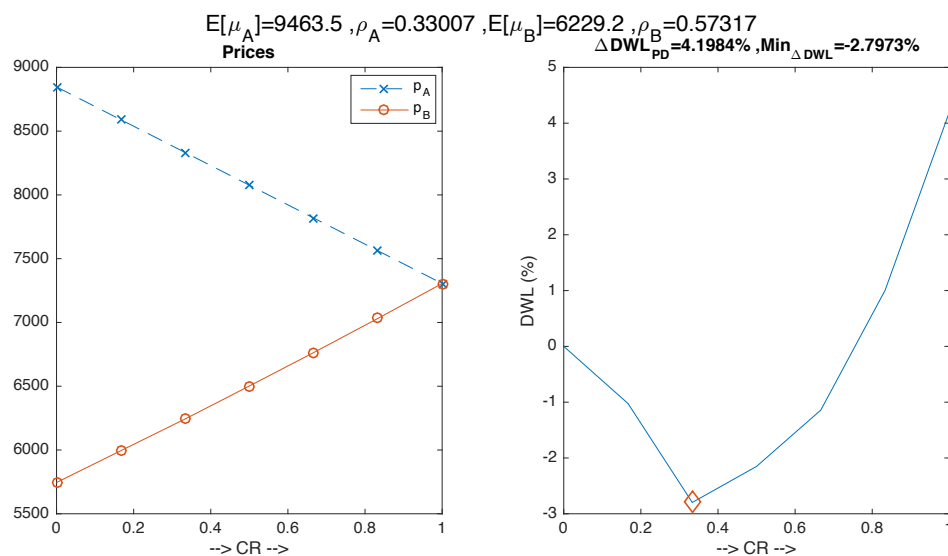


Figure 23: An intermediate constraint on PD when the high-cost set has less significant adverse selection.

results in approximately the same amount of DWL as full CR. However, the optimal constraint on PD achieves a reduction in DWL of about 7%.

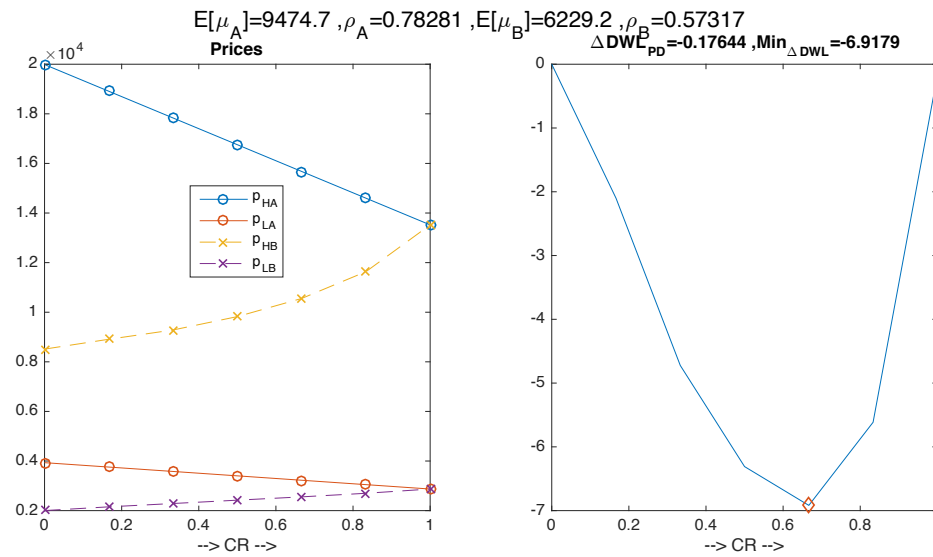


Figure 24: An intermediate constraint on PD when the high-cost set has greater adverse selection.

Figure 25 considers the case where the high-cost consumer subset A has less significant adverse selection ( $\rho_B > \rho_A$ ). In this case, full PD achieves a reduction in DWL of about 6.5% relative to full CR. However, the optimal intermediate level of price constraint achieves a reduction in DWL of about 1% relative to full PD.

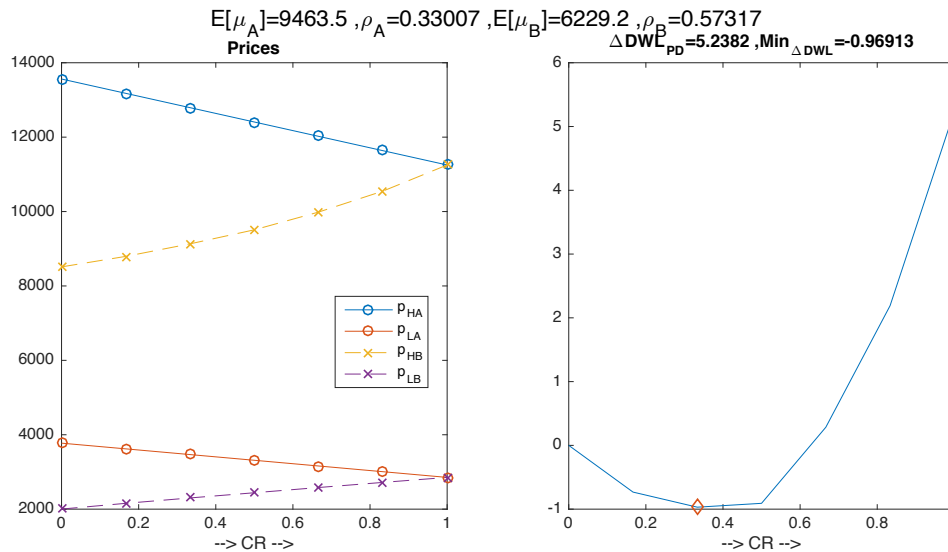


Figure 25: An intermediate constraint on PD.

## I Micro-foundations for willingness to pay

Suppose  $u = EX(u) + IV(u)$ , where  $EX(u) \geq 0$  are expenditures and  $IV(u)$  is “insurance value.” Suppose  $c(u) = EX(u) + ADM$  where  $ADM \geq 0$  is an administrative “load.” If, as in HHW,  $IV(u) > ADM = 0$ , then  $u > c(u)$ ,  $\forall u$ , so the optimal price is  $p^{**} = 0$ . If for some  $u$ ,  $IV(u) < ADM$ , then the socially optimal price is  $p^{**} > 0$ . Also, if  $IV(u) < 0$  for some  $u$  (e.g., due to moral hazard), then again the socially optimal price is  $p^{**} > 0$ .

Suppose that there exist some individuals for whom  $u = c(u) = 0$ . For instance, these individuals are risk neutral and have zero expected expenditures (as in HHW). Then, selection cannot be globally advantageous ( $c'(p) < 0, \forall p$ ) since  $c(u) \geq 0$ . However, selection can be advantageous over a policy-relevant range of prices, and can be globally adverse.<sup>55</sup>

The calibration presented in Appendix H provides an explicit expression for  $u$  for the case when individuals have CARA utility and are subject to Gaussian wealth shocks.

<sup>55</sup>This also occurs in Einav, Finkelstein and Cullen (2010). In Einav, Finkelstein and Cullen (2010), those who are the least willing to purchase insurance correspond to  $Q = 1$ . If there are people with zero cost and risk neutral, they will be the ones with lowest WTP for insurance and their cost is zero, so  $MC(Q = 1) = 0$ . Therefore I cannot have advantageous selection ( $MC'(Q) > 0$ ) everywhere. However, I can nonetheless have adverse selection ( $MC'(Q) < 0$ ) everywhere.