Imperfect Private Information in Insurance Markets

Adam Solomon[*]

April 15, 2022

Abstract

It is well known that private information can impair the functioning of insurance markets, and widely documented that individuals misperceive their private information. But these two facts are rarely analyzed jointly. I demonstrate theoretically that when contracts endogenously respond to misperceptions the covariance between risk type and risk misperception is central. The implications for equilibrium and welfare differ starkly according to according to this covariance, and are in contrast to assuming perfect private information or a uniform misperception across the population. I study a variety of risks - mortality, health, disability and long-term care - and find disparate patterns. While for some of these risks misperception is uniform across the population (e.g. long-term care and catastrophic medical risk), misperceptions often covary with risk in an 'S' shape: high risks overestimate their risk while low risks understimate theirs (e.g. mortality and disability risk). This suggests that welfare in some insurance markets is impaired by individuals' imperfect perception of their private risk type, while it is improved in others.

[*] Department of Economics, Massachusetts Institute of Technology. Email: adamsol@mit.edu. I thank Jim Poterba, Amy Finkelstein, Stephen Morris, Dmitry Taubinsky, Nathan Hendren, Hunt Allcott, Frank Schilbach, Rafael Veiel and Rob Townsend for very helpful discussions and comments.

1 Introduction

Accurately understanding and insuring risk is vital to economic well-being. A problem in the smooth functioning of insurance markets is the presence of private information known to individuals. Yet there is substantial evidence that individuals widely and systematically misperceive their risk in a variety of contexts.[1] I study the effect of private information on market equilibrium and welfare when that information is imperfectly perceived.

In this paper I analyze misperceptions of a variety of risks: mortality risk, long-term care risk, disability risk and catastrophic health risk. My insight is that high risk individuals and low risk individuals have different misperceptions, the contracts offered by firms will respond to these misperceptions, and sometimes these misperceptions will increase welfare while at other times they will decrease it.

I develop theory as to how misperceptions affect welfare when the contract menu is endogenous. Throughout welfare is utilitarian and accounts only for experienced utility (according to the real probabilities) not perceived utility (according to the misperceived probabilities). The primary setting is the binary loss framework of Rothschild and Stiglitz (1976) generalized to allow for misperceptions of risk types between and within risk classes. I study how the equilibrium changes with the introduction of risk misperceptions and what the welfare consequences are. I find that misperceptions by high risk types have a material effect on the equilibrium , while those by low risk types do not.

I show theoretically that the covariance between risk type and risk mispercetion has particular implications for equilibrium and welfare. The implications are distinct from if there was a common misperception for all risk types, or no misperception at all, or the opposite covariance between risk type and misperception. Since misperceptions endogenously determine which contracts are offered in equilibrium, I find that certain misperception patterns can enhance welfare while others can diminish it.

I investigate the robustness of these results to alternative market structures and equilibrium concepts. In this setting the equilibrium notions of Riley (1979) and Azevedo and Gottlieb (2017) are coincident with Rothschild and Stiglitz (1976). I then study the planner's problem and derive qualitatively identical conclusions to the competitive market. As a special case of the planner's problem I highlight the dynamics in the Wilson-Miyazaki-Spence (MWS) constrained efficient allocation. Assuming a monopoly instead of a competitive market leads to errors by all types mattering, but maintains that some over-perceptions by high types are welfare enhancing, while under-perceptions are not.

I empirically study the misperception of private information about mortality risk, long-term care risk, disability risk and catastrophic health risk. I use data from the Health and Retirement Study (HRS), a biennial panel survey mostly of older Americans that has run since 1990. The first crucial datum is a measure of what individuals subjectively perceive their risk to be. This comes from HRS questions such as: "What is the percent chance that you will live to be 75 or more?"

[1] See, for example, Handel et al. (2020), Handel and Kolstad (2015), Mueller et al. (2018) and Handel et al. (2018).

(mortality risk), or ""What about the chances that your health will limit your work activity during the next 10 years?" (disability risk).

Second, I compare this to an estimate of objective risk of these outcomes occuring. The HRS collects a rich set of medical and demographic covariates correlated with mortality, disability, long-term care and medical outcomes, and has run for long enough to directly observe the relevant outcomes for much of the sample. Using these predictors and outcomes, I can make an accurate and valid out-of-sample prediction of what the true prospective risks were for individuals when they were interviewed at earlier ages. This 'machine learned' LASSO prediction has substantially greater out-of-sample accuracy than individuals subjective perceptions or classical predictive algorithms such as OLS or logit.

I compare my objective prediction with the subject's subjective perception. I find that when there are individuals with high true risk they on average under-perceive their risk, and that individuals with low true risk, when they exist, over-perceive their risk. This shows that the misperception of risk is not uniform in the population but systematically varies with risk type. This typically 'S' shaped pattern of misperception is consistent with the behavioural economics literature, going back to Tversky and Kahneman (1992), in which individuals act as if probabilities are bent toward the middle. But for some risks almost everyone is low risk, and so only the pessimistic half of the 'S' shape occurs.

I interpret these quantitative findings through the lense of the theory. In situations with low risk and pessimism throughout, such as catastrophic health risk, welfare is enhanced by misperceptions relative to the adversely selected second best. For other risks, such as long-term care, the optimism of the high risks is more quantitatively significant than the pessimism of the low types. This pattern implies an even greater welfare loss than the direct effects of suppressed high type demand would suggest. These welfare conclusions are more subtle and sometimes of opposite sign than assuming a uniform misperception across the population allows for. Critically, the mechanism - equilibrium contracts endogenously responding to misperceptions - is often absent in the literature that fixes the contracts offered.

To test the robustness of the quantitative findings, I present analyses that take the subjective perception as an accurate representation of individuals' beliefs, as well as a number of analyses that account for measurement error in these elicitions. Measurement error seems particularly important given the well-known pattern in the data of individuals "rounding" their responses to focal numbers. I find that rounding error explains almost all of the misperception for disability risk, but that the qualitative patterns regarding mortality, long-term care and health risk remain robust to this accounting for rounding error.

This paper contributes to multiple literatures. The first contribution is to behavioural mechanism design, specifically to the study of departures from perfect rationality in an insurance context. The problem under perfect rationality was first studied in the canonical papers by Akerlof (1970), Rothschild and Stiglitz (1976), Stiglitz (1977), and to overcome the possible non-existence of equilibria alternative equilibrium notions were proposed by Riley (1979), Wilson (1977), Spence (1978), Miyazaki (1977) and most recently by Azevedo and Gottlieb (2017).

Working within the Rothschild and Stiglitz (1976) framework, there has been work that relaxes the informational assumptions. Young and Browne (2000) shows that the qualitative features remain in a non-EU framework, Chassagnon and Villeneuve (2005) characterizes the constrained-efficient frontier when risk misperceptions similar to this paper are made, and Sandroni and Squintani (2007) study a version of this model where optimistic high types are indistinguishable from low types and so cannot be screened apart.

In contrast, assuming a monopolistic rather than competitive market structure as in Akerlof (1970), Jeleva and Villeneuve (2004) studies misperceptions of risk as in this paper, and obtain a monopolistic analogue to the pooling result I obtain in section A.4. Sandroni and Squintani (2013) extend their earlier work to the monopolistic setting. A key departure of this paper is that I do not ex-ante rule out over-insurance resulting from misperceptions and as such obtain the possibility of welfare enhancing misperceptions.

More recently, based on the influential framework by Einav et al. (2010), a literature has documented empirically and theoretically the impact of choice frictions and misperceptions within a setting with fixed contracts. For example, in the setting of health insurance, studies include Abaluck and Gruber (2016a), Abaluck and Gruber (2016b), Handel and Kolstad (2015), Handel et al. (2015), Handel et al. (2018), Ericson et al. (2020), Einav et al. (2015), Bhargava et al. (2017), Ketcham et al. (2012), Fang et al. (2008), in the context of unemployment insurance Landais et al. (2017) and Mueller et al. (2018), and with regards to tax design Allcott et al. (2019) and Stantcheva (2020). Theoretically, Spinnewijn (2013) and Spinnewijn (2015) make related contributions by studying screening problems with biased agents. This paper builds on these contributions by allowing for misperceptions to endogenously determine the contracts offered in equilibrium.

Of particular note are Mueller et al. (2018) and Handel et al. (2020) who study how choice frictions and risk misperceptions vary across the population. The former documents how optimism is concentrated in those with a high risk of long spells of unemployment and derives implications for the optimal public unemployment insurance system. The latter demonstrates that choice quality between health insurance policies differs markedly across demographic, economic and occupational strata. They focus on the distributional implications that these frictions introduce. In these settings the contract menu is fixed, and so this channel by which misperceptions can affect the equilibrium is closed down by assumption. I focus on that channel and ask how the distribution of risk misperceptions with respect to risk affect the contract menu, choices and welfare that prevail in equilibrium.

Finally, this work speaks to the accuracy and predictive content in subjective elicitations, particularly within the Health and Retirement Study. This is a literature initiated by Hurd and McGarry (1995) and Hurd and McGarry (2002) that shows that short-run survival outcomes are predicted by responses to the longer-term mortality questions asked in the HRS. A drawback of the early papers was that long-run survival outcomes such as those asked about in the HRS (e.g. survival to 75, or 85) were not observable. A contribution of this paper is to examine how accurate those predictions were ex-post. Also, there is a literature using various econometric technqiues to de-round elicitations, for example Gan et al. (2003) and Kleinjans and Soest (2014). The typical finding is that even after a structural model to de-round elicitations is estimated, the quantitative changes are minimal and, in the context of the models explored in those papers, the qualitative changes essentially null.

The rest of this paper is organized as follows. Section 3 presents the empirical analysis, within which 3.1 describes the ideal data, 3.2 details and compares the actual data, 3.3 explains how the prediction algorithm for risk works, 3.4 sets out my solutions to the rounding error present in the data and 3.5 presents the empirical results. Section 2 contains the theoretical framework that accounts for the impact of risk misperceptions on the equilibrium contracts and welfare. Different market structures are explored in sections 2.2, A.2 and A.1. Finally, section 4 concludes.

2 Theory

There is a unit mass of agents, all of whom are endowed with wealth $w > 0$ and face a stochastic loss of size $l > 0$. There are two types of agents. The fraction $\alpha_H \in (0, 1)$ of the population are high risk types, for whom the probability of the loss occurring is $p_H \in (0, 1)$. The complementary $\alpha_L = 1 - \alpha_H$ of the agents are low risk for whom the loss occurs with probability $p_L < p_H$.[(2)]

The risk of loss occurring for each agent is private information unobservable to the insurer. Unlike in the canonical model of Rothschild and Stiglitz (1976) where their private risk of loss is perfectly known by each agent, here the agents may misperceive this probability. High type agents believe their risk of loss to be q_H where we allow $q_H \neq p_H$, and similarly for low type agents. For a given consumption bundle in the *loss* (L) and *no loss* (NL) states of the world, denoted $\boldsymbol{c} = (c^{NL}, c^L)$, each agent with *perceived* risk $q_i \in \{q_L, q_H\}$ receives value

$$V(q, \boldsymbol{c}) = (1 - q_i)u\left(c_{NL}\right) + q_i u\left(c_L\right),$$

where $u(\cdot)$ is the VnM utility function assumed to be twice continuously differentiable, with $u'(c) > 0$ and $u''(c) < 0$.

The supply side of this market consists of a continuum of identical insurers. Each contract offered by an insurer specifies consumption in each state of the world. A typical contract is labelled $\boldsymbol{c} = (c^{NL}, c^L)$. Selling such a contract to a consumer with *true* risk $p_i \in \{p_L, p_H\}$ results in a profit of

$$\begin{aligned}\pi(p_i, \boldsymbol{c}) &= (1 - p_i)\left(W - c^{NL}\right) - p_i(c_L - (W - l)) \\ &= W - (1 - p_i)c^{NL} - p_i(c^L + l).\end{aligned}$$

Firms know the true and perceived risks, p_i and q_i as well as the relative sizes α_i, for each

[(2)] All of the analysis that follows extends straightforwardly to any finite number of types. For clarity of exposition I discuss only the two type case throughout.

$i \in \{H, L\}$ of each group in the population, but they cannot identify which group a particular individual falls into.

As shorthand, write the marginal rate of substitution as (the absolute value of) the slope of the indifference curve of an individual with (real or perceived) risk q indifferent to contract $\boldsymbol{c}$ as

$$MRS(q, \boldsymbol{c}) = \frac{1-q}{q} \frac{U'(c^{NL})}{U'(c^{L})},$$

and note the slope of the zero profit line when individuals of average type p are insured is $\frac{1-p}{p}$.

2.1 Equilibrium with Misperceptions

The equilibrium concept I use, following Rothschild and Stiglitz (1976), Sandroni and Squintani (2007), Sandroni and Squintani (2013), is of a *locally-competitive* equilibrium. This captures the idea that firms experiment with new policies that are similar to the current market offering.

Definition 1. *A set of contracts $\mathcal{C}$ is a* ***locally-competitive equilibrium*** *when: 1. No contract $\boldsymbol{c} \in C$ makes strictly negative profits; 2. There exists $\epsilon > 0$ such that every possible contract $\boldsymbol{c}'$ with $\|\boldsymbol{c} - \boldsymbol{c}'\| < \epsilon$ for each $\boldsymbol{c} \in C$ does not make strictly positive profits if offered in addition to $\mathcal{C}$.*

For the rest of the paper, when I write equilibrium I mean locally-competitive equilibrium, unless otherwise specified. In this context with single-dimensional types and no further heterogeneity between agents, the (typically unique) local equilibrium coincides with the alternative equilibrium relaxations used by Riley (1979) and Azevedo and Gottlieb (2017). Moreover, the local equilibrium is the unique Nash equilibrium, when the latter exists. A typical sufficient condition for existence is that the mass of low types is not too large, although I will not explore that issue further.

To understand the forces at work once misperceptions are introduced, the natural point of comparison is the equilibrium in the model of Rothschild and Stiglitz (1976) which assumes no misperceptions, that is, in this terminology, $q_H = p_H, q_L = p_L$. The unique equilibrium is illustrated in figure 1.

The high types purchase the contract marked $\boldsymbol{c}_H$, which offers full insurance whilst earning the insurers zero profits. The high types indifference curve through $\boldsymbol{c}_H$ is marked IC_H. The low types receive contract $\boldsymbol{c}_L$ defined by the intersection of IC_H with the zero profit line for the low type.

The key qualitative features of the RS equilibrium are: They high risk types receive their first best, full insurance contract (subject to the contract breaking even). The low risk types receive only partial insurance, with this distortion due to the information rents that accrue to the high types due to private information. In order for the high types to self-select into the contract designed for them, the low-type contract must be made sufficiently unattractive. For a high risk type who values insurance highly, a downward distortion in the low risk contract will successfully deter the high types from swapping contracts.

These forces are not specific to the RS model. As shown in subsection A, the incentive compatibility constraint always distorts down the contract received by the low types. In essentially any

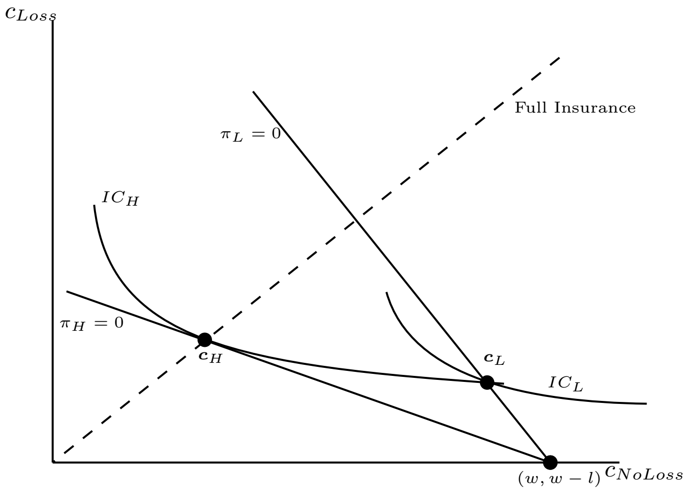


Figure 1: Rothschild Stiglitz Equilibrium

competitive model this incentive compatibility constraint must be present. Only if the government compelled everyone to reveal their types, or compelled insurers to offer just one defined contract, could an incentive compatibility constraint be avoided.

Now with the primary forces in hand I move to an analysis of misperceptions. The equilibrium characterization is stated in the following proposition. All proofs are in the appendix.

Proposition 1. *Suppose $\|\boldsymbol{p} - \boldsymbol{q}\|_\infty < \xi$, with $q_H > q_L$. For small enough ξ,*[(3)] *in the unique locally-competitive equilibrium, high risk individuals buy the contract $\boldsymbol{c}_H$ that solves:*

$$MRS(q_H, \boldsymbol{c}'_H) = \frac{1 - p_H}{p_H} \text{ and } \pi(p_H, \boldsymbol{c}'_H) = 0,$$

and low risk individuals receive the contract $\boldsymbol{c}_L$ that solves

$$V(q_H, \boldsymbol{c}'_H) = V(q_H, \boldsymbol{c}'_L) \text{ and } \pi(p_L, \boldsymbol{c}'_L) = 0.$$

The equilibrium with misperceptions is illustrated in comparison to the equilibrium without misperceptions in figure 2.1. Figrue 2.1 illustrates the resulting equilibrium contracts when the high type makes a downward, or optimistic error, by believing that their risk is lower than in truth: $q_H < p_H$. The eqilibrium contracts with such a misperception are labeled $\boldsymbol{c}'_H, \boldsymbol{c}'_L$ in figure 2.1, without misperceptions are labelled as $\boldsymbol{c}_H, \boldsymbol{c}_L$. Per proposition 1, $\boldsymbol{c}'_H$ is defined as the point of tangency between the zero profit line and the high type's indifference curve IC'_H according to their *perceived* risk q_H. Then $\boldsymbol{c}'_L$ is defined bu the intersection of IC'_H and the zero profit line for

[(3)] When I assume ξ is small (the exact bound is given in the proof), qualitatively I am assuming that the contract defined by the intersection of IC_H and $\pi_L = 0$ is the low type's preferred contract out of all seperating contracts that break even and the high type weakly doesn't prefer. A simplified version of the case where IC_H doesn't bind and the low type can receive a contract they prefer to $\boldsymbol{c}'_L$ without tempting the high types to pretend to be low types is treated in Sandroni and Squintani (2007) and Sandroni and Squintani (2013). I will not pursue it further here.

the low types. This ensures that the high type is indifferent between the two contracts offered in equilibrium, as proposition 1 specifies.

The effects of misperceptions in this model are felt through two channels. First, an individual who misperceives their risk will, directly, have a higher or lower perceived utility from a given contract. In this case, the optimism of the high types means, loosely, they have a lesser demand for insurance. In equilibrium this lesser demand leads to an equilibrium contract featuring less insurance than without a misperception.

Second, an individual who misperceives their risk will find contracts designed for the other type to be more or less attractive. Depending on the incentive constraints, the other types contract may then change in response to this misperception. In this case, the optimisim of the high types leads them to find the original low type contract, $\boldsymbol{c}_L$, more attractive than before. To maintain a seperating equilibrium, this leads to the low types new contract, $\boldsymbol{c}'_L$ being distorted away from full insurance even further. This is because the binding incentive constraint is the high type's. I refer to this effect as the **externality** of a misperception, and in the illustrated case it is a negative externality.

So the misperception by the high type has both a direct and indirect effect on the menu of contracts offered in equilibrium. On the other hand, the low type might themselves have made a misperception, yet per proposition 1 this will have no effect on the equilibrium contract menu. That is, only the misperceptions of the high types have a direct or indirect effect on the equilibrium contract menu, with the externality of that indirect effect flowing downwards and affecting the low types.[(4)]

Proposition 1 establishes that the effect of a misperception on the equilibrium contract menu depends crucially on: Who makes the misperception (high or low type) and what type of misperception is made (optimism or pessimism). An important consequence of proposition 1 is that knowing only the level of aggregate misperception in the population, or knowing the misperception made by the marginal type (in a context of insurance purchase) is not sufficient for welfare calculations. In the following section I will make these consequences precise.

2.1.1 Welfare Consequences of Misperceptions

The introduction of misperceptions into this model of competitive insurance purchase has four key implications for welfare that this section will explore. First, that misperceptions of risk may increase or may decrease welfare, depending on the who makes the misperceptions and in what directions those misperceptions are. Second, that knowing the level of average level of misperception in a population is not welfare-sufficient. Third, knowing the misperception made by (almost) any particular type is not welfare-sufficient. Fourth, to increase insurance coverage by a type of people, sometimes interventions (risk-rating, education, subsidies etc) should be directed to infra-marginal

[(4)] Proposition 1 easily generalizes to multiple types, maintaining the assumption that their perceived risks are ordered in the same way as their true risks. With multiple types, and 'small' misperceptions, only the highest type can have a direct effect on the entire equilibrium, whereas all types but the lowest exert an indirect effect on the contract offered to all lower risk types than themselves.

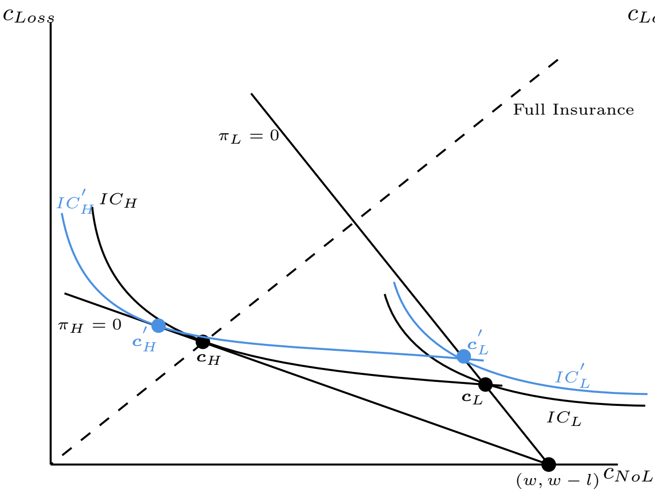


Figure 2: Pessimistic Error by the High Types

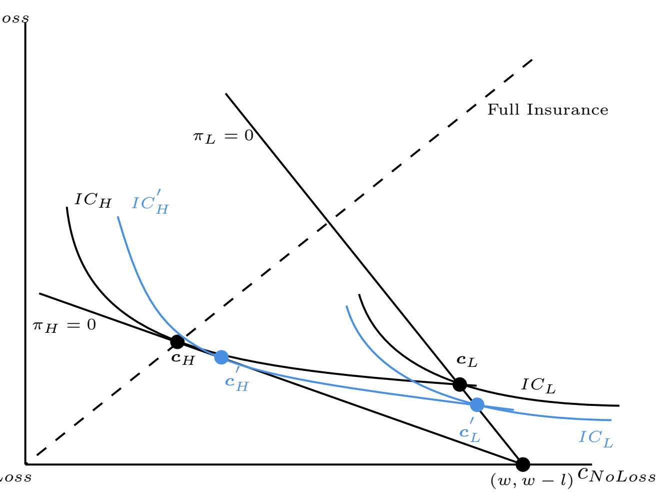


Figure 3: Optimistic Error by the High Types

types.

To make these statements precise, I define a utiliatarian welfare function:

$$W(\boldsymbol{c}_L, \boldsymbol{c}_H) = \rho V(p_H, \boldsymbol{c}_H) + (1-\rho)V(p_L, \boldsymbol{c}_L). \tag{2.1}$$

The parameter $\rho \in [0, 1]$ is the Pareto weight. Higher ρ means the planner weights the experienced utility of the high types relatively more than the low types. Setting $\rho = \alpha_H$, the proportion of high types in the population is natural but not required in the following analysis.

I assume the planner evaluates welfare according to *realized* expected utilities (i.e. according to p, not q). This is justified if a planner cares about utilities that are experienced when the risk is realized (or not). It would not be correct if the planner cared about utility received during a forward-looking evaluation of that risk.

The primary result of this section that drives all others are that some misperceptions impair welfare while others improve it. These two situtations are illustrated in the left and right panels, figures 3 and 2.

In figure 3, the high types make an optimistic misperception. They think the probability of loss is lower than it actually is. As illustrated, the new equilibrium, relative to the pre-misperception, features the high types buying $\boldsymbol{c}'_H$ that offers less insurance than their no-misperception contract $\boldsymbol{c}_H$. But their original contract $\boldsymbol{c}_H$ was first-best for them, and their new contract will deliver a loss of realized utility due to the optimistic error. Due to the binding incentive compatability constraint of the high type, the low type's new contract $\boldsymbol{c}'_L$ also offers less insurance relative to $\boldsymbol{c}_L$, whilst both contract still just break-even. This is even further from full insurance for the low type, and so the high type's error also causes a decrease in the utility that the low type will realize, despite their not making any error.

In figure 2, the high types make a small pessimistic misperception. They think their probability of loss is slightly higher than it truly is. Inversely to figure 1, they are now over-insured relative to their no-misperception first-best contract. This causes a loss of realized utility, but only a second-order loss by the envelope theorem, since along the zero profit line $\boldsymbol{c}_H$ was a global maximum. In contrast, the incentive compatability constraint distorting down the low types coverage has been loosened, such that the new contract $\boldsymbol{c}'_L$ delivers strictly higher utility than $\boldsymbol{c}_L$. Since $\boldsymbol{c}_L$ is away from the first-best, this improvement is first order. Overall, a first order improvement in the low type's utility outweighs a second order decline in the high type's utility, leading to an overall welfare gain.

Hence, there exist misperceptions that strictly increase, strictly decrease, or have no effect at all on, the overall welfare in the market. The above discussion is summarized in the following proposition.

Proposition 2. *Suppose the high types perceived risk is $q_H = p_H + \xi$. Then $\frac{\partial W}{\partial \xi} > 0$ when evaluated at $(\boldsymbol{c}_L, \boldsymbol{c}_H)$, the undistorted contracts. In particular, welfare increases for small positive ξ, and decreases for negative ξ.*

In contrast to the effects, positive or negative, of errors by the high type, any misperceptions by the low types have no effect at all on the equilibrium menu of contracts, nor on welfare. This is because the low types contract is pinned down by a zero-profit condition according to true probabilities, and the incentive compatability constraint of the *high* type. Neither of these are affected by a misperception by the low type. This observation, in addition to proposition 2, makes the following immediate. Define the average misperception in the population as

$$\bar{\xi} = \alpha_H \xi_H + \alpha_L \xi_L$$

where $\xi_H = q_H - p_H, \xi_L = q_L - p_L$.

Corollary 1. *For any given (small) $\bar{\xi}$ there exist (ξ_L, ξ_H) and (ξ'_L, ξ'_H), both of which generate an average error of $\bar{\xi}$, but with the former causing a welfare loss, and the latter a welfare gain, relative to $\boldsymbol{\xi} = \mathbf{0}$.*

There are important implications of this result. Some studies[5] document misperceptions that occur on average in the population. Statements about welfare are then typically made with reference to these average misperceptions. The result above shows that such welfare implications are possibly wrong, and definitely unfounded. The welfare implications depend on, at least, which risk class makes which type of error, and analysis that does not take this into account might be missing an important channel.

Were this analysis to be embedded in a larger framework of demand for a particular type of insurance, one could define which risk-type was marginal. If the marginal type is found to be optimistic (hence under-demanding insurance), one might suggest an intervention be targeted at the marginal type. For example, an educational or informational campaign that aims to better

[5] For example, O'Dea and Sturrock (2020)

inform the marginal type as to their true risk. The analysis above shows that this targeting of an intervention might be misguided. In particular, should the marginal type be a low(er) type, no amount of information provision to that type will affect the insurance they purchase in equilibrium. This is because, for all but the highest type, the insurance purchased in equilibrium depends on the misperceived risk *only of the higher types.* So if high and low types are both making a misperception, and a planner corrects the latter but not the former, then the contract purchased by the latter in equilibrium will not change.

The general implication, that non-local or non-marginal interventions are sometimes necessary to for the contract and utility of the marginal types to change, is summarized in the following.

Corollary 2. *Suppose a planner can correct the misperception of some types. The equilibrium contract received by the low types will change if and only if the misperception of the high type is changed.*

2.2 The Planner's Problem

To be sure that the conclusions drawn from the previous section are not overly sensitive to the particularities of the Rothschild and Stiglitz (1976) setup, in this section I study the planner's problem. I will show that the patterns obtained identically track the Rothschild and Stiglitz (1976) style formulation from section 2.1.

I continue to assume that the planner has preferences as in (2.1) and can implement any incentive compatible and budget-balanced menu of contracts. Their problem is then:

$$\max_{\boldsymbol{c}_H,\boldsymbol{c}_L} \rho V(p_H,\boldsymbol{c}_H) + (1-\rho)V(p_L,\boldsymbol{c}_L) \tag{2.2}$$

$$\text{such that} \tag{2.3}$$

$$V(q_H,\boldsymbol{c}_H) \geq V(q_H,\boldsymbol{c}_H) \tag{2.4}$$

$$V(q_L,\boldsymbol{c}_L) \geq V(q_L,\boldsymbol{c}_H) \tag{2.5}$$

$$\alpha_H \pi(p_H,\boldsymbol{c}_H) + \alpha_L \pi(p_L,\boldsymbol{c}_L) = 0 \tag{2.6}$$

$$(\boldsymbol{c}_H,\boldsymbol{c}_L) \in A = \left\{ \boldsymbol{c}_H,\boldsymbol{c}_L : MRS_H \leq \frac{1-p_H}{p_H} \wedge c_L^{NL} \geq c_L^L \right\}.. \tag{2.7}$$

The usual restriction on the contract space is that $c_{NL} \geq c_L$ for both types and is called an indemnity constraint (as in Netzer and Scheuer (2014)) such that the loss cannot make an individual better off. I wish to keep this in spirit - not allowing individuals to harm themselves through perceived over-insurance - but I do not want to shut down the wedge between subjective and objectively defined optimal insurance. So I define the set of acceptable contracts, A, such that $\boldsymbol{c}$ can only allow for small over-insurance for the high type, and I will assume that errors are small, such that the only over-insurance that can be bought is that which improves perceived utility for the high type.

This rules out equilibria where, according to the misperceived risk, a type could be made better off with a profit-neutral reduction of insurance. If no errors are made, $\boldsymbol{q} = \boldsymbol{p}$, this reduces to the standard indemnity constraint. If one omits this restriction, uniqueness of the equilibria that follow might be lost, and the new equilibria will be qualitatively similar but possibly inverted.

Equations (2.4) and (2.5) are incentive compatability constraints for the high and low risk types respectively, and equation (2.6) is an aggregate resource constraint which allows for cross-subsidization between types, unlike in the competitive equilibrium in which each contract individually broke even. I re-emphasize that whilst incentive constraints are evaluated according to subjectively perceived probabilities q_i, welfare and the resource constraint are evaluated according to objective probabilities p_i.

The first proposition characterizes the constraints that bind at the planners optimum.

Proposition 3. *At the planner's optimum, constraint (2.4) binds, while (2.5) is slack. Constraint (2.7) binds in that* $MRS_H = \frac{1-p_H}{p_H}$ *at the optimum.*

This demonstrates that the qualitative structure of the planners solution mirrors the competitive equilibrium. The high type will receive full insurance or infinitesimally more than full insurance if they make an upward error. The low type will receive partial insurance constrained and distorted by the high type's incentive compatibility constraint.

Initially the contract space has four free parameters. Proposition 3 shows that there is at most on degree of freedom left. It will be helpful to think in terms of the cross-subsidization from low types to high types, so I define:

$$\chi = p_H c_H^L + (1 - p_H) c_H^{NL} - (w - l p_H) \geq 0.$$

Given a level of cross-subsidization χ, this defines the profit/loss level to be earned from each type's contract. In combination with the binding part of A, this defines the high type contract. The incentive constraint of the high type then fixes the low type's contract. By this logic thinking in terms of χ as the final free-parameter is valid and parsimonious.

The next lemma relates the Pareto weight ρ to the degree of cross-subsidization χ that prevails in the optimum.

Lemma 1. *Suppose there are two planners with different Pareto weights:* $\rho_1 > \rho_2$. *Then at the respective optima,* $\chi^*(\rho_1) \geq \chi^*(\rho_2)$, *holding strictly when* $\boldsymbol{c}_H \neq \boldsymbol{c}_L$.

When marginally more weight is placed on the high types utility, higher cross-subsidization from the low types to the high types is optimal for the planner. This result is intuitive, but is important to keep in mind for the slightly less intuitive results in the next proposition. The next propostion, the main one of this section, establishes that the qualitative features of the equilibrium in section 2.1 remain true for the planners problem.

Proposition 4. *Fix any* ρ. *At the planner's optimum the following hold for small enough* $\xi > 0$*:*

- *For upward misperceptions by the high types $q_H = p_H + \xi$, welfare increases and cross-subsidization decreases relative to no misperceptions.*
- *For downward misperceptions by the high types $q_H = p_H - \xi$, welfare decreases and cross-subsidization increases relative to no misperceptions.*
- *Misperceptions by the low types have no impact on the optimum.*

As ρ varies the entire range of constrained Pareto efficient outcomes are selected by the planner. The proposition above shows that for any particular ρ, welfare increases (decreases) but cross-subsidization decreases (increases) when there is an upward (downward) misperception is made by the high type. Intuitively, this is because the set of implementable contracts has strictly gotten bigger (smaller) due to the misperception loosening (tighetening) the high types IC constraint. The change to the IC constraint allows for welfare to strictly rise or fall. Similarly to the RS equilibrium, because the low type's IC constraint never binds at the optimum, misperceptions by them have no effect on the optimum or on welfare.

To understand the cross-subsidization result, note that without misperceptions the optimum features a net zero welfare change if one more dollar of subsidy were to go from the low types to the high types. The low types would strictly be worse off, the high types better off, but at the optimum these cancel out. Now with an upward misperception, the incentive constraint is loosened and at the previous level of cross-subsidization the low type is better off. But because the low type is better off and closer to full insurance, the marginal dollar of cross-subsidization is even less desirable to them given their relatively improved contract. The marginal dollar of cross-subsidization is, to the first order, of the same value to the high type as before the error. So the net value of the marginal dollar of cross-subsidization is strictly lower than without misperceptions, and hence less cross-subsidization prevails in equilibrium.

2.2.1 The Miyazaki-Wilson allocation as a particular constrained efficient allocation

Issues with the existence of the Rothschild-Stiglitz competitive equilibrium have lead to the use of alternative equilibrium notions. A leading alternative is that of Wilson (1977), Miyazaki (1977) and later Spence (1978). Since Miyazaki (1977) the MW allocation has been thought of as a particular solution to the planner's problem, micro-founded by the restricted deviation set in Wilson (1977).

In this context, with misperceptions, the game-theoretic logic of Wilson (1977) and the constrained efficient allocation of Miyazaki (1977), Spence (1978), Netzer and Scheuer (2014) and others are not coincident. In this section, I study the latter interpertation, which is more persuasive and interesting. In appendix A.2 I consider the game-theoretic interpretation more in line with Wilson (1977).

For now, write $(\boldsymbol{c}_H^{RS}, \boldsymbol{c}_L^{RS})$ for the allocations that prevail in the Rothschild-Stiglitz equilibrium in section 2.1. Then define an expanded program (2.3)* as (2.3) subject to (2.4) through (2.7) with the constraint $V(q_H, \boldsymbol{c}_H) \geq V(q_H, \boldsymbol{c}_H^{RS})$ appended. That is, (2.3)* is program (2.3) with the extra requirement that the high types do no worse than the competitive equilibrium. This is defined by analogy to Netzer and Scheuer (2014). We can then write:

Definition 2. *Denote the allocation* $(\boldsymbol{c}_H, \boldsymbol{c}_L)$ *that solves* (2.3)* *as the MW allocation.*

The following then follows directly from the preceding results of this section.

Corollary 3. *The MW allocation has the following properties:*

1. *The constraints bind as in Proposition 3.*
2. *(Small) upward misperceptions by the high type increase welfare, downward misperceptions decrease welfare. Misperceptions by the low type do not affect welfare.*
3. *Amongst the set of constrained efficient allocations with positive cross-subsidization, the MW allocation features the minimal level of cross-subsidization.*

2.3 Discussion

The conclusion that prevails throughout these multiple market structures and equilibrium assumptions is that covariance between risk and risk misperception is central to judge the welfare implications of risk misperceptions. This is because contracts endogenously response to misperceptions. This is in contrast to much of the literature, in which WTP for an insurance contract is affected by misperceptions, but the insurance contract itself is not. Consequently, if the set of contracts offered is held fixed when miserpceptions are considered, an important mechanism is missed. Misperceptions in risk affect not only who buys an insurance contract but what that insurance contract looks like.

The analysis shows that how individual misperceptions map to their equilibrium contract is not straightforward. Errors by higher risk types propagate downwards to affect all other contracts. This externality shows that to determine the welfare of those buying a particular contract it is necessary to consider the misperceptions of those buying the particular contract and all those buying more generous contracts.

Further, the welfare implications of distortions in demand due to risk misperception are ambiguous. Some misperceptions lead to welfare improvements, others lead to welfare declines, and others still have no effect at all. It is an empirical question which misperceptions have positive or negative welfare consequences, and hence which might bebest corrected or left alone by a welfare maximizing planner. I analyze multiple risks to demonstrate the range of welfare conclusions and patterns of misperception that exist.

To preview, those findings will span the set of possible outcomes. For disability risk, misperceptions are minimal. For catastrophic medical risk, the high risk types are accurate in their beliefs, and the lower risk types over-perceive their risk, implying an unambiguous welfare improvement. For long-term care and mortality risk, the high risk types under-perceive their risk, which reduces their welfare and those further down the risk distribution.

3 Empirical Analysis

3.1 Ideal Data and Quantities of Interest

The theoretical analysis in section 2 generates the novel insight that it is not sufficient to measure the misperception averaged over the whole population. I need to know, for each true risk type p_i, what their perceived risk q_i is.

In this section I study a variety of risks for which a measure of objective and subjective risk are attainable. I consider mortality risk (death by the age of 75), disability risk (health-limiting impairment or health problem), long-term care risk (nursing home stay) and catastrophic medical risk (out-of-pocket costs above \$ 8000).

The first key datum is a measure of an individuals perception of their risk, q_i. Ideally, I would observe this perception as relevant to an actual insurance purchase. Unfortunately, the best I can do is to observe this quantity in this setting of a survey, albeit a long-established and reputable survey. An issue with this datum is that there might be 'elicitation error' due only to the survey that wouldn't be present in an actual insurance purchase. I develop methods to test for robustness to this.

Secondly, I need a measure of the individuals true propensity for the risk to occur, p_i. I need a measure of the true probability of death by age 75, or of a work-limiting health problem and so on. This propensity needs to be predicted such that the conditional distribution of $q_i \mid p_i$ can be analyzed. While any one person might have personal information that does not enter my prediction, comparing my (unbiased) objective predictions for a group to their subjective perceptions of that risk will answer the question: Does that group have *on average* correct beliefs about their risk.

To make that prediction, I need a set of risk-relevant covariates. All of these risks are health and lifestyle related, and so natural candidates include medical history, measures of day-to-day functionality, longevity information about the individuals' parents and other close family, employment history, income and wealth data. Then I need to know how this risk was realized for at least some of the population of interest. More details are given in section 3.3. By splitting the sample into a test set and training set, valid out-of-sample probabilities of risk can be retroactively predicted. These probabilities will be (my noisy estimate of) p_i. They can then be compared to subjective assessments the probability, q_i, to do inference on moments of the distribution $q_i \mid p_i$.

3.2 Data

The data come from the Health and Retirement Study (HRS), a biennial panel survey of Americans who are (mostly) older than 50 years. It contains the three vital ingredients described above.

Subjective Perception of Risk

1. Mortality risk: The subjective probability of mortality by age 75 is elicited from all respondents aged 65 and younger. The survey asks: *"What is the percent chance that you will live to be 75 or more"*. This will form the raw estimate of q_i.

2. Long-term care risk: All respondents aged 65 and older are asked *"What is the percent chance that you will move to a nursing home in the next five years?"*
3. Disability risk: All respondents in waves 1 through 6 are asked: *"What about the chances that your health will limit your work activity during the next 10 years?"*
4. Catastrophic medical risk: from wave 10 onwards all respondents are asked *[Regarding health insurance, excluding premiums] what are the chances that you will spend more than $ 8,000"* during the coming year?[6]

Realization of Risk

1. Mortality risk: The outcome of the risk of death by age 75 is known for approximately half of the respondents. Those who attrited[7] before age 75 or who are still younger than age 75 are excluded. Those that remain are assigned a death indicator $D_i = 1$ is they died before age 75 and $D_i = 0$ if not.
2. Long-term care risk: The outcome is coded as 1 for those that entered a nursing home for an overnight stay in the past 5 years and for whom there are three waves of surveys after the subjective perception was elicited so that the outcome is observable. Those that have not had an overnight stay in a nursing home but for whom the outcome observable are coded as 0, and anyone who has not responded to three more survey waves is excluded.
3. Disability risk: Those that respond yes in the next 10 years to the question "Do you have any impairment or health problem that limits the kind or amount of paid work you can do?" are coded as 1, and those who always answer no are coded as 0. Those for whom 10 years of responses are not recorded are excluded.
4. Catastrophic medical risk: out-of-pocket medical expenses in each wave are computed as the sum of out-of-pocket hospital costs, nursing home costs, doctor visit costs, dentist costs, outpatient surgery costs, monthly prescription costs, home health care costs and special facilities costs. These are surveyed and measured for the last two years, and I assume these are equally likely to have occured in each of the two years to cohere with the subjective elicitation which asks about the next 12 months.

[6] I restrict the sample to all those who had only public insurance (Medicare or Medicaid) in the wave in which the subjective perception was elicited and the wave afterwards in which the risk outcome was observed. This is so risk being predicted and that being realized are about the same insurance contract. This neglects the fact that those with high projected out-of-pocket costs may select into a private insurance plan due to the anticipated costs. However, the patterns that will obtain - pessimism about the risk amongst all risk types - would be increased if I included those who select into more insurance when they perceive a high change of out-of-pocket costs. It is unlikely that those who under-perceive the chance of catastrophic out-of-pocket costs are those selecting into more insurance and hence those excluded by my sample restriction.

[7] The survey has run since 1992, with relatively low attrition.See, for example, Banks et al. (2011), who finds low attrition and no correlation between attrition and major medical conditions.

The survey collects a rich variety of health data, employment data, family history and other data relevant to the prediction of these risks. The entire set used in the prediction algorithm is specified in appendix C.1. The prediction algorithm is described in the next section.

Data from all waves of the HRS are potentially included in the prediction and inference, subject to some questions being asked only in a subset of waves. Prediction is done at the individual-wave level, so that a respondent's responses as a 55 year old might be included while their responses as a 63 year old might not. The two essential requirements are: The individual-wave responses must be early enough so that the risk can be definitely observed. In the case of mortality risk this means they must have been born early enough that they would be 75 or older by the final wave. For the other risks it means sufficiently many more waves of the survey must have occured (and been responded to) so that the time horizon relevant to the risk (12 months, 5 years or 10 years) has been observed.

Moreover, the covariates that are used by the prediction routine must be present. The latter restriction rules out waves 1-3 in which a different and less comprehensive set of medical information was collected, and everyone from waves 4 onwards with some of these data missing[8]. Although a different set of variables will be used for supplementary analysis for mortality risk in wave 1, which is explored in section 3.5.1. Prediction is done in long form, so that an individual at different ages may appear as multiple observations. To control for unobserved individual level heterogeneity the standard errors are clustered by individual, and in the panel specification individual fixed effects are differenced out.

The HRS has some advantages and disadvantages. The primary advantage is the availability of the three key classes of data above. Few surveys collect data on subjective elicitations of this many short and medium run risks. Fewer still have gone on for long enough that the realization of the longer term risks, such as mortality by 75, are known. Whilst some progress could be made using, for example, hazard modelling if only short-run mortality was known, a distinction needs to be made between subjective elicitations with high short run accuracy (e.g. if an individual has a terminal disease their prediction is likely to be very accurate and resolved soon) and high long run accuracy. This is to my knowledge, one of the first studies to assess the accuracy of long-run mortality predictions over the entire duration to which the prediction pertains. For no other reason than the lack of time, this has not been possible until quite recently.

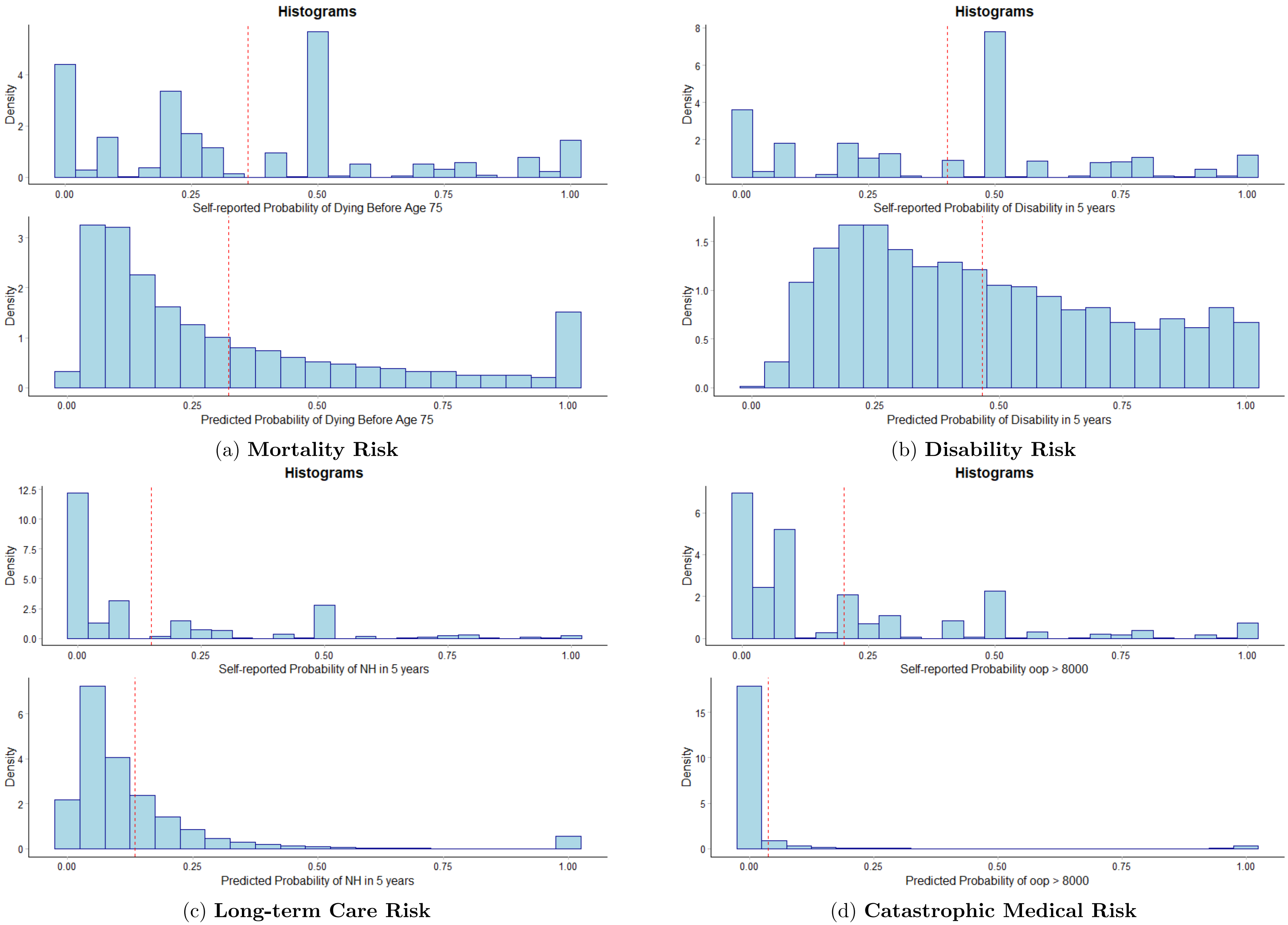
A key disadvantage is the lack of incentivization for giving accurate or thoughtful answers. There are some checks in place to make sure the respondents are not answering randomly, but the survey is still cheap talk. It is this uncertainty as to how well the elicitations from the HRS correspond to true private estimates of risk that presents an empirical problem in the analysis and to which I will return later.

[8] See appendix C for the full list of variables used the prediction routine

3.3 Prediction of True Risk

Since the true probability of these risks are epistemically unobservable researcher the best I can do is make a prediction $\hat{p}_i$ that has favourable properties for inference. For each risk, I begin by taking the set of individuals for whom the risk is definitely realized or not. I randomly split them into a training set and a test set with equal probability. This is done on the individual level, not the individual-wave level. I train various models to predict the risk realization as a function of covariates. Inference is then performed on all the observations that were left out of the training set, i.e. the test set. Sample-splitting of this form is needed to avoid overfitting.

My preferred prediction routine is a LASSO logit model with over 400 variables and interactions that seemed, ex-ante, relevant to mortality prediction. I will refer to this as the 'logit' or 'lasso logit' model. I also compared this to a logit model run on all 400 variables. The details are in appendix D but qualitatively the conclusions are identical.



(a) **Mortality Risk**

(b) **Disability Risk**

(c) **Long-term Care Risk**

(d) **Catastrophic Medical Risk**

Figure 5: Histograms of self-reported and predicted probabilities for each risk.
Note: Throughout only the leave-out 'training' sample are included.

Figure 5 show histograms of the self-reported and lasso logit predicted probabilities of each of the four risks.

The averages (the red vertical dashed lines) of the self-reports and the predictions are quite close for mortality risk and long-term care risk, while there is overall pessimism for catastrophic medical risk and optimism for disability risk. However, for all risks, the self-reports display much more bunching than the predictions, due to the propensity of survey respondents to round to focal numbers such as 0%, 50%, 100% as well as multiples of 10 or 20%. Dealing with the bias potentially introduced by this rounding is critical.

Summary statistics for 15 variables with the most predictive power for mortality (in the sense of reducing the mean-squared prediction error of the training set) are provided in table 1. The mean and standard deviation are reported for those for whom the various risks are realized and those for which they are not. The variables are listed in order of their predictive power for mortality risk.

Being a current or former smoker, a measure of assistance needed with daily mobility activities, self health assessments as well as various major diseases are most predictive of all the risks. Before proceeding to do inference it is critical to take into account individuals' rounding of the subjective perceptions of risk. The following section describes this procedure.

Variable	Survive to 75	Death before 75	No Disability	Disability	No LTC	LTC	< 8000 OOP	≥ \$8000 OOP
	0.18	0.39	0.14	0.18	0.10	0.08	0.14	0.10
Smoker (current)	(0.38)	(0.49)	(0.35)	(0.38)	(0.30)	(0.27)	(0.35)	(0.31)
	0.67	1.45	0.24	1.19	1.15	2.33	1.09	2.14
Assistance needed for daily mobility tasks$^{(9)}$	(1.11)	(1.61)	(0.58)	(1.41)	(1.45)	(1.84)	(1.47)	(1.88)
	0.09	0.25	0.07	0.16	0.21	0.26	0.19	0.25
Diabetes (ever diagnosed)	(0.29)	(0.43)	(0.25)	(0.37)	(0.40)	(0.44)	(0.39)	(0.43)
	2.50	3.28	2.13	3.01	2.91	3.45	2.85	3.47
Self-health Rating (1=Excellent, 5=Poor)	(1.09)	(1.19)	(0.89)	(1.08)	(1.10)	(1.12)	(1.11)	(1.14)
	0.57	0.45	0.54	0.57	0.56	0.66	0.57	0.64
Female	(0.5)	(0.5)	(0.50)	(0.50)	(0.50)	(0.47)	(0.49)	(0.48)
	0.59	0.77	0.54	0.60	0.58	0.56	0.57	0.55
Smoker (ever)	(0.49)	(0.42)	(0.50)	(0.49)	(0.49)	(0.50)	(0.49)	(0.5)
	0.10	0.25	0.08	0.20	0.28	0.40	0.22	0.37
Heart Disease	(0.30)	(0.43)	(0.26)	(0.40)	(0.45)	(0.49)	(0.41)	(0.48)
	71.2	69.8	71.7	71.1	71.9	71.1	71.7	71.8
Father's age (current or age of death)	(14.4)	(1431)	(14.1)	(14.3)	(14.8)	(14.7)	(14.4)	(14.2)
	0.19	0.31	0.32	0.56	0.95	0.98	0.62	0.73
Covered by any government health insurance	(0.39)	(0.46)	(0.47)	(0.5)	(0.23)	(0.14)	(0.49)	(0.44)
	6.5	11.9	5.2	10.1	9.9	14.7	9.7	20.0
Doctor Visits (past year)	(9.8)	(27.5)	(7.4)	(17.7)	(17.2)	(27.6)	(18.3)	(40.1)
	0.10	0.29	0.02	0.20	0.20	0.63	0.20	0.60
Assistance needed for fine-motor tasks$^{(10)}$	(0.36)	(0.63)	(0.16)	(0.50)	(0.53)	(0.93)	(0.54)	(0.94)
	0.02	0.07	0.01	0.05	0.08	0.26	0.07	0.21
Home health care utilized (past two years)	(0.13)	(0.25)	(0.1)	(0.22)	(0.27)	(0.44)	(0.25)	(0.40)
	1.02	1.51	0.51	1.54	1.30	1.99	1.27	1.93
Assistance needed for large-muscle tasks$^{(11)}$	(1.27)	(1.42)	(0.85)	(1.23)	(1.29)	(1.32)	(1.33)	(1.40)
	0.02	0.08	0.01	0.06	0.09	0.22	0.08	0.19
Stroke (ever)	(0.15)	(0.27)	(0.12)	(0.24)	(0.29)	(0.42)	(0.27)	(0.39)
	0.14	0.30	0.11	0.24	0.27	0.52	0.25	0.52
Overnight hospital stay (last two years)	(0.35)	(0.46)	(0.31)	(0.43)	(0.44)	(0.50)	(0.43)	(0.50)
	0.06	0.12	0.07	0.10	0.17	0.19	0.13	0.20
Cancer (ever)	(0.23)	(0.32)	(0.25)	(0.30)	(0.37)	(0.39)	(0.33)	(0.40)
Sample Size								
Observations (Ind × wave)	25,027	10,940	161,780	161,191	275,151	42,846	118,865	3,552

Table 1: Summary statistics for the most (mortality) predictive variables. Note: Means are reported, and standard deviations are in parentheses.

3.4 Accounting for Rounding Error

When individuals are faced with an HRS interviewer and asked to consider their mortality prospects, there are broadly two types of errors they might make. One is a genuine misperception, of the kind I am interested in throughout this paper. They might truly think their risk prospects are better or worse than they actually are, and this misperception would persist even in a perfect elicitation of their belief. These I will call **demand-relevant errors**. On the other hand, the response given in the HRS might be infected with what I will call **elicitation noise** or **elicitation error**. By this I mean a discrepancy between the individuals demand-relevant perception of their risk and their reported perception of their risk that is due only to the artefact of the survey. In particular, there might be errors made by the individual only because the survey is unincentivized, or the individual is tired and so rounds their answer, or because no actual insurance purchase is being made. If such errors could be corrected, hypothetically, if we observed the individual's perception in a demand-relevant high stakes scenario, then I deem such errors merely elicitation noise.

The challenge is that I only want to do inference on the demand-relevant misperceptions after filtering out the elicitation noise (primarily rounding) but these quantities cannot be disentangled. Only the sum is observed by looking at the difference between an individual's predicted risk and their elicited perception of that risk from the HRS. A

I will attempt to overcome this in multiple ways. First, from the individuals elicitation I will classify them into a rounding category, and then attribute as much of the observed misperception to elicitation error as is consistent with the rounding category. I will do this in an adverserial (to my conclusions) way and then argue that the residual (conservatively small) true misperception still generates the same qualitative patterns as the observed misperception. Second, I will leverage a change in scale on which the probabilities were elicited in the first wave of the HRS. In that wave, the unit to which everyone is rounding is unambiguous and so there will be much less uncertainty that the rounding category is correct. This check can only be performed for mortality risk as the other subjective elicitations were not surveyed in the first wave. Third, I will study within person variation across time, so that any individually fixed elicitation errors are removed, as well as applying the rounding procedure to be described in detail, so that plausibly only time varying true-misperception remains. Using all these methods similar qualitative conclusions will obtain.

Even though actual insurance purchases are available in the data for some insurance markets (e.g. life insurance), I will not use them to infer private beliefs. This is because insurance markets are already distorted in multiple ways such that private perceptions of risk cannot be easily extracted from demand information. For example, in the life insurance market there is a litany of health conditions that would lead to rejection if one tried to buy insurance. As such, life insurance holding

[9] Count 0-5 of tasks for which assistance is needed. The tasks are: Walking several blocks, walking one block, walking across the room, climbing several flights of stairs and climbing one flight of stairs.

[10] Count 0-3 of tasks for which assistance is needed. The tasks are: Picking up a dime, eating, and dressing.

[11] Count 0-4 of tasks for which assistance is needed. The tasks are: Sitting for two hours, getting up from a chair, stooping or kneeling or crouching, pushing or pulling a large object

is *negatively* correlated with elicited mortality risk perception, contrary to intuition.[12] Put another way, the insurance markets are in a second-best state due to many factors. To isolate the impact of the one factor under study - misperceptions in risk - we cannot look at the current equilibrium and isolate which distortion is due to this factor and which is due to others on which we do not have data.

The de-rounding procedure works as follows. As can be seen in the top panel of figure **??**, many of the self-reports are multiples of 50%, 20% or 10%. I assume that a report of 40% means that the individual has rounded to the nearest 20%. So even if the 40% represents rounding due to impatience or inability to narrow it down more precisely, I can be sure the true subjective belief, if perfectly reported, lies between 30% and 50%. To adjust the 40% report, I consider what the predicted risk for that individual is. If the individuals predicted risk is $\hat{p}_i \in [30\%, 50\%]$, then I set their "de-rounded elicitation" $\tilde{q}_i$ to be $\tilde{q}_i = \hat{p}_i$ such that, after accounting for rounding in this way, their true misperception $\tilde{q}_i - \hat{p}_i$ is zero. Alternatively, if their predicted risk is $\hat{p} < 30\%$ then I set $\tilde{q}_i = 30\%$, and if their (predicted) true risk is $\hat{p} > 50\%$ then I set $\tilde{q}_i = 50\%$.

In words: If the gap between an individuals raw elicitation and their predicted risk is fully explainable by rounding error, then I assume their demand-relevant misperception is zero. If the gap between an individuals raw elicitation and their (predicted) true risk is not fully explainable by rounding error, then I attribute as much as consistent with their rounding category to be due to rounding, and call the remainder (a conservative bound on) true misperception.

I categorize individuals into rounding categories as follows. If an individual is designated an $x\%$ rounder, then this means I allow for up to $x/2\%$ error in each direction, as in the example above. I categorize:

- If $q_i \in \{0, 50\%, 100\%\}$, then I assume a 50% rounder.
- If $q_i \in \{25\%, 75\%\}$, then I assume a 25% rounder.
- If $q_i \in \{10\%, 20\%, 30\%, 40\%, 60\%, 70\%, 80\%, 90\%\}$, then I assume a 20% rounder.
- If $q_i \in \{5\%, 15\%, \ldots, \%95\}$, then I assume a 10% rounder.
- Everyone else, I assume no rounding.

According to this procedure the de-rounded perceptions $\tilde{q}_i$ are computed. In the next section inference will be performed as if either subjective beliefs are q_i or as if they are $\tilde{q}_i$. Probably the former will understate the elicitation error and the latter will overstate it, with the true private beliefs lying somewhere in between. But what we will find is that the qualitative conclusions hold in both cases, and thus would be true if we could observe the demand-relevant subjective perceptions devoid of any elicitation noise.

[12] Note though that conditional on not having a rejectable condition - life insurance holding is correlated with perception of risk, indicating that there is demand-relevant private information in the HRS elicitation. See, for example, Hendren (2013).

3.5 Results

Figure 5 shows that average misperception is small for most risks. I interrogate how this misperception is distributed over the risk distribution. Is everyone accurate on average, or uniformly pessimistic or optimistic, or is there some variation uncorrelated with risk, or is there a clear relationship between risk and risk misperception?

For each risk, I begin with the following two OLS specifications:

$$q_i = \beta \hat{p}_i^{Logit} + \epsilon_i, \tag{3.1}$$

$$\tilde{q}_i = \beta \hat{p}_i^{Logit} + \epsilon_i. \tag{3.2}$$

The dependent variable is either the raw elicitation of subjective risk q_i or the de-rounded version of that elicitation $\tilde{q}_i$. The independent variable is the logit predicted true risk. The quantities of interest are $E(q_i \mid p_i = 0)$, $E(q_i \mid p_i = 1)$ and $\beta \approx \frac{\partial q_i}{\partial p_i}$, as well as the analogues for $\tilde{q}$. These are the average error at the top of the risk distribution, the average error at the bottom of the risk distribution, and the slope of the conditional mean, interpretable as the partial effect on the subjective elicitation of a change in the predicted true risk.

That the true risk p_i is not observed and at best proxied by the prediction $\hat{p}_i$ introduces Berkson measurement error into the specification. However, OLS will still deliver unbiased estimates of the quantities of interest. This is because the prediction error $\eta_i = p_i - \hat{p}_i$ will be orthogonal to the prediction, $\hat{p} \perp \eta$, unlike in the classical measurement error setting in which $p \perp \eta$. [13] Moreover, since the self-report q is included in the prediction of p, we have $E(\eta q) = 0$.

Hence, by using $E(\eta q) = E(\eta \hat{p}) = 0$ mechanically the slope coefficient from the OLS specification will satisfy

$$\hat{\beta} = \frac{Cov(q, \hat{p})}{Var(\hat{p})} = \frac{Cov(q, p - \eta)}{Var(p - \eta)} = \frac{Cov(q, p)}{Var(p)},$$

showing that this measurement error due to prediction does not cause bias in the slope coefficient nor in the intercept.

The estimates for these three quantities for each of the two specifications are in table 2. Standard errors are clustered by respondent throughout and only the test set is used for inference.

The results are visually illustrated in figure 7. For each of the risks, the horizontal axis is the prediction probability of that risk occuring. Individuals are binned according to that predicted probability and for each bin the self-reported probabilities (blue), the self-reported probabilities corrected for rounding (red) and the ex-post risk outcomes (black) are computed and plotted.

The better the prediction routine, the closer the black dots should be to the dashed 45-degree line. If the black dots are coincident with the 45 degree line this would mean the predicted probabilities are on average equal to the ex-post outcomes in the test set. As you can see, this is broadly true, although with worse performance at the tails and for certain risks.

[13] Figure **??** provides evidence for this orthogonality, as the difference between predicted probabilities and true frequencies is zero on average and uncorrelated with the prediction probabilities, as it is zero at all predicted values.

The most conservative estimate of risk misperception and its relationship to risk is found by comparing the red dots (self-reports corrected for elicitation error) with the black dots (ex-post risk outcomes). I summarize the patterns below.

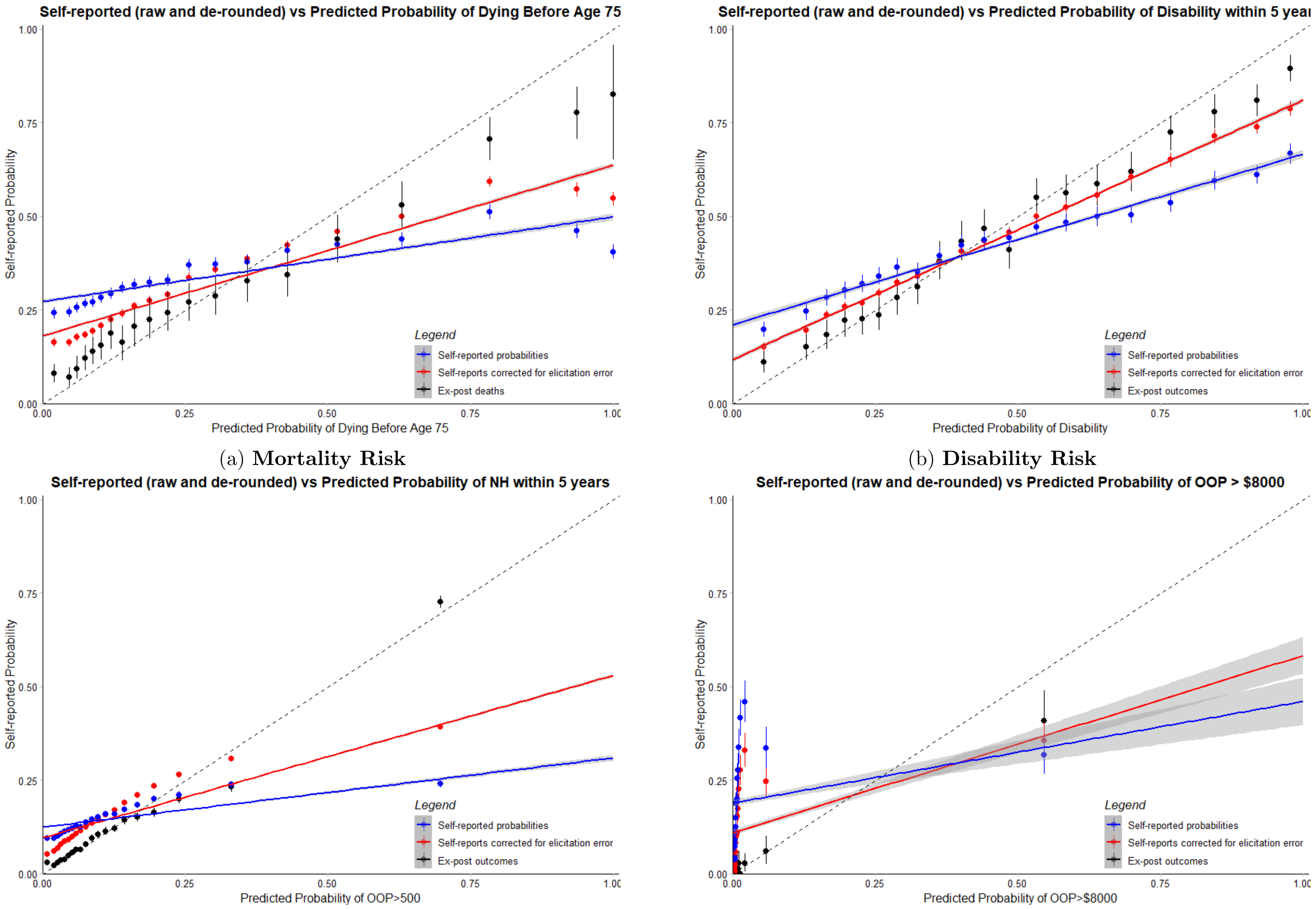
1. Mortality risk: The misperceptions exhibit an 'S' shape. The low risk individuals are pessimistic and overestimate their risk. The high risk indivduals are optimistic and underestimate their risk. The middle 60% of the risk distribution have accurate perceptions, or at least I cannot reject the hypothesis that their misperceptions are entirely due to rounding noise due to the survey. Per the theory, this suggests that mortality (e.g. life insurance) and longevity (e.g. life annuities) risk markets have under-demand for the high risk types, approximately correct demand for the middle risk types, and over-demand for the low risk-types.

2. Long-term care risk: The misperceptions also exhibit an 'S' shape, although the magnitude of pessimism for the low risk types is very low. Those over-perceptions are very small relative to the under-perceptions of the highest risk types, the top 5% of whom who think they are up to 50% less risky on average than in actuality. These patterns point to a severe downward distortion in the demand for those who should have the highest WTP for long-term care insurance and close to accurate demand the the bottom 80% or so of the market.

3. Disability risk: Individuals' perceptions of their disability risk is broadly accurate. Since the risk is short-term (within 5 years) and perhaps often progressive (e.g. someone might know their back pain has been degenerating for a while), it is unsurprising that people have almost correct perceptions. In fact, once rounding error is allowed for, there is essentially no difference between perceptions and average ex-post outcomes. This demonstrates the importance of accounting for rounding noise in the context of an unincentivized survey such as the HRS.

4. Catastrophic medical risk: All risk types are pessimistic or accurate in their perceptions of catastrophic out-of-pocket costs. This suggests that individuals demand for the most generous health insurance contracts is higher than their actual risk would suggest. This pattern is in contrast to individuals undervalueing health insurance relative to their risk [14] that is often observed in the literature. This highlights the need to separate the analysis based on different risks (catastrophic health risk vs moderate health risk) as well as different parts of the risk distribution. Here we find that when it comes to catastrophic risk, relevant to the sickest part of the population, there is pessimism.

[14] See, for example, Finkelstein et al. (2019).



(a) **Mortality Risk**

(b) **Disability Risk**

(c) **Long-term Care Risk**

(d) **Catastrophic Medical Risk**

Figure 7: Predicted probability of each risk versus self-reported probability and ex-post risk realizations
Note: Only those in the training set are included. Individuals are sorted by their predicted probability, then divided into 20 evenly sized bins. In each bin the average self-report and average risk realization rate, with standard errors, are graphed.

	Mortality Risk		Disability Risk		LTC Risk		Health Risk	
Quantity	Raw	Derounded	Raw	Derounded	Raw	Derounded	Raw	Derounded
	q_i	$\tilde{q}_i$	q_i	$\tilde{q}_i$	q_i	$\tilde{q}_i$	q_i	$\tilde{q}_i$
$E(q_i \mid p_i = 0)$	0.25*** (0.01)	0.14*** (0.004)	0.21*** (0.01)	0.12*** (0.004)	0.12*** (0.002)	0.10*** (0.001)	0.18*** (0.003)	0.10*** (0.002)
β	0.38*** (0.02)	0.62*** (0.02)	0.45*** (0.01)	0.69*** (0.01)	0.18*** (0.01)	0.43*** (0.01)	0.34*** (0.03)	0.53*** (0.02)
$E(q_i \mid p_i = 1)$	0.61*** (0.02)	0.76*** (0.01)	0.66*** (0.01)	0.81*** (0.01)	0.31*** (0.01)	0.53*** (0.01)	0.51*** (0.03)	0.62*** (0.020)
Observations (Ind × wave)	4,736	4,736	15,706	15,706	74,851	74,851	17,194	17,194
Individuals	2,090	2,090	4,156	4,156	9,637	9,637	6,023	6,023

Table 2: Results from estimating equations (3.1) and (3.2). Notes: *** means significant at the 1% level against 0 (first row) or 1 (second and third rows). Only those in the training set are included.

Table 2 shows that the misperceptions are often not constant over the risk distribution. For example, Before de-rounding, the lowest mortality risk individuals are pessimistic, overstating their mortality prospects by 25% depending on the specification. But the highest mortality risk individuals are optimistic, understating their mortality prospects by 37%. As we move from the bottom to the top of the risk distribution, an extra 1% of predicted true risk is associated with only 0.38% greater subjective elicitation. With pessimism at the bottom, optimism at the top and attenuated response in the middle, at most risk levels there is error on average, but not uniform error as is often assumed.

The pattern can be summarized quite cleanly. When the objective risk is low, individuals over-perceive it. When the objective risk is high, individuals under-perceive it. But for some insurance decisions, for example the difference in medical insurance between a catastrophic contract and no insurance, the relevant risk is low for everyone. And so only half of the 'S' shaped pattern is relevant and pessimism throughout is observed. Further detail on the coherence between this pattern of misperception and the typical probability distortions observed in psychology and behavioral economics are in section 3.7.

3.5.1 Mortality Risk Evidence from Wave 1

A concern with the evidence presented above is that the rounding categories into which I have assigned individuals are incorrect. I have attempted to be conservative and err in the direction that would be adversarial to my results (by assigning individuals to coarser categories whenever in doubt). Nevertheless this concern might persist. To circumvent this concern, I offer evidence below that comes only from the responses in the first wave of the HRS.

In the first wave of the HRS, the same question was asked of survival to age 75, but the responses were constrained to be an integer from 0-10. These responses are then aligned with later HRS waves by treating the integer response of x as signifying a response of x in 10, or $10x$ percent[15]. So a response of 4 (out of 10) in wave one is equivalent to a response of 40% in any subsequent wave. Interpreted this way, the unit of rounding is clear. I assume that an individual considered all possibilities and rounded to the nearest 10%.

The same prediction and inference exercise is performed , except now restricted to the individuals for whom a prediction and subjective elicitation can be obtained in wave 1. This restricts the sample size considerably and there is more noise in the prediction and inference. The results, again with the two different specifications (raw and derounded elicitations) are presented in table 3.

The patterns are the same as table 2. In all specifications I observe pessimism at the bottom of the risk distribution, optimism at the top and attenuation as we move from the bottom to the top. I continue reject the hypothesis of constant misperception (in particular, of zero misperception).

[15] Study (2018)

	Mortality Risk	
Quantity	Raw q_i	Derounded $\tilde{q}_i$
$E(q_i \mid p_i = 0)$	0.26*** (0.01)	0.23*** (0.01)
β	0.31*** (0.04)	0.38*** (0.03)
$E(q_i \mid p_i = 1)$	0.57*** (0.03)	0.61*** (0.03)
Observations (Individuals)	1,420	1,420

Table 3: Results from estimating equations (3.1) and (3.2) restricted to wave 1. Notes: *** means significant at the 1% level against 0 (first row) or 1 (second and third rows). Only those in the training set are included.

3.5.2 Evidence from the Panel Data

The HRS is a panel data set that has been collected for almost 30 years. Up to this point I have pooled all observations without meaningfully using the panel structure, except when clustering standard errors. In this section I offer within-person evidence using the panel data to reinforce the conclusions so far. The questions such evidence can answer include: Is the within-person time variation in subjective perceptions concordant with the within-person time variation in predicted objective risk?

Introducing a time subscript, I model the relationship between objective risk and subjective perception as:

$$q_{it} = \alpha + \beta p_{it} + u_i + \epsilon_{it}. \tag{3.3}$$

I estimate (3.3) by time-demeaning within person, such that u_i drops out and β is identified using only within person variation. In particular, any time-invariant optimism or pessimism or unobservable factor drops out. It is very unlikely there is a true time invariant unobservable. Suppose an individual had some disease that was unobserved to the HRS. If at age 60 say this disease decreased the probability of survival by 10%, it is unlikely at age 62 the decreased probability of survival is still 10%, because the individual is now 2 years closer to 75. So the primary purpose of the fixed effect is to remove a hypothetical baseline optimism or pessimism, such that an individual always inflates or deflates their perception by a constant amount.

The first specification in table 4 is the fixed effect regression of (3.3) with raw elicitation q_i used. The second specification is again (3.3) with de-rounded $\tilde{q}_i$ used.

	Mortality Risk		Disability Risk		LTC Risk		Health Risk	
Quantity	Raw q_i	Derounded $\tilde{q}_i$	Raw q_i	Derounded $\tilde{q}_i$	Raw q_i	Derounded $\tilde{q}_i$	Raw q_i	Derounded $\tilde{q}_i$
β	0.10*** (0.03)	0.41*** (0.02)	0.52*** (0.02)	0.73*** (0.02)	0.14*** (0.01)	0.36*** (0.01)	0.12*** (0.02)	0.34*** (0.02)
Observations (Ind × wave)	5,252	5,252	15,706	15,706	74,851	74,851	17,194	17,194
Individuals	2,241	2,241	4,156	4,156	9,637	9,637	6,023	6,023

Table 4: Fixed Effect estimation of equation (3.3) by taking differences over 1 wave (2 years).

Note: *** means significant at the 1% level against 0 (first row) or 1 (second and third rows).

As before similar patterns obtain. Disability risk is quite well understood by individuals, with almost three quarters (after de-rounding) of within-person variation in objective risk updated into subjective elicitations. This speaks to the relative predictability of disability risk, as opposed to the rarer but more severe shocks that dramatically move health and mortality risk.

For the other three risks - mortality, long-term care and catastrophic health - table 4 shows that individuals do not appropriately update their risk perceptions. Even accounting for rounding bias, a 1% change in objective risk is reflected in an, at most, 0.41% change in subjective elicitation.

In combination with the main results in table 2, I conclude that the patterns of misperception are not simply due to baseline bias that is anchored to. Rather the attenuation pattern we see in the three risks except for catastrophic medical risk are due to within person attenuation. Whether individuals' risk changes favorably or unfavorably the full magnitude of this change isn't factored in, resulting in subjective perceptions that are bent toward the middle. Even though catastrophic medical risk exhibits pessimism throughout the risk distribution, the same within-person mechanism is likely the cause. There is baseline pessimism and since most movements are in the direction of lower out-of-pocket risk (either due to treatment or the upgrading of insurance), an underreaction to the lowered risk increases the pessimism.

3.6 Welfare and Market Implications

I have presented robust evidence that individual's misperceive various risks and that this misperception often covaries with the risk. The theory developed in this paper allow me to suggestively interpret these patterns in terms of welfare.

Beginning with mortality risk, while on average they are pessimistic about their mortality prospects (see Figure 5) this misses the heterogeneity in misperceptions. Figure 7 shows that the most mortal 40% of the population are on average *optimistic* about their mortality prospects, while the least mortal 60% are pessimistic. The latter group is unsurprising and consistent with, for example, O'Dea and Sturrock (2020).

The evidence here suggests that individuals underestimating their longevity affects not just individual WTP for longevity insurance products such as longevity, but distorts the amount of insurance offered in the equilibrium contract menu. And the longevity pessimism distorts not only the contracts that the highest longevity types buy, but exerts an externality through incentive constraints on all other contracts in the menu. This is consistent with the annuity market being heavily skewed toward products that offer minimal insurance, with the most most popular products being 80-90% bonds, as per Poterba and Solomon (2021). The welfare harm is transmitted through both the direct, demand suppression channel, and the indirect contract distortion mechanism.

Next, long-term care risk features a large majority of low and medium risk individuals who, roughly, accurately perceive their risk. But the top 5-15% exhibit substantial optimism by underperceiving their risk. This explains why long-term care insurance provides little coverage in comparison to potential care costs. Indeed Brown and Finkelstein (2007) find that more comprehensive policies are available but not purchased: The riskiest still choose partial insurance owing to their misperceptions. This cascades down the contract menu, and even the unbiased majority must purchase contracts distorted toward under-coverage due to the errors by the high risk type. Welfare is very likely harmed by this endogenous adjustment to the contract menu.

Finally, and in contrast, catastrophic medical risk is never over-perceived. Rather the highest risk types are unbiased in their perception, and almost all lower risk types think it is more likely than in actuality. This would mean that in a health insurance context, the most generous contracts ('platinum' contracts in the parlance of the Affordable Care Act) are correctly valued. But less generous contracts are over-valued. However, this is good for welfare, as those contracts are already distorted away from full insurance due to adverse selection. When the medium risk types over-value their medium coverage contracts, this allows the low risk types to move a little closer to full insurance, as per figure 2. This goes some way to correcting the distortion due to incentive constraints in the screening problem, and is likely welfare enhancing.

In sum, while all these risk markets exhibit misperceptions, it is not immediately clear what the welfare implications are. This paper develops the required model to do welfare analysis. I find that while misperceptions of risk likely depress welfare in many risk markets (for example, longevity and long-term care insurance) it might increase welfare in others such as medical insurance.

3.7 Cognitive Foundations

I have shown that there is a strong covariance between risk type and risk misperception. The general pattern is that when there are high risk types, they under-perceive their risk, and when there are low-risk types they over-perceive this risk. This typically generates an 'S' shaped pattern. But with an important exception: when the risk in question is almost always low, such as catastrophic medical risk, then only the pessimism remains.

This empirical pattern is broadly consistent with the experimental literature on how individuals perceive probabilities. Beginning with Tversky and Kahneman (1992) and summarized in Enke and Graeber (2019), in many contexts people have held (or acted as if they held) subjective probabilities that are 'S' shaped. For low probabilities, the subjective probability is above the objective, while for high probabilities the reverse is true. This is true in the evaluation of lotteries, choice under ambiguity, Bayesian updating and more.

The main cognitive foundation consistent with this pattern is confusion causing movement toward the middle as a default. While the findings in this paper are independent of the precise cause of the misperception, it is worth noting some possibilities. Fischhoff and Bruine De Bruin (1999) show that individuals who have no idea about the answer often respond with 50% when faced with a probability scale. If these confused individuals are spread throughout the risk distribution, this might mechanically generate this pattern.[16] A continuous version of this dynamic is posited by Enke and Graeber (2019) in which individuals choose something between their true belief and some

[16] The HRS asks a follow-up question to those who respond with 50%, querying whether they truly think it is equally likely or are just not sure. Excluding those who respond they are not sure does not change the empirical findings above.

cognitive default based on their degree of cognitive uncertainty. The more cognitively uncertain choose closer to the default, which might too be 50% on this scale.

Alternative cognitive causes might include insufficient updating away from a mean based on personal characteristics. This is consistent with the panel evidence shown in section 3.5.2. Even as time passes, and these risks usually change due to aging or as medical conditions arise or resolve or as mechanically the number of years left until the age of 75 fall, individuals on average fail to update their subjective perceptions sufficiently.

This paper provides evidence for the same pattern in a new setting. In individual's subjective perceptions of their own mortality and their long-term care risk, the probabilities too are 'S' shaped: They are compressed toward the middle. Novelly, this is true in the present context of probabilities that are fundamentally personal or private information, not simply in a carefully controlled lab setting or in the context of a future piece of public information such as a macroeconomic statistic. It is reassuring though that the literature has broadly concluded that this pattern actually reflects distorted views of probabilities, not just noise induced by surveys.

4 Conclusion

In this paper, through a variety of theoretical models, I analyzed the equilibrium impact of these type-differential risk misperceptions. I studied the equilibrium in terms of contracts offered and welfare accrued. The conclusion was that errors by the higher risk types are more consequential than the lower risk types. Pervasively I found that a misperception in a given direction has a different impact depending on which risk type misperceives and in which direction they misperceive their risk.

Empirically, using data from the HRS I then studied a variety of risks: mortality, long-term care, disability and catastrphic medical risk. I found diverse patterns. Mortality and long-term care risk misperception is widespread but not uniform in the population. High risk individuals under-perceive their risk, and low risk individuals over-perceive their risk. This pattern, interpreted through the theory, implies a welfare loss above and beyond the second-best adverse selection equilibrium. Conversely, disability risk perceptions are largely accurate, with any deviations explainable by rounding error. Finally, catastrophic medical risk is correctly perceived by the riskiest, and over-perceived by everyone else. This suggests a welfare improvement relative to the no-misperception equilibrium. I demonstrate these findings, except for disability risk as noted, to be robust to multiple econometric specifications and a generous accounting for elicitation error.

I have shown that adverse selection and misperceptions interact and impact welfare through a distortion in contracts offered. The welfare imapct of misperceptions are ambiguous and hence an empirical question. And theoretical or empirical work that assumes the set of contracts offered do not endogenously respond to demand-side misperceptions miss this important mechanism and welfare impact. Future research could ideally get closer to actual insurance purchases and claims than I have been able to. Confirming that these misperceptions are in fact demand relevant, and perhaps intervening to correct misperceptions and analyze the resulting demand changes are interesting avenues to pursue.

Bibliography

Abaluck, J. and J. Gruber (2016a, August). Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance. *The American economic review 106*(8), 2145–2184.

Abaluck, J. and J. Gruber (2016b, December). Improving the Quality of Choices in Health Insurance Markets. Technical report.

Akerlof, G. A. (1970). The Market for Lemons: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019, August). Regressive Sin Taxes, with an Application to the Optimal Soda Tax*. *The Quarterly Journal of Economics 134*(3), 1557–1626.

Azevedo, E. M. and D. Gottlieb (2017). Perfect Competition in Markets With Adverse Selection. *Econometrica 85*(1), 67–105.

Banks, J., A. Muriel, and J. Smith (2011). Attrition and health in ageing studies: evidence from elsa and hrs. *Longitudinal and Life Course Studies 2*(2), 101–126.

Bhargava, S., G. Loewenstein, and J. Sydnor (2017, August). Choose to Lose: Health Plan Choices from a Menu with Dominated Option*. *The Quarterly Journal of Economics 132*(3), 1319–1372.

Brown, J. R. and A. Finkelstein (2007, November). Why is the market for long-term care insurance so small? *Journal of Public Economics 91*(10), 1967–1991.

Chassagnon, A. and B. Villeneuve (2005). Optimal Risk-Sharing under Adverse Selection and Imperfect Risk Perception. *The Canadian Journal of Economics / Revue canadienne d'Economique 38*(3), 955–978.

Dubey, P. and J. Geanakoplos (2002). Competitive pooling: Rothschild-stiglitz reconsidered. *The Quarterly Journal of Economics 117*(4), 1529–1570.

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating Welfare in Insurance Markets Using Variation in Prices. *The Quarterly Journal of Economics 125*(3), 877–921.

Einav, L., A. Finkelstein, and P. Schrimpf (2015, April). The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D. *The Quarterly Journal of Economics 130.*

Enke, B. and T. Graeber (2019, November). Cognitive Uncertainty. Working Paper 26518.

Ericson, K. M., P. Kircher, J. Spinnewijn, and A. Starc (2020, May). Inferring Risk Perceptions and Preferences Using Choice from Insurance Menus: Theory and Evidence. *The Economic Journal* (ueaa069).

Fang, H., M. P. Keane, and D. Silverman (2008). Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. *Journal of Political Economy 116*(2), 303–350.

Finkelstein, A., N. Hendren, and M. Shepard (2019, April). Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts. *American Economic Review 109*(4), 1530–1567.

Fischhoff, B. and W. Bruine De Bruin (1999). Fifty–Fifty=50%? *Journal of Behavioral Decision Making 12*(2), 149–163.

Gan, L., M. Hurd, and D. Mcfadden (2003, March). Individual Subjective Survival Curves. *Analyses in the Economics of Aging*.

Handel, B., I. Hendel, and M. D. Whinston (2015). Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk. *Econometrica 83*(4), 1261–1313.

Handel, B. R. and J. T. Kolstad (2015, August). Health Insurance for Humans: Information Frictions, Plan Choice, and Consumer Welfare. *American Economic Review 105*(8), 2449–2500.

Handel, B. R., J. T. Kolstad, T. Minten, and J. Spinnewijn (2020, September). The Social Determinants of Choice Quality: Evidence from Health Insurance in the Netherlands. Technical report.

Handel, B. R., J. T. Kolstad, and J. Spinnewijn (2018, November). Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets. *The Review of Economics and Statistics 101*(2), 326–340.

Hendren, N. (2013). Private Information and Insurance Rejections. *Econometrica 81*(5), 1713–1762.

Hurd, M. and K. McGarry (2002, February). The Predictive Validity of Subjective Probabilities of Survival. *Economic Journal 112*, 966–985.

Hurd, M. D. and K. McGarry (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *The Journal of Human Resources 30*, S268–S292.

Jeleva, M. and B. Villeneuve (2004, May). Insurance contracts with imprecise probabilities and adverse selection. *Economic Theory 23*(4), 777–794.

Ketcham, J. D., C. Lucarelli, E. J. Miravete, and M. C. Roebuck (2012, May). Sinking, Swimming, or Learning to Swim in Medicare Part D. *American Economic Review 102*(6), 2639–2673.

Kleinjans, K. J. and A. V. Soest (2014). Rounding, Focal Point Answers and Nonresponse to Subjective Probability Questions. *Journal of Applied Econometrics 29*(4), 567–585.

Landais, C., A. Nekoei, P. Nilsson, D. Seim, and J. Spinnewijn (2017, October). Risk-Based Selection in Unemployment Insurance: Evidence and Implications. Technical report, Centre for Economic Performance, LSE.

Miyazaki, H. (1977). The rat race and internal labor markets. *The Bell Journal of Economics 8*(2), 394–418.

Mueller, A. I., J. Spinnewijn, and G. Topa (2018, November). Job seekers' perceptions and employment prospects: Heterogeneity, duration dependence and bias. Working Paper 25294, National Bureau of Economic Research.

Netzer, N. and F. Scheuer (2014). A game theoretic foundation of competitive equilibria with adverse selection. *International Economic Review 55*(2), 399 – 422.

O'Dea, C. and D. Sturrock (2020, August). Survival pessimism and the demand for annuities. Working Paper 27677, National Bureau of Economic Research.

Poterba, J. M. and A. Solomon (2021, March). Discount Rates, Mortality Projections, and Money's Worth Calculations for US Individual Annuities. *NBER Working Paper Series*.

Riley, J. (1979). Informational Equilibrium. *Econometrica 47*(2), 331–59.

Rothschild, M. and J. Stiglitz (1976, November). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics 90*(4), 629–649.

Sandroni, A. and F. Squintani (2007, December). Overconfidence, Insurance, and Paternalism. *American Economic Review 97*(5), 1994–2004.

Sandroni, A. and F. Squintani (2013, September). Overconfidence and asymmetric information: The case of insurance. *Journal of Economic Behavior & Organization 93*, 149–165.

Shi, S. Y. (1988, January). Rothschild-Stiglitz competitive insurance market under quasilinear preferences. *Economics Letters 27*(1), 27–30.

Spence, M. (1978, December). Product differentiation and performance in insurance markets. *Journal of Public Economics 10*(3), 427–447.

Spinnewijn, J. (2013). Insurance and Perceptions: How to Screen Optimists and Pessimists. *The Economic Journal 123*(569), 606–633.

Spinnewijn, J. (2015). Unemployed But Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association 13*(1), 130–167.

Stantcheva, S. (2020, August). Understanding Tax Policy: How Do People Reason? Technical report.

Stiglitz, J. E. (1977). Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies 44*(3), 407–430.

Stiglitz, J. E., J. Yun, and A. Kosenko (2018, June). Characterization, existence, and pareto optimality in insurance markets with asymmetric information with endogenous and asymmetric disclosures: Revisiting rothschild-stiglitz. Working Paper 24711, National Bureau of Economic Research.

Study, H. . R. (2018). public use dataset. produced and distributed by the university of michigan with funding from the national institute on aging (grant number nia u01ag009740).

Tversky, A. and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Wilson, C. (1977, December). A model of insurance markets with incomplete information. *Journal of Economic Theory 16*(2), 167–207.

Young, V. R. and M. J. Browne (2000, December). Equilibrium in Competitive Insurance Markets Under Adverse Selection and Yaari's Dual Theory of Risk. *The Geneva Papers on Risk and Insurance Theory 25*(2), 141–157.

A Theoretical Extensions

A.1 Monopoly

As another robustness check that the particular equilibrium concept is not driving the qualitative results, here I consider the provision of insurance by a monopolist instead of a competitive market or planner. As standard, the monopolists problem is to, writing the endowment as $\mathbf{0} = (w, w-l)$,:

$$\max_{\boldsymbol{c}_H, \boldsymbol{c}_L} \sum_{i \in \{L,H\}} \alpha_i \pi(p_i, \boldsymbol{c}_i) \tag{A.1}$$

$$\text{such that} \tag{A.2}$$

$$V(q_H, \boldsymbol{c}_H) \geq V(q_H, \boldsymbol{c}_H) \tag{A.3}$$

$$V(q_L, \boldsymbol{c}_L) \geq V(q_L, \boldsymbol{c}_H) \tag{A.4}$$

$$V(q_H, \boldsymbol{c}_H) \geq V(q_H, \mathbf{0}) \tag{A.5}$$

$$V(q_L, \boldsymbol{c}_L) \geq V(q_L, \mathbf{0}). \tag{A.6}$$

The equilibrium configuration is a direct adaptation of the various parts of Jeleva and Villeneuve (2004) and is summarized below. Using their terminology: when I say one type receives optimal insurance means that the type is receiving their first best contract **fixing the level of profit** from that types contract. Or geometrically, it means that the type's indifference curve is tangent to the iso-profit line at their equilibrium contract.

Proposition 5 (Jeleva and Villeneuve (2004))**.** *Suppose $q_H > q_L$ and errors are small. In the unique optimum:*

1. *(A.6) and (A.3) bind.*
2. *Types are always seperated.*
3. *H types obtain optimal insurance. If $q_H > p_H$ this is over-insurance, if $q_H < p_H$ then under insurance.*
4. *L types are sub-optimal partial insurance, perhaps none.*
5. *There is a threshold $\overline{\gamma_H}$ such that low types receive no insurance if $\gamma_H > \overline{\gamma_H}$ and are served if not.*

The case where the low type is not served is not particularly interesting. In such a case low type errors clearly have no impact. The high types participation constraint binds according to their perceived probability, so a pessimistic error is bad and an optimistic error good according to their objective welfare as calculated by the planner.

The more interesting dynamics occur when both types are served. The monopolist faces a trade-off: They can extract more profit from the low types by offering them more and more insurance but this requires less (perhaps negative) profits to be earned from the high types due to the binding incentive constraint. Conversely, they can extract more profits from the high types by raising the price of optimal insurance and pushing the high types closer to their outisde option, but for the same incentive reasosn the low types contract must then feature even less insurance and profit (perhaps zero of both) so as to not tempt the high type. These forces will help with interpreting the welfare implications of errors.

First, consider a small upward misperception by the high type with no error by the low type. The low type will continue to receive utility equal to their outside option and so their welfare will not change. The high types incentive constraint has now loosened. By how much it has loosened depends on where their risk was to begin with. When they were already high risk, the loosening is greater (at the limit, consider flat indifference curves) and so per dollar of extra profit to be extracted from the low type the incentive cost to the high type is lower. This motivates the (sufficient) condition $2p_H > 1 + p_L$. The higher risk the high type is, the higher the maginal loosening in incentive cost is when they make a small error, and so the temptation to earn more profits from the low types whilst making the high types slightly better off is stronger.

On the other hand, suppose the low type makes a small error. The effect on their welfare is unambiguous. If the low types think they are riskier than they are, the monopolist will sell them insurance they don't need, still restrict them to zero utility - but according to their perceived pessimistic probability - which the planner will then determine as detrimental to their welfare as defined by the objective probability. On the other hand, if the low types are optimistic, their zero perceived utility earned will actually be an objective improvement in the planners eyes.

But the effect of a low type error on a high type is more subtle. To charge more for insurance to the low types (their IR curve has shifted left), the high types must be made better off. If the high types are very different from the low types in risk, then the insurer can charge the low types more for the same loss payout whilst not affecting their profit from the high types much at all (at the limit, again, consider high types having horizontal indifference curves). If the high types are closer in risk to the low types, then extracting profit from the low type will require a greater marginal loss from the high types. Hence, the (sufficient) condition $\frac{p_H(1-p_L)}{p_L(1-p_H)} < \frac{u'(w)}{u'(w-l)}$ says that when p_H is small and p_L large, i.e. their difference is small, the low types naivete can be exploited at lower change to high types profit.

Proposition 6. *For a small error, the effects on welfare for the high and low types are:*

For an error by the high type $\frac{\partial V_H(p_H,\boldsymbol{c}^*)}{\partial q_H} \mid_{\boldsymbol{q}=\boldsymbol{p}} > 0$ *if* $2p_H > 1 + p_L$, $\frac{\partial V_L(p_L,\boldsymbol{c}^*)}{\partial q_H} \mid_{\boldsymbol{q}=\boldsymbol{p}} = 0$.. *For an error by the low type* $\frac{\partial V_H(p_H,\boldsymbol{c}^*)}{\partial q_L} \mid_{\boldsymbol{q}=\boldsymbol{p}} > 0$ *if* $\frac{p_H(1-p_L)}{p_L(1-p_H)} < \frac{u'(w)}{u'(w-l)}$, $\frac{\partial V_L(p_L,\boldsymbol{c}^*)}{\partial q_L} \mid_{\boldsymbol{q}=\boldsymbol{p}} < 0$..

As compared to the competitive market dynamics and the planners optimal choice problem, the key difference here is that errors by both types matter for welfare. But the main qualitative insight holds: errors that make types more disparate (which, loosely speaking, might 'increase' the amount of private information) can actually be good for welfare by weakening the incentive constraints. When there is no endogenous contract adjustment, this channel is closed down and the welfare conclusions might be changed.

A.2 Wilson-Miyazaki equilibrium, alternative interpretation

Two concerns that the have been repeatedly levelled at the Rothschild and Stiglitz (1976) equilibrium concept are that sometimes no equilibrium exists and even when the equilibrium does exist it need not be pareto optimal. To address these concerns a loosening of the equilibrium notion was developed by Wilson (1977) and subsequently further studied by Miyazaki (1977), Spence (1978) and Netzer and Scheuer (2014) amongst others. When deciding whether a given menu is an equilibrium, instead of allowing any profitable deviations, the Wilson-Miyazaiki (WM) concept only considers deviations that continue to make a non-negative profit even after all contracts rendered unprofitable by the initial deviation are withdrawn. Thus, firms anticipate best responses to their deviations in this limited manner. A formal definition is given in Wilson (1977).

There are two ways to interpret the MW equilibrium. In game-theoretic terms as described above or as a particular contrained efficient solution to the planner's problem. THe literature has largely taken the latter as canonical (see e.g. Netzer and Scheuer (2014)). For that reason that constrained efficient solution is priveleged in the main body of the paper. In this appendix I study the game-theoretic interpretation of the MW equilibrium for theoretical completeness.

From now on I focus on the maximization problem that defines the MW equilibrium. Although there is no planner performing this maximization, the market behaves as if the (subjectively (mis-)perceived)) welfare of the low risk type is being maximized subject to constraints. Consider the program:

$$\max_{\boldsymbol{c}_H,\boldsymbol{c}_L} V(q_L, \boldsymbol{c}_L) \tag{A.7}$$

$$\text{such that} \tag{A.8}$$

$$V(q_H, \boldsymbol{c}_H) \geq V(q_H, \boldsymbol{c}_L) \tag{A.9}$$

$$V(q_L, \boldsymbol{c}_L) \geq V(q_L, \boldsymbol{c}_H) \tag{A.10}$$

$$V(q_H, \boldsymbol{c}_H) \geq V(q_H, \boldsymbol{c}_H^{RS}) \tag{A.11}$$

$$\alpha_H \pi(p_H, \boldsymbol{c}_H) + \alpha_L \pi(p_L, \boldsymbol{c}_L) = 0 \tag{A.12}$$

$$(\boldsymbol{c}_H, \boldsymbol{c}_L) \in A = \left\{ \boldsymbol{c}_H, \boldsymbol{c}_L : MRS_H \leq \frac{1 - p_H}{p_H} \wedge c_L^{NL} \geq c_L^L \right\}. \tag{A.13}$$

This generalizes the program in Miyazaki (1977) and Netzer and Scheuer (2014) by allowing for q_H and q_L to differ from p_H and p_L. In addition I make one more restriction on the equilibrium. The usual restriction on the contract space is that $c_{NL} \geq c_L$, for both types and is called an indemnity constraint (as in Netzer and Scheuer (2014)) such that the loss cannot make an individual better off. I wish to keep this in spirit - not allowing individuals to harm themselves through perceived over-insurance - but I do not want to shut down the wedge between subjective and objectively defined optimal insurance. So I define the set of acceptable contracts, A, such that $\boldsymbol{c}$ can only allow for small over-insurance for the high type, and somce I will assume that errors are small, such that the only over-insurance that can be bought is that which improves perceived utility for the high type. This assumption rules out non-infinitesimal over-insurance for the low types, so the standard indemnity constraint can remain for them.

This rules out equilibria where, according to the misperceived risk, a type could be made better off with a profit-neutral reduction of insurance. If no errors are made, $\boldsymbol{q} = \boldsymbol{p}$, this reduces to the standard indemnity constraint. If one omits this restriction, uniqueness of the equilibria that follow might be lost, and the new equilibria will be qualitatively similar but possibly inverted.

Proposition 7. *For small errors $\xi = \|\boldsymbol{q} - \boldsymbol{p}\|,$, the unique solution to the program (A.7) has (A.12), (A.9) and (A.13) binding and coincides with the the unique MW equilibrium also constrained by (A.13) and (**??**).*

There is a continuum of contract pairs that satisfy the three constraints in the above program. They differ only in their degree of cross-subsidization from low-type to high type (never the other way). It will be useful to think in terms of this cross-subsidization, so I define:

$$\chi = p_H c_H^L + (1 - p_H) c_H^{NL} - (w - l p_H) \geq 0.$$

And so program (A.7) can be written as a univariate maximization over χ where the constraints above and definition of χ implicitly define the contracts as functions of χ, e.g. $c_H^{NL} = c_H^{NL}(\chi)$ and so on. The solution can then be thought of as the optimal amount of cross-subsiziation from low types to high types to maximise the perceived utility of the former.

When $\chi = 0$ this reduces to the RS equilibrium. When $\chi > 0$ the equilibrium features a profit-making contract offered to the low types and a loss-making contract offered to the high types that makes both better off relative to $\chi = 0$. RS is the particular case where the low types are not willing to cross-subsidize the high types so that the high types will receive more in both states of the world and the low types can move closer to full insurance owing to the relaxed IC constraint. This happens, most intuitively, when there are too many high types, so that the cross-subsidization is too expensive to justify. On the other hand when there are many low types, the RS equilibrium, even if it existed, would be inefficient as the high types could all pay a small subsidy to get a relatively large relaxation of the high types IC constraint.

The prior sections study the comparative statics when $\chi = 0$. Here I focus on the case where, prior to errors being made, there is a non-zero amount of cross-subsidization: $\chi > 0$.

It will turn out that in that situation the movement of χ with q_H or q_L is welfare sufficient. In particular, welfare increases with cross-subsidization. And so if some errror increases the cross-subsidization in the equilibrium then welfare will increase, and vice versa. And so if I sign $\frac{\partial}{\partial q_H}\chi^*(q_H)$ or similarly for q_L the welfare implications will follow.

Proposition 8. *Suppose without errors $\chi^* > 0$. Then if a small error is made: welfare increases in χ, $\frac{\partial}{\partial q_H}\chi^*(q_H) < 0$ and $\frac{\partial}{\partial q_L}\chi^*(q_L) > 0$.*

To understand the intuition for this result, one should keep in mind that the χ^* in the MW equilibrium is pushed higher and higher so long as the low types are still willing to give a dollar to the high types to relax the incentive constraint. The question is how an error by either type changes the incentive for the low types to give the final dollar of cross-subsidy to the high types.

First suppose the high types make a small upward error, $q_H > p_H$. Then, holding the no-error χ^* fixed, the high types contract will move up and left on the same iso-profit line, as in figure 3 (except with the iso-profit line not necessarily earning zero). The low types will then also receive closer to full insurance and be better off even before the cross-subsidy is adjusted. But because the low type is now closer to full insurance, the marginal utility from relaxing the cross-subsidy is now lessened, and so the final dollar of cross-subsidy that was optimal with no errors is now longer optimal with this error. There is a counterveiling force: that the marginal relaxation of the incentive constraint with a dollar of cross-subsidy is larger with the higher q_H (as the indifference curves are flatter) but this is dominated by the first effect. As such the new $\chi^* \mid_{q_H > p_H} < \chi^* \mid_{q_H = p_H}$. But welfare is evalauted according the objective probabilities, and $\chi^* \mid_{q_H = p_H}$ was optimal according to objective probabilities, and so the final foregone dollar of cross-subsidy is welfare decreasing.

Conversely, but similarly, if the low types make a downward error, $q_L < p_L$, they are no longer willing to pay for the final dollar of cross-subsidy, even though it is welfare optimal. This explains the sign of $\frac{\partial}{\partial q_L}\chi^*(q_L) > 0$.

These results reverse the logic from the RS equilibrium: errors that make a type think they are closer to the other type are welfare enhancing, errors that spread the difference between perceived types are welfare reducing. The intuition here is that **type-spreading errors are substitutes for cross-subsidies.** That is, iin the RS equilibrium where cross-subsidies are impossible, the only mechanism by which the incentive constraint can be loosened is through errors that spread the types apart. This is what was found in proposition 2. In the MS equilibrium, cross-subsidies are present without errors. Those cross-subsidies loosen the incentive constraint by as much as the low type is willing to pay. When we add an error which also loosens the incentive constraint, the low types cross-subsidize a little less. In this sense when cross-subsidies are present they are crowded out, suboptimally, when errors are introduced.

This highlights a crucial distinction: the impact of errors depends on whether or not contracts are currently cross-subsidizing each other.

A.3 Heterogeneity in misperceptions within risk classes, seperating equilibrium

Amongst the simplifications in the basic model in section 2 is that everyone in a given risk class makes the same misperception. Clearly this is unrealistic: For any given level of true risk we would expect some people to be optimistic, some to be pessimistic, and others accurate. In this section we demonstrate how the equilibrium notion generalizes in when there is heterogeneity of misperception with a risk class. The key forces that determine the equilibrium, and hence that determine the welfare consequences of the misperceptions, remain qualitatively unchanged. Throughout we focus on the case where there are two risk classes, and within each 2 different classes of misperception are made. The analysis naturally extends to any finite number of risk classes and within those any finite number of perceived risk classes, as will be clear.

Amongst those with true risk p_H, we define $q_{H,O}$ and $q_{H,P}$ as two different perceived risk classes: The former being relative optimists, the latter relative pessimists, such that $q_{H,O} < q_{H,P}$. Similarly amongst those with true risk p_L we analogously define $q_{L,O}$ and $q_{L,P}$. I write $\alpha_{H,O}, \alpha_{H,P}, \alpha_{L,O}, \alpha_{L,P}$ for the proportion of the population that falls into each of these classes. In this section I study the case in which

$$q_{H,P} > q_{H,O} > q_{L,P} > q_{L,O}. \tag{A.14}$$

This leads to qualitatively similar outcomes as in the core model. In the following subsection A.4, we study the case in which $q_{H,O} < q_{L,P}$ and find this to be a novel mechanism for pooling in equilibrium - a rare theoretical foundation for a ubiquitous observed phenomenon.

The equilibrium that obtains if we assume (A.14) is illustrated in figure 8. It is formalized in the following proposition.

Proposition 9. *Suppose $\|\boldsymbol{p}-\boldsymbol{q}\|_\infty < \xi$, and A.14 holds. For small enough ξ, in the unique locally-competitive equilibrium, high risk pessimistic individuals buy the contract $\boldsymbol{c}_{H,P}$ that solves:*

$$MRS(q_{H,P}, \boldsymbol{c}_{H,P}) = \frac{1-p_H}{p_H} \text{ and } \pi(p_H, \boldsymbol{c}_{H,P}) = 0,$$

high risk optimistic individuals buy the contract $\boldsymbol{c}_{H,O}$ that solves:

$$MRS(q_{H,O}, \boldsymbol{c}_{H,O}) = \frac{1-p_H}{p_H} \text{ and } \pi(p_H, \boldsymbol{c}_{H,O}) = 0,$$

and low risk individuals receive the contract $\boldsymbol{c}_L$ that solves

$$V(q_{H,O}, \boldsymbol{c}_{H,O}) = V(q_{H,O}, \boldsymbol{c}_L) \text{ and } \pi(p_L, \boldsymbol{c}_L) = 0.$$

The high risk types, both optimistic and pessimistic, obtain their perceived first best contract subject to zero profit being earned. Naturally, $\boldsymbol{c}_{H,P}$ features more insurance than $\boldsymbol{c}_{H,O}$ as the pessimists think the risk is more likely then the optimists and so have a greater demand for insurance.

The low risk types all receive the same insurance, and as in 1 this equilibrium contract features partial insurance distorted away from the first best owing to the information rents that accrue to the high types. As I noted above, misperceptions by the low-types have no effect on the equilibrium menu, and hence heterogeneity in misperceptions by the low types also have no effect.

The equilibrium menu is separating in the sense that all true high types are separated from all true low types, as opposed to the following subsection. The incentive constraint that binds to determine the low types contract is that of the most optimistic high type. But this is optimistic only relative to the pessimistic high types, not necessarily relative to the truth. That is, if $p_H > q_{H,O}$

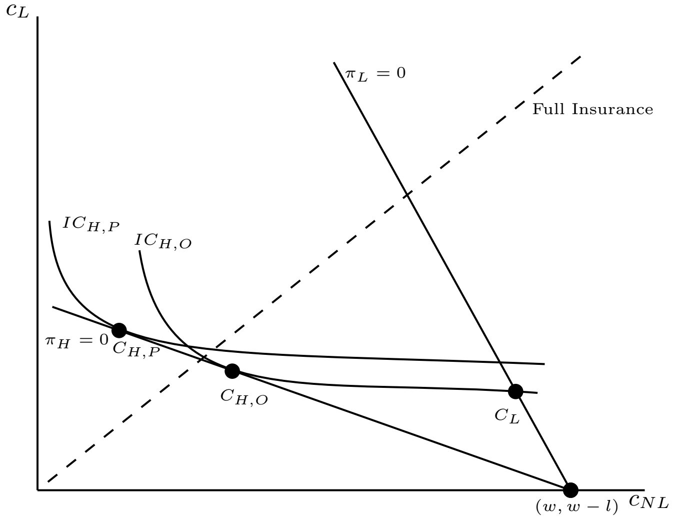


Figure 8: Heterogenous perceptions within risk class, seperating equilibrium

(as is illustrated in figure 8) in which case the misperception will certainly create a welfare loss in the market in the same way as proposition 2. Or, if $p_H < q_{H,O} < q_{H,P}$ then if $q_{H,O} - p_H$ is small enough then the misperceptions can lead to a welfare improvement as also discussed in proposition 2.

Broadly, when the heterogeneity within risk classes is ordered as in (A.14), all the dynamics apply as in the core model and proofs therein, with the slight caveat that the relevant incentive compatibility constraint for determining a lower types contract is that of the most optimistic amongst the higher types.

A.4 Partial pooling in the Rothschild and Stiglitz (1976) framework

The assumption (A.14) states that despite heterogeneity within risk classes, the most pessimistic low type still thinks they are lower risk than the most optimistic high type. As the empirical analysis speaks to, this is likely unrealistic, with each risk class having a large amount of heterogeneity in perceived risk, such that the most pessimistic low types perceive their risk to be higher than the most optimistic high types. In this section I study the effect of such an ordering. Specifically, I assume:

$$q_{H,P} > q_{L,P} > q_{H,O} > q_{L,O}. \tag{A.15}$$

I will find that under a mild assumption, formalized below, that in equilibrium those with perceived risk $q_{L,P}$ are pooled together with those with perceived risk $q_{H,O}$. Define the average risk if those with perceived risk $q_{L,P}$ and $q_{H,O}$ are pooled together:

$$p_{HL} = \frac{\alpha_{L,P} p_L + \alpha_{H,O} p_H}{\alpha_{L,P} + \alpha_{H,O}}.$$

The partial-pooling equilibrium that obtains is:

Proposition 10. *Suppose A.15 holds, $\alpha_{H,P} \geq \alpha_{H,O}$ and $MRS(q_{L,P}, \boldsymbol{c}_{HL}) < \frac{1-p_{HL}}{p_{HL}}$.*[17] *In the*

[17] This assumption rules out a deviation to a contract that offers less insurance, is preferred by both pooled types

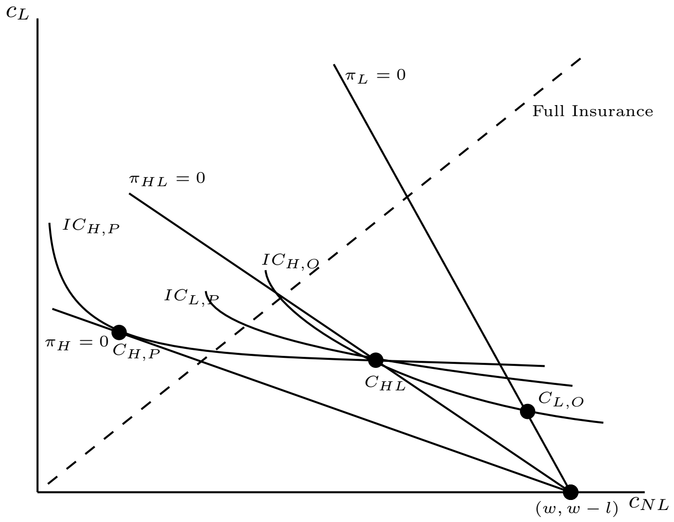


Figure 9: Heterogenous perceptions within risk class, pooling equilibrium

unique locally-competitive equilibrium, high risk pessimistic individuals buy the contract $\boldsymbol{c}_{H,P}$ *that solves:*

$$MRS(q_{H,P}, \boldsymbol{c}_{H,P}) = \frac{1 - p_H}{p_H} \text{ and } \pi(p_H, \boldsymbol{c}_{H,P}) = 0,$$

high risk optimistic ***and*** *and low risk pessimistic individuals buy the contract* $\boldsymbol{c}_{HL}$ *that solves:*

$$V(q_{H,P}, \boldsymbol{c}_{H,P}) = V(q_{H,P}, \boldsymbol{c}_{HL}) \text{ and } \pi(p_{HL}, \boldsymbol{c}_{HL}) = 0,$$

and low risk optimistic individuals receive the contract $\boldsymbol{c}_{L,O}$ *that solves*

$$V(q_{H,O}, \boldsymbol{c}_{HL}) = V(q_{H,O}, \boldsymbol{c}_{L,O}) \text{ and } \pi(p_L, \boldsymbol{c}_{L,O}) = 0.$$

This equilibrium is illustrated in Figure 9. The pessimistic high types still receive their perceived first-best contract subject to zero profits being made. Then, most notably, the pessimistic low types and the optimistic high types are pooled together. The contract they are pooled into is defined so that the pessimistic high types are indifferent between in and their contract, as standard.

It is worth considering why this specific heterogeneity (equation (A.15) in misperception allows for the fundamental no-pooling result in Rothschild and Stiglitz (1976) to not hold here. Typically, any pooling contract is not an equilibrium because a deviating contract offering slightly less insurance, attracting only the lower risk type would make a profit. Here due to the fact that $q_{H,O} < q_{L,P}$, a contract offering less insurance is more attractive to the *higher* risk type, due to their large misperception. So you cannot 'cream-skim' in the normal way by offering less insurance than the pooling contract. Here, inversely, to seperate the two pooled types and only attract the lower (true) type, a contract with more insurance than the pooled contract needs to be offered. This is because the lower of the two true types being pooled has a higher perceived type, per (A.15).

So skimming the lower type out of the pool can be done with a contract offering more insurance. , But such a deviation will also attract the pessimistic high types, who are even higher perceived risk

(but no one else) and makes a positive profit. Alternatively we could use the Riley notion since such a deviation is not 'safe'. This is a minor assumption unrelated to the substance of these two different true risk classes being pooled together, as the discussion makes clear.

than the optimistic low type being skimmed out. The assumption that $\alpha_{H,P} > \alpha_{H,O}$ means that this cream skimming contract replaces the optimistic high types with a larger mass of pessimistic high types, thereby making a loss.

This result resolves the empirically unsatisfactory no-pooling prediction of the Rothschild-Stiglitz model. It shows that misperceptions of risk that cause 'overlap' of perceived risk types as in (A.15) can generate equilibrium pooling. Alternative mechanisms that generate pooling in this framework include: departing from the expected utility framework as in Shi (1988); quantity caps as in Dubey and Geanakoplos (2002); or endogenous sharing of information by firms that allows for partial non-exclusivity of contracts (e.g. Stiglitz et al. (2018)).

Even in this differently structured equilibrium, the qualititive insights of the core model hold true. Misperceptions made by the higher types still cascade down and exert an externality on the equilibrium contract offered to lower types. The welfare consequences are still ambiguous, with the pooled contract being worse for the pessimistic low type and better for the optimistic high type than the no-misperception equilibrium.

B Proofs

B.1 Proof of Proposition 1

Proof. We prove a more general claim with three types, H, M and L. The result follows by allowing M and L to have the same low risk. The claim is:

Proposition. *Suppose $\|\boldsymbol{p}-\boldsymbol{q}\|_\infty < \xi$, with $q_H > q_L$. For small enough ξ, in the unique locally-competitive equilibrium, high risk individuals buy the contract $\boldsymbol{c}_H$ that solves:*

$$MRS(q_H, \boldsymbol{c}'_H) = \frac{1-p_H}{p_H} \text{ and } \pi(p_H, \boldsymbol{c}'_H) = 0,$$

medium risk individuals receive the contract $\boldsymbol{c}_M$ that solves

$$V(q_H, \boldsymbol{c}'_H) = V(q_H, \boldsymbol{c}'_M) \text{ and } \pi(p_M, \boldsymbol{c}'_M) = 0,$$

and low risk individuals receive the contract $\boldsymbol{c}_L$ that solves

$$V(q_M, \boldsymbol{c}'_M) = V(q_M, \boldsymbol{c}'_L) \text{ and } \pi(p_L, \boldsymbol{c}'_L) = 0.$$

Step 1. No pooling

First, note an intuitive but repeatedly used fact: for types i, j if $q_i > q_j$ then the equilibrium contract that i buys (and perhaps others) must offer weakly more insurance than the equilibrium contract j buys: $\boldsymbol{c}_i \succsim \boldsymbol{c}_j$. If i and j pool this is trivial. If not, then either the contracts are not comparable with respect to $\succsim$, in which case one dominates the other and an IC constraint must break, or $\boldsymbol{c}_j \succ \boldsymbol{c}_i$. In that case, we must have $V(q_j, \boldsymbol{c}_j) \geq V(q_j, \boldsymbol{c}_i)$, but also that $V(q_i, \boldsymbol{c}_j) > V(q_i, \boldsymbol{c}_i)$ since $MRS(q_i, \boldsymbol{c}_j) < MRS(q_j, \boldsymbol{c}_j)$ and $\boldsymbol{c}_j \succ \boldsymbol{c}_i$. This contradicts the i types IC constraint. Hence higher types must receive weakly more insurance that lower types in equilibrium. Denote this observation by **(*)**

The impossibility of pooling between H and M types, and between M and L types, and between all three types, in a RS equilibrium, and hence in a locally-competitive equilibrium, follows by almost identical arguments to R-S. We need only check that the deviating contract that attracts the relatively lower type in each situation does not attract the currently un-pooled type, if relevant. For the latter two cases this is obvious. For the first, where we conjecture one contract that pools and one that attracts only L, $\boldsymbol{c}_{HM}, \boldsymbol{c}_L$, note simply that the the low typs IC constraint cannot bind. If it did, then the medium types IC constraint wouldn't bind (as the indifference curves are distinct), and so a deviation to offering $\boldsymbol{c}_L + (-\epsilon, \epsilon r)$ with $r \in (MRS(q_L, \boldsymbol{c}_L), \frac{1-p_L}{p_L})$ will by construction improve utility, earn positive profits, yet not attract any M or H types for small enough $\epsilon > 0$. So the low types IC constraint does not bind, so a contract marginally more attractive to the medium type will not attract the low type, and hence the pooling between high and medium types cannot be sustained. [18]

Finally, H and S cannot pool in any locally competitive equilibrium. If they did, then by the observation (*) above the contracts must satisfy $\boldsymbol{c}_{HS} \succ \boldsymbol{c}_M$ or the opposite. In the former case, we contradict (*) with S types being more insured than M types. In the latter, we contradict the same (*) with M types being more insured than H types. This establishes that pooling cannot be sustained in any configuration.

Step 2. Any allocation except the solution to the maximization problem in the Proposition cannot be a local equilibrium

Step 1 shows that an equilibrium menu must feature three seperating contracts with $\boldsymbol{c}_H \succ \boldsymbol{c}_M \succ \boldsymbol{c}_L$.

We first verify that the incentive constraints cannot bind upwards. Assume that $\boldsymbol{c}$ underinsures.[19] For a contradiction suppose $V(q_M, \boldsymbol{c}_M) = V(q_M, \boldsymbol{c}_H)$. This implies $V(q_H, \boldsymbol{c}_M) < V(q_H, \boldsymbol{c}_H)$. Hence, for small enough $\epsilon > 0$, there is a new contract $\boldsymbol{c}'_M = \boldsymbol{c}_M + (-\epsilon, \epsilon r)$ with $r \in \left(MRS(q_M, \boldsymbol{c}_M), \min\left\{\frac{1-q_M}{q_M}, MRS(q_L, \boldsymbol{c}_L)\right\}\right)$ that strictly improves on the utility of the medium types, yet makes positive profits as it doesn't attract any other types. Hence the IC constraint preventing the medium type from reporting themselves as high doesn't bind in any equilibrium.

Similarly, we cannot have $V(q_L, \boldsymbol{c}_L) = V(q_L, \boldsymbol{c}_M)$. Assume that $\boldsymbol{c}_L$ features underinsurance.[20]. If we did have $V(q_L, \boldsymbol{c}_L) = V(q_L, \boldsymbol{c}_M)$, then we must have $V(q_M, \boldsymbol{c}_L) < V(q_M, \boldsymbol{c}_M)$. This also implies that $V(q_H, \boldsymbol{c}_L) < V(q_H, \boldsymbol{c}_M)$ as the MRS curves are always shallower for H types. Hence with $r \in \left(MRS(q_L, \boldsymbol{c}_L), \frac{1-q_L}{p_L}\right)$, then $\boldsymbol{c}_L + (-\epsilon, \epsilon r)$ will make low types strictly better off, earning strictly positive profits by just attracting low types so long as $\epsilon > 0$ is small enough.

Then, suppose the H types received any contract except the specific contract whilst maintaining zero profits. Then we must have $MRS\left(q_H, \boldsymbol{c}_H\right) \neq \frac{1-p_H}{p_H}$. In either case there is a contract slightly north-west or south-east of the current contract (depending on whether the inequality is $>$ or $<$, that offers positive profits and strictly higher utility to the H types without attracting any lower types when $\epsilon > 0$ is small since their IC constraints do not bind upwards.

Then, suppose that $V(q_H, \boldsymbol{c}_H) > V(q_H, \boldsymbol{c}_M)$. Then the deviation from we considered M's IC constraint (pretending to be H) is a profitable, utility improving deviation here as well (assuming underinsurance at $\boldsymbol{c}_M$)[21] . Similarly, suppose that $V(q_M, \boldsymbol{c}_M) > V(q_M, \boldsymbol{c}_L)$. Then the deviation from when we considered Ls IC constraint suffices here as well. Note that that deviation relies on L

[18] Note an $r \in (MRS(q_L, \boldsymbol{c}_L), \frac{1-p_L}{p_L})$ exists when q_L receives less than full insurance. If L is has ful insurance, their IC constraint cannot bind as they are at their first-best. If they are strictly over-insured then they can move toward full insurance with a positive profit made.

[19] If not, then since $MRS\left(q_M, \boldsymbol{c}_M\right) > \frac{1-p_M}{p_M}$ for overinsured contracts, the M types indifference curve is always above $(1-p_M)(W - c_M^{NL}) + p_M(c_M^L - (W - l)) = 0$ for any contract that features more insurance than $\boldsymbol{c}_M$, and hence cannot intersect with $(1-p_H)(W - c_H^{NL}) + p_H(c_H^L - (W - l)) = 0$ at such contract, and hence the IC constraint for the medium type cannot bind.

[20] If not, then the logic of the footnote above applies

[21] If not, and $\boldsymbol{c}_M$ is overinsured, then and since Ls IC constraint does nto bind, M can be moved toward full insurance slightly at a profit and utility gain since $MRS\left(q_M, \boldsymbol{c}_M\right) > \frac{1-p_M}{p_M}$ for overinsured $\boldsymbol{c}_M$

being underinsured, which must be the case since they can always be offered less insurance without affecting the IC constraints.

This shows that the given menu is the only possible equilibrium configuration.

Step 3. Existence

It remains to verify that the proposed menu is indeed a locally competitive equilibrium.

Consider first $\boldsymbol{c}_H$. At $\boldsymbol{c}_H$ we have $MRS(q_H, \boldsymbol{c}_H) = \frac{1-p_H}{p_H}$. Further, since $U'' < 0$ these curves are tangent.This directly means that any contract attracting only H types cannot both make a non-negative profit and also strictly improve upon H's utility. As above, the IC constraints of the lower types do not bind upward, and so no local deviation will attract anyone but the H types. It follows there are no local deviations near $\boldsymbol{c}_H$.

Consider next $\boldsymbol{c}_M$. Recall that $MRS(q_M, \boldsymbol{c}_M) < \frac{1-p_M}{p_M}$ for small enough ξ. Consider a deviation in the direction $(-1, r)$ from $\boldsymbol{c}_M$ with $r \in \left[MRS\left(q_M, \boldsymbol{c}_M\right), \frac{1-p_M}{p_M}\right]$. Since $MRS\left(q_M, \boldsymbol{c}_M\right) > MRS\left(q_M, \boldsymbol{c}_M\right)$ this contract attracts high and medium types, but for small ϵ this will be loss making since $p_{HM} > p_M$ and so the pooled contract will make profits arbitrarily close to $(1-p_{HM})(W - c_M^{NL}) + p_{HM}(c_M^L - (W-l)) < (1-p_M)(W - c_M^{NL}) + p_M(c_M^L - (W-l)) = 0$. Similarly, any contract deviations with $r \in [MRS\left(q_M, \boldsymbol{c}_M\right), MRS\left(q_M, \boldsymbol{c}_M\right))$ will attract only high types and also earn negative profits. With $r > MRS\left(q_M, \boldsymbol{c}_M\right)$ no one will demand the contract.

Then if we consider deviations of the form $(1, r)$ from $\boldsymbol{c}_M$, if $r > -\frac{1-p_M}{p_M}$ then the contract will make negative profit if it attracts medium types, and hence also if it attracts medium and high or just high types. If $r < -\frac{1-p_M}{p_M}$ in particular $r < -MRS\left(q_M, \boldsymbol{c}_M\right) < MRS\left(q_H, \boldsymbol{c}_H\right)$ for small ξ and so no medium or high types will swap to it, and for small enough deviations neither will low types. This shows that there are no local deviations from $\boldsymbol{c}_M$. Similar working rules out local deviations from $\boldsymbol{c}_L$. [22]

From this we can conclude that the proposed menu is a locally competitive equilibrium , and from step 2 that there are no others. This concludes the proof.

■

B.2 Proof of Proposition 2

Proof. First we suppose that the error made is a small upward error by the high types. Proposition 2 shows that the high types contract can be found as the solution to the constrained maximization problem:

$$\hat{V}_H(\boldsymbol{q}) \equiv \max_{\boldsymbol{c}} V(q_H, \boldsymbol{c}) \quad \text{s.t } \pi_H(p_h, \boldsymbol{c}) = 0.$$

By the envelope theorem, for the multiplier $\lambda > 0$ we have

$$\frac{\partial \hat{V}}{\partial q_H} = \frac{\partial V}{\partial q_H} + \lambda \frac{\partial g}{\partial q_H} = \frac{\partial V}{\partial q_H} = u(c_H^L) - u(c_H^{NL}).$$

In particular, at $q_H = p_H$ we have $c_H^L = c_H^{NL}$ and hence :

$$\frac{\partial \hat{V}}{\partial \xi}|_{q_H = p_H} \frac{\partial \hat{V}}{\partial \xi}|_{q_H = p_H} = 0.$$

For low types, we can write their equilibrium contract, for a small error, as the solution to the (degenerate) maximization program:

[22] This all assumes that $\boldsymbol{c}_M$ features underinsurance, which is true for small enough ξ.

$$\hat{V}_L(\boldsymbol{q}) \equiv \max_{\boldsymbol{c}} V(q_L, \boldsymbol{c}) \quad \text{s.t } \pi_L(p_L, \boldsymbol{c}) = 0$$
$$\text{and } h_L \equiv V(q_H, \boldsymbol{c}_H) - V(q_H, \boldsymbol{c}) = 0.$$

This is degenerate in the sense that the constraints exactly pin down the solution, and hence the maximization is superfluous. Nevertheless, with multiplier $\lambda > 0$ on the IC constraint we immediately get that:

$$\frac{\partial \hat{V}_L}{\partial q_H} = u(c_H^L) - u(c_H^{NL}) - (u(c_L^L) - u(c_L^{NL})).$$

At $q_H = p_H$ we have $c_H^L = c_H^{NL}$ and $c_L^L < c_L^{NL}$ and so

$$\frac{\partial \hat{V}_L}{\partial q_H}|_{q_H=p_H} = \frac{\partial \hat{V}_L}{\partial \xi}|_{q_H=p_H} > 0.$$

It follows that for a small change ξ, we have

$$\frac{\partial W}{\partial \xi} > 0 \text{ when } \xi > 0, \text{ and } \frac{\partial W}{\partial \xi} < 0 \text{ when } \xi < 0,$$

which completes the proof.

■

B.3 Proof of Propositions 3 and 4

Proof. First, we begin with IC_H. Suppose IC_H doesn't bind. If the optimum is pooling then this is an immediate contradiction. So the optimum must seperate the types. Then IC_L must bind else one type can be made better off holding cross-subsidization fixed (since both cannot be at their tangency point, by the assumption of small errors, as then both individuals tangency points are arbitrarily close to full insurance and one will be strictly preferred to the other by both types).

If $\boldsymbol{c}_L \succ \boldsymbol{c}_H$ we have an immediate contradiction as the indifference curve of the high type is strictly shallower than the low type, so if $V_L(\boldsymbol{c}_L, q_L) = V_L(\boldsymbol{c}_L, q_L)$ then it must be that $V_H(\boldsymbol{c}_L, q_H) > V_L(\boldsymbol{c}_H, q_H)$, contracting IC_H. On the other hand, if $\boldsymbol{c}_H \succ \boldsymbol{c}_L$ then, since $(\boldsymbol{c}_H, \boldsymbol{c}_L) \in A$ and the contracts are seperate, the low type must receive less than full insurance, $c_L^{NL} > c_L^L$. But, fixing a level of cross-subsidization, their optimal contract from the planner's perspective is full insurance. They can then be made better off by moving along the profit line $\pi_L = -\chi$ toward full insurance. That IC_H will continue to be slack for small enough movements, giving a contradiction to welfare optimality. Hence IC_H must bind. It immediately follows that unless there is pooling, IC_L does not bind.

Then suppose there is a small error by the high type only. Whereas the no-error optimum for the high type featured $c_H^{NL} = c_H^L$, now slight over-insurance $c_H^{NL} + \epsilon = c_H^L$ for some small $\epsilon > 0$ is feasible, according to the constraints A. Holding constant the level of cross-subsidazation, the proof of Proposition 2 shows that this slight over-insurance for the high types and a correspondingly better contract for the low types owing to the relaxed IC_H strictly improves welfare. This proves the first welfare property (with respect to the high type misperceptions) and shows that part of requirement (2.7) binds, in that $MRS_H = \frac{1-p_H}{p_H}$.

Now if the low type makes a small error, that IC_H binds and IC_L doesn't immediately demonstrates the welfare property with respect to misperceptions by the low type.

It remains to establish the cross-subsidization results. First, the comparative static results for $c_H^{NL}(\chi), c_H^L(\chi), c_L^{NL}(\chi), c_L^L(\chi)$ due to the implicit function theorem in the proof of proposition 8 hold here. So we calculate

$\frac{\partial}{\partial \chi} Welfare$,, substitute in comparative statics (B.56) through (B.63), and then differientiate with respect to q_H to get the cross partial, and then evaluate at $\boldsymbol{q} = \boldsymbol{p}$ to study small errors. The resulting expression is

$$\frac{(p_L - 1)p_L(\rho - 1)\left(u'(c_L^L) - u'(c_L^{NL})\right)\left((a_H - 1)u'(c)\left((p_L - 1)u'(c_L^L) - p_L u'(c_L^{NL})\right) + a_H u'(c_L^L)u'(c_L^{NL})\right)}{(a_H - 1)\left(p_H(p_L - 1)u'(c_L^L) - (p + H - 1)p_L u'(c_L^{NL})\right)^2}.$$

The denominator is negative as $a_H < 1$. The numerator is positive and hence the cross partial is negative. That means the return to a little more cross-subsidization decreases in q_H. The claims follow. ■

B.4 Proof of Lemma 1

Proof. Similarly to the final part of the proof above, we substitute into the welfare function the comparative static results for $c_H^{NL}(\chi), c_H^L(\chi), c_L^{NL}(\chi), c_L^L(\chi)$ from (B.56) through (B.63), and then evaluate $\frac{\partial^2}{\partial \rho \partial \chi} Welfare \mid_{\boldsymbol{q}=\boldsymbol{p}}$.

This yields

$$\frac{(p_H - p_L)\left((a_H - 1)u'(c)\left((p_L - 1)u'(c_L^L) - p_L u'(c_L^{NL})\right) + a_H u'(c_L^L)u'(c_L^{NL})\right)}{(a_H - 1)\left(p_H(p_L - 1)u'(c_L^L) - (p_H - 1)p_L u'(c_L^{NL})\right)}.$$

The denominator is easily seen to be positive, and the numerator as well, making the whole expression positive. This shows that the return to cross-subsidaztion increases in ρ. The results follow.

■

B.5 Proof of Proposition 9

Proof. Essentially identical to the proof of proposition 1. ■

B.6 Proof of Proposition 10

Proof. Essentially identical to the proof of proposition 1 except that it must be established that the optimistic high types and pessimistic low types must be pooled together.

Suppose not. First, suppose that the are not pooled together. The pessimistic high types must then have their own contract that earns zero profits.[23] As such both true low types must individually earn zero profits. But then we cannot have both that $V(q_{L,P}, \boldsymbol{c}_{LP}) \geq V(q_{H,O}, \boldsymbol{c}_{HO})$ and the reverse since we have $\boldsymbol{c}_{HO} \gtrsim \boldsymbol{c}_{LP}$ whilst $q_{H,O} < q_{L,P}$ contradicting the fact established at the beginning of proposition 1.

So they must be pooled together and the contract must earn zero profits. If not another firm will offer a slightly more attractive contract. That the constraints must bind in the way described is clear and arguments as in proposition 1 apply.

[23] This is because they cannot be pooled with the high pessimistic types since a deviation along $\pi_H = 0$ but tangent for the high optimistic types will attract just them as the low pessimistic types and high pessimistic types prefer the high pessimistic types contract. They cannot be pooled with the lowest types for the standard reasons.

It remains to show that the contract defined by the zero profit line from the pool of high optimists and low pessimists and the high pessimists binding IC constraint is in fact an equilibrium. The usual deviation away from pooling is to just attract the lower risk type by offering slightly less insurance. This doesn't work here as offering slightly less insurance attracts the high optimists but not the low pessimists as the latter think they are higher risk than the former, despite being objectively lower risk. So to attract the objectively low risk pessimists, slightly more insurance needs to be offered. But given that $V(q_{H,O}, \boldsymbol{c}_{H,O}) = V(q_{H,O}, \boldsymbol{c}_L)$, offering slightly more insurance that attracts the low pessimists will also attract the high optimists, since $(A.15)$ holds. Then, because of the additional assumption that $\alpha_{H,P} > \alpha_{H,O}$ such a contract will make less than the zero profit from $\boldsymbol{c}_{HL}$ since the group of high optimists is being replaced by the larger pool of high pessimists. Hence there is no local deviation and the equilibrium is as posited.

■

B.7 Proof of Proposition 6

First I calculate the change in welfare with respect to a small high type error. The process for a low type is similar and not explicated at length.

Proof. The three binding constraints are respectively the tangency constraint for the high type, the incentive constraint for the high type, and the participation constraint for the low type:

$$\frac{(1-q_H)u'(c_H^{NL})}{q_H u'(c_H^L(c_H^{NL}))} = \frac{1-p_H}{p_H} \tag{B.1}$$

$$\text{IC} = q_H u(c_H^L(c_H^{NL})) + (1-q_H)u(c_H^{NL}) = q_H u(c_L^L(c_H^{NL})) + (1-q_H)u(c_L^{NL}(c_H^{NL})) \tag{B.2}$$

$$\text{IR} = q_L u(c_L^L(c_H^{NL})) + (1-q_L)u(c_L^{NL}(c_H^{NL})) = q_L u(w-l) + (1-q_L)u(w). \tag{B.3}$$

Given these, I can implicitly express each of c_H^L, c_L^L, c_L^{NL} in terms of c_H^{NL} in which case the monopolists maximization problem is solely a function of c_H^{NL}. As such, from the implicit function theorem I have

$$\frac{\partial c_H^{NL}}{\partial qH} = -\frac{\partial^2 \Pi/\partial qH \partial c_H^{NL}}{\partial^2 \Pi/\partial^2 c_H^L}.$$

Recalling that

$$\Pi(c_H^{NL}) = \alpha_H \left[(1-p_H)\left(W - c_H^{NL}\right) - p_H(c_H^L(c_H^{NL}) - (W-l))\right] + \alpha_H \left[(1-p_L)\left(W - c_L^{NL}(c_H^{NL})\right) - p_L(c_L^L(c_H^{NL}) - (W$$

differentiating we have

$$\frac{\partial^2 \Pi(c_H^{NL})}{\partial qH \partial c_H^{NL}} = \alpha_H \left[p_H \frac{\partial^2}{\partial qH \partial c_H^{NL}}(c_H^L(c_H^{NL}))\right]$$
$$+\alpha_H \left[(1-p_L)\left(\frac{\partial^2}{\partial qH \partial c_H^{NL}} c_L^{NL}(c_H^{NL})\right) - p_L(\frac{\partial^2}{\partial qH \partial c_H^{NL}} c_L^L(c_H^{NL})))\right]$$

and

$$\frac{\partial^2 \Pi(c_H^{NL})}{\partial (c_H^{NL})^2} = \alpha_H \left[p_H \frac{\partial^2}{\partial (c_H^{NL})^2}(c_H^L(c_H^{NL}))\right] + \alpha_H \left[(1-p_L)\left(\frac{\partial^2}{\partial (c_H^{NL})^2} c_L^{NL}(c_H^{NL})\right) - p_L(\frac{\partial^2}{\partial (c_H^{NL})^2} c_L^L(c_H^{NL})))\right]$$

To calculate $\frac{\partial c_H^L(c_H^{NL})}{\partial c_H^{NL}}, \frac{\partial c_L^L(c_H^{NL})}{\partial c_H^{NL}}, \frac{\partial c_L^{NL}(c_H^{NL})}{\partial c_H^{NL}}$ differientiate each of the binding constraints with respect to c_H^{NL} to get respectively:

$$0 = \frac{(q_H - 1)\left(u''(c_H^{NL})u'(c_H^L(c_H^{NL})) - (c_H^L)'(c_H^{NL})u'(c_H^{NL})u''(c_H^L(c_H^{NL}))\right)}{q_H u'(c_H^L(c_H^{NL}))} \tag{B.4}$$

$$q_H(c_H^L)'(c_H^{NL})u'(c_H^L(c_H^{NL})) + (1-q_H)u'(c_H^{NL}) = q_H(c_L^L)'(c_H^{NL})u'(c_L^L(c_H^{NL})) + (1-q_H)(c_L^{NL})'(c_H^{NL})u'(c_L^{NL}(c_H^{NL})) \tag{B.5}$$

$$0 = q_L(c_L^L)'(c_H^{NL})u'(c_L^L(c_H^{NL})) + (1-q_L)(c_L^{NL})'(c_H^{NL})u'(c_L^{NL}(c_H^{NL})) \tag{B.6}$$

Solving this system yields

$$\frac{\partial c_L^L(c_H^{NL})}{\partial c_H^{NL}} = -\frac{(q_L - 1)\left((q_H - 1)u'(c_H^{NL}) - \frac{q_H u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2}{u'(c_H^{NL})u''(c_H^L(c_H^{NL}))}\right)}{(q_L - q_H)u'(c_L^L(c_H^{NL}))} \tag{B.7}$$

$$\frac{\partial c_L^{NL}(c_H^{NL})}{\partial c_H^{NL}} = -\frac{q_L\left((q_H - 1)u'(c_H^{NL}) - \frac{q_H u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2}{u'(c_H^{NL})u''(c_H^L(c_H^{NL}))}\right)}{(q_L - q_H)u'(c_L^{NL}(c_H^{NL}))} \tag{B.8}$$

$$\frac{\partial c_H^L(c_H^{NL})}{\partial c_H^{NL}} = \frac{u''(c_H^{NL})u'(c_H^L(c_H^{NL}))}{u'(c_H^{NL})u''(c_H^L(c_H^{NL}))} \tag{B.9}$$

To get the cross partials, differientiate with respect to q_H:

$$\frac{\partial^2}{\partial qH \partial c_H^{NL}} c_L^L(c_H^{NL})) = \frac{(q_L - 1)\left(q_L u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2 - (q_L - 1)u'(c_H^{NL})^2 u''(c_H^L(c_H^{NL}))\right)}{(q_H - q_L)^2 u'(c_H^{NL})u''(c_H^L(c_H^{NL}))u'(c_L^L(c_H^{NL}))} \tag{B.10}$$

$$\frac{\partial^2}{\partial qH \partial c_H^{NL}} c_L^{NL}(c_H^{NL})) = \frac{q_L\left(q_L u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2 - (q_L - 1)u'(c_H^{NL})^2 u''(c_H^L(c_H^{NL}))\right)}{(q_H - q_L)^2 u'(c_H^{NL})u''(c_H^L(c_H^{NL}))u'(c_L^{NL}(c_H^{NL}))} \tag{B.11}$$

$$\frac{\partial^2}{\partial qH \partial c_H^{NL}} c_H^L(c_H^{NL})) = 0. \tag{B.12}$$

Hence we have

$$\begin{aligned}\frac{\partial^2 \Pi(c_H^{NL})}{\partial q_H \partial c_H^{NL}} = &\frac{\alpha_L\left(q_L u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2 - (q_L - 1)u'(c_H^{NL})^2 u''(c_H^L(c_H^{NL}))\right)}{(q_H - q_L)^2 u'(c_H^{NL})u''(c_H^L(c_H^{NL}))u'(c_L^L(c_H^{NL}))u'(c_L^{NL}(c_H^{NL}))} \\ &\times \left((p_L - 1)q_L u'(c_L^L(c_H^{NL})) - p_L(q_L - 1)u'(c_L^{NL}(c_H^{NL}))\right)\end{aligned}$$

On the other hand, differentiating (B.7) with respect to c_H^{NL} again, and using the values from $(B.7)$ yields

$$\frac{\partial^2}{\partial(c_H^{NL})^2}c_L^L(c_H^{NL})) = \Bigg[(q_L-1)((1-q_L)u''(c_H^L(c_H^{NL}))u''(c_L^L(c_H^{NL}))(q_H u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2 \tag{B.13}$$

$$-(q_H-1)u'(c_H^{NL})^2u''(c_H^L(c_H^{NL})))^2 \tag{B.14}$$

$$-(q_H-q_L)u'(c_L^L(c_H^{NL}))^2(-q_H u^{(3)}(c_H^L(c_H^{NL}))u''(c_H^{NL})^2u'(c_H^L(c_H^{NL}))^3 \tag{B.15}$$

$$+q_H u'(c_H^L(c_H^{NL}))^2u''(c_H^L(c_H^{NL}))^2(u^{(3)}(c_H^{NL})u'(c_H^{NL})+u''(c_H^{NL})^2) \tag{B.16}$$

$$-(q_H-1)u'(c_H^{NL})^2u''(c_H^{NL})u''(c_H^L(c_H^{NL}))^3))\Bigg] \tag{B.17}$$

$$/\left[(q_H-q_L)^2u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3u'(c_L^L(c_H^{NL}))^3\right] \tag{B.18}$$

$$\frac{\partial^2}{\partial(c_H^{NL})^2}c_L^{NL}(c_H^{NL})) = \Bigg[-q_L((q_H-q_L)u'(c_L^{NL}(c_H^{NL}))^2(-q_H u^{(3)}(c_H^L(c_H^{NL}))u''(c_H^{NL})^2u'(c_H^L(c_H^{NL}))^3 \tag{B.19}$$

$$+q_H u'(c_H^L(c_H^{NL}))^2u''(c_H^L(c_H^{NL}))^2(u^{(3)}(c_H^{NL})u'(c_H^{NL})+u''(c_H^{NL})^2) \tag{B.20}$$

$$-(q_H-1)u'(c_H^{NL})^2u''(c_H^{NL})u''(c_H^L(c_H^{NL}))^3) \tag{B.21}$$

$$+q_L u''(c_H^L(c_H^{NL}))u''(c_L^{NL}(c_H^{NL}))(q_H u''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2-(q_H-1)u'(c_H^{NL})^2u''(c_H^L(c_H^{NL})))^2)\Bigg] \tag{B.22}$$

$$/\left[(q_H-q_L)^2u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3u'(c_L^{NL}(c_H^{NL}))^3\right] \tag{B.23}$$

$$\frac{\partial^2}{\partial(c_H^{NL})^2}c_H^L(c_H^{NL})) = \frac{u'(c_H^L(c_H^{NL}))(u^{(3)}(c_H^{NL})u'(c_H^{NL})u''(c_H^L(c_H^{NL}))^2-u^{(3)}(c_H^L(c_H^{NL}))u''(c_H^{NL})^2u'(c_H^L(c_H^{NL})))}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3}. \tag{B.24}$$

And so the second order total derivative of profit is

$$\frac{\partial^2\Pi(c_H^{NL})}{\partial(c_H^{NL})^2} = \frac{\gamma \mathrm{H}(p_H-1)u'(c_H^L(c_H^{NL}))(u^{(3)}(c_H^{NL})u'(c_H^{NL})u''(c_H^L(c_H^{NL}))^2 - u^{(3)}(c_H^L(c_H^{NL}))u''(c_H^{NL})^2u'(c_H^L(c_H^{NL})))}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3} \tag{B.25}$$

$$- \frac{(1-\alpha_H)p_L(q_L-1)(1-q_L)u''(c_H^L(c_H^{NL}))u''(c_L^L(c_H^{NL}))(q_Hu''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2}{(q_H-q_L)^2u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3u'(c_L^L(c_H^{NL}))^3} \tag{B.26}$$

$$- \frac{-(q_H-1)u'(c_H^{NL})^2u''(c_H^L(c_H^{NL})))^2}{(q_H-q_L)^2u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3u'(c_L^L(c_H^{NL}))^3} \tag{B.27}$$

$$- \frac{(q_H-q_L)u'(c_L^L(c_H^{NL}))^2(-q_Hu^{(3)}(c_H^L(c_H^{NL}))u''(c_H^{NL})^2u'(c_H^L(c_H^{NL}))^3}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3} \tag{B.28}$$

$$- \frac{q_Hu'(c_H^L(c_H^{NL}))^2u''(c_H^L(c_H^{NL}))^2(u^{(3)}(c_H^{NL})u'(c_H^{NL})}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3} \tag{B.29}$$

$$+ \frac{u''(c_H^{NL})^2) - (q_H-1)u'(c_H^{NL})^2u''(c_H^{NL})u''(c_H^L(c_H^{NL}))^3))}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3} \tag{B.30}$$

$$+ \frac{(1-\alpha_H)(1-p_L)q_Lq_Lu''(c_H^L(c_H^{NL}))u''(c_L^{NL}(c_H^{NL}))(q_Hu''(c_H^{NL})u'(c_H^L(c_H^{NL}))^2}{(q_H-q_L)^2u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3u'(c_L^{NL}(c_H^{NL}))^3} \tag{B.31}$$

$$+ \frac{(q_H-1)u'(c_H^{NL})^2u''(c_H^L(c_H^{NL})))^2}{(q_H-q_L)^2u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3u'(c_L^{NL}(c_H^{NL}))^3} \tag{B.32}$$

$$+ \frac{(q_H-q_L)u'(c_L^{NL}(c_H^{NL}))^2 - q_Hu^{(3)}(c_H^L(c_H^{NL}))u''(c_H^{NL})^2u'(c_H^L(c_H^{NL}))^3}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3} \tag{B.33}$$

$$+ \frac{q_Hu'(c_H^L(c_H^{NL}))^2u''(c_H^L(c_H^{NL}))^2(u^{(3)}(c_H^{NL})u'(c_H^{NL})}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3} \tag{B.34}$$

$$+ \frac{u''(c_H^{NL})^2) - (q_H-1)u'(c_H^{NL})^2u''(c_H^{NL})u''(c_H^L(c_H^{NL}))^3)}{u'(c_H^{NL})^2u''(c_H^L(c_H^{NL}))^3}. \tag{B.35}$$

Now, knowing how a change in q_H affects the optimized value of c_H^{NL}, I calculate how this will affect the optimized values of c_H^L, c_L^L, c_L^{NL} . The optimized values solve the same constraints of course, now written as dependant on q_H. Differentiating them with respect to q_H and substituting the known expression for $\frac{\partial^2\Pi(c_H^{NL})}{\partial(c_H^{NL})^2}$ yields a system that implicitly defines $\frac{\partial(c_L^L)^*}{\partial q_H}, \frac{\partial(c_L^{NL})^*}{\partial q_H}, \frac{\partial(c_H^L)^*}{\partial q_H}$.

Solving these gives, evaluated now at $\boldsymbol{p} = qvec$:

$$\frac{\partial (c_L^L)^*}{\partial q_H} = \frac{(p_L - 1)}{(p_H - p_L)u'(c_L^L)} \tag{B.36}$$

$$\times\left(\frac{u'(c)^2u'(c_L^L)^2u'(c_L^{NL})^2\left(u'(c_L^{NL}) - u'(c_L^L)\right)}{u'(c_L^{NL})^2\left((p_H - p_L)u''(c)u'(c_L^L)^2\left(u'(c_L^L) - u'(c_L^{NL})\right) - (p_L - 1)u'(c)^2u''(c_L^L)u'(c_L^{NL})\right) + p_Lu'(c)^2u'(c_L^L)^3u''(c_L^{NL})}\right. \tag{B.37}$$

$$\left. - \frac{u'(c)^2}{(p_H - 1)u''(c)} + u(c_L^L) - u(c_L^{NL})\right) \tag{B.38}$$

$$\frac{\partial (c_L^{NL})^*}{\partial q_H} = \frac{p_L}{(p_H - p_L)u'(c_L^{NL})}\left(u'(c)^2u'(c_L^L)^2u'(c_L^{NL})^2\left(u'(c_L^{NL}) - u'(c_L^L)\right)\right) \tag{B.39}$$

$$/\left(u'(c_L^{NL})^2\left((p_H - p_L)u''(c)u'(c_L^L)^2\left(u'(c_L^L) - u'(c_L^{NL})\right) - (p_L - 1)u'(c)^2u''(c_L^L)u'(c_L^{NL})\right)\right. \tag{B.40}$$

$$\left. + p_Lu'(c)^2u'(c_L^L)^3u''(c_L^{NL}) - \frac{u'(c)^2}{(p_H - 1)u''(c)} + u(c_L^L) - u(c_L^{NL})\right) \tag{B.41}$$

$$\frac{\partial (c_H^L)^*}{\partial q_H} = \left((p_H^2 - p_L)u'(c)u''(c)u'(c_L^L)^2u'(c_L^{NL})^2\left(u'(c_L^L) - u'(c_L^{NL})\right) + u'(c)^3\left[p_Lu'(c_L^L)^3u''(c_L^{NL})\right.\right. \tag{B.42}$$

$$\left.\left. - (p_L - 1)u''(c_L^L)u'(c_L^{NL})^3\right]\right) \tag{B.43}$$

$$/\left((p_H - 1)p_Hu''(c)(u'(c_L^{NL})^2((p_H - p_L)u''(c)u'(c_L^L)^2\left(u'(c_L^L) - u'(c_L^{NL})\right)\right. \tag{B.44}$$

$$\left. - (p_L - 1)u'(c)^2u''(c_L^L)u'(c_L^{NL})) + p_Lu'(c)^2u'(c_L^L)^3u''(c_L^{NL}))\right). \tag{B.45}$$

Now we can evaluate the change in welfare at the new optimum versus the old optimum: Directly from the constraints we have

$$\frac{\partial W_L^*}{\partial q_H} \mid_{\boldsymbol{q}=\boldsymbol{p}} = 0.$$

For the high type:

$$\frac{\partial W_H^*}{\partial q_H} |_{\boldsymbol{q}=pvec} = \alpha_H p_H (c_H^L)'(q_H) u'(c_H^L) + \alpha_H (1-p_H)(c_H^{NL})'(q_H) u'(c_H^{NL}) \tag{B.46}$$

$$= \Bigg(\alpha_H (2p_H - p_L - 1) u'(c)^2 u''(c) u'(c_L^L)^2 u'(c_L^{NL})^2 \left(u'(c_L^L) - u'(c_L^{NL}) \right) \tag{B.47}$$

$$+ \alpha_H u'(c)^4 \left(p_L u'(c_L^L)^3 u''(c_L^{NL}) - (p_L - 1) u''(c_L^L) u'(c_L^{NL})^3 \right) \Bigg) \tag{B.48}$$

$$/ \Bigg((p_H - 1) u''(c) (u'(c_L^{NL})^2 \Big[(p_H - p_L) u''(c) u'(c_L^L)^2 \left(u'(c_L^L) - u'(c_L^{NL}) \right) \tag{B.49}$$

$$- (p_L - 1) u'(c)^2 u''(c_L^L) u'(c_L^{NL}) + p_L u'(c)^2 u'(c_L^L)^3 u''(c_L^{NL}) \Big] \Bigg) \tag{B.50}$$

The denominator can be seen to be negative, since $u'(c_L^L) - u'(c_L^{NL}) > 0$ and utility is concave. As such, the welfare change will be positive when, the numerator is negative. A sufficient condition for this is $2pH > 1 + pL > 0$, which proves the claim.

The calculations for a change in the low types error are completely analogous and omitted for brevity.

■

B.8 Proof of Proposition 7

Proof. First I show that each of the constraints must hold else the allocation cannot be an MW equilibrium.

First the budget constraint (A.12). If negative profit is earnt the contract will be withdrawn. If positive profit is made and the equilibrium is pooling, then consumption in each state of the world can be increased, and everyone will swap. If the equilibrium is seperating, then increase consumption for both contracts to make everyone better off whilst keeping the IC constraint binding. The old contracts will be withdrawn, the new contracts will be demanded by the appropriate types. Hence (A.12) must hold in the MW equilibrium.

Next, suppose that (**??**) does not hold. Write the level of profit earned from each type as $\pi_H = \pi(p_H, \boldsymbol{c}_H), \pi_L = \pi(p_L, \boldsymbol{c}_L)$. First, suppose IC_H binds and the equilibrium is separating. This means IC_L does not bind. By assumption (A.13), this means that $\boldsymbol{c}_H$ is down and left of the tangency point of iso-profit line for π_H with the appropriate H indifference curve, Then suppose a firm deviated to offer the contract $\boldsymbol{c'} = \boldsymbol{c} - (\epsilon, \frac{\epsilon(1-p_H)}{p_H}$. The high types are better off, as the slope of their utility function is monotone along an iso-profit line with a maximum at the point of tangency, whilst the low types still buy their old contract, and total profit has not changed by construction. So no contracts are withdrawn, and this is a valid deviation. So (**??**) must hold. Suppose the equilibrium is pooling, then the deviation to each types optimal contract subject to the pooling profit iso-profit line still breaks even and makes the high type strictly better off, and the low type weakly better off. All types change, so no contracts are withdrawn, so this is a valid deviation. Hence (**??**) must hold here by contradiction.

Suppose IC_L binds and the equilibrium is seperating. As $\boldsymbol{c}_H$ is down and right of the optimal insurance point on the π_H isoprofit line, then $MRS(q_H, \boldsymbol{c}_H) < \frac{1-p_H}{p_H}$ and also $MRS(q_H, \boldsymbol{c}_H) < MRS(q_L, \boldsymbol{c}_H)$ and so there exists a slope $\psi \in ([MRS(q_H, \boldsymbol{c}_H), \frac{1-p_H}{p_H}) \cap (MRS(q_H, \boldsymbol{c}_H) < MRS(q_L, \boldsymbol{c}_H))$ such that the contract $\boldsymbol{c} + \epsilon(-1, \xi)$ is preferred by H types to $\boldsymbol{c}_H$, not preferred by low types to $\boldsymbol{c}_L$

and costs less. This deviation shows the initial menu was not an MW equilibrium and hence (**??**) must hold.

Next, suppose that (A.11) does not hold, i.e. $MRS(q_H, \boldsymbol{c}_H) < V(q_H, \boldsymbol{c}_H^{RS})$. If a firm then offers a contract arbitrarily close (but earning positive profit) to $\boldsymbol{c}_H^{RS}$ all the high types will swap to it, and even if the low types prefer (perhaps after other contracts are withdrawn) it then profit will be even higher. So this is a valid deviation and hence (A.11) must hold.

Finally, suppose that (A.9) does not hold. Then as in RS the low types can be made better off with positive profit without inducing the high types to swap. So it must be that (A.9).

Now I show that the maximization has a unique solution. After which I will show that the MW equilibrium is exactly that unique solution. The three binding constraints mean that this is reducible to a one-dimensional problem. In particular, we can index the problem by the cross-subsidy χ. The constraint (A.11) implies $\chi \geq 0$. Moreover, an upper bound on feasible χ is the maximum that which induces pooling: $\chi_1 = (p_H - P_L)l$ since for any higher χ the incentive constraint for the high type cannot possibly bind, or that which hits the indemnity constraint for the low type (**??**), χ_2. Write $\overline{\chi} = \max\{\chi_1, \chi_2\}$. Moreover, a choice of $\chi \in [0, \overline{\chi}]$ uniquely defines $\boldsymbol{c}_H$ and $\boldsymbol{c}_L$. To see this: once a χ is chosen, the iso-profit for each type is pinned down. The high types iso-profit, together with (A.13), uniquely pins down $\boldsymbol{c}_H$ as the MRS is monotonic along an iso-profit line. Then (A.9) and the low types isoprofit line intersect at two places. One of them is up and left of $\boldsymbol{c}_H$ and violates (**??**). Hence only the intersection that offers the low type less insurance remains. So $\boldsymbol{c}_L$ is pinned down.

Given all this, the problem becomes:

$$\max_{\chi \in [0, \overline{\chi}]} V(q_L, \boldsymbol{c}_L(\chi)) \tag{B.51}$$

so that all the constraints hold with $\boldsymbol{c}_H, \boldsymbol{c}_L$ implicit functions of χ. The existence of a solution follows by Weierstrass' theorem. Uniqueness follows directly from the proof of uniqueness in Netzer and Scheuer (2014), the fact that errors are small and so the solutions to this problem are arbitrarily close to the problem without errors, and continuity of the second derivative. In particular, they showed that the second derivative is globally concave, which then holds here.

This shows that any MW equilibrium must satisfy the given constraints, and that a unique solution to the optimization problem exists. It remains to show that only the amongst the allocations that satisfy the constraints, only that which maximizes the objective is an MW equilibrium. Suppose not. Then suppose a firm does offer the menu that satisfies the constraints and maximizes the objective. By definition the low types will prefer to switch to the new contract. That contract makes weakly positive profit, by constraint (A.11). The high types will either remain in their old contract or swap to the newly offered contract for them. In the former case the deviating firm will make zero profits, in the latter they will make positive profits. No withdrawal of contracts will make the low types switch as their subjective utility is being maximized. This shows their is a MW deviation, and so the supposition was false. Hence the objective must be maximized. This concludes the proof.

■

B.9 Proof of Proposition 8

Proof. THe argument above establishes that to sign $\frac{\partial}{\partial q_H}\chi^*(q_H)$ when without errors it is the case that $\chi^* > 0$ the comparative statics with respect to q_H can be computed in the following way. (The result for q_L is entirely similar and omitted for brevity.

The following binding constraints (and a definition) implicitly define $c_H^{NL} = c_H^{NL}(\chi)$ and similarly for c_H^L, c_L^L, c_L^{NL}. :

$$\frac{u'((\chi))}{u'(c_H^L(\chi))} = \frac{(1-p_H)q_H}{p_H(1-q_H)} \tag{B.52}$$

$$q_H u(c_H^L(\chi)) + (1-q_H)u(c_H^{NL}(\chi)) = q_H u(c_L^L(\chi)) + (1-q_H)u(c_L^{NL}(\chi)) \tag{B.53}$$

$$w - l(p_H + p_L) = (1-\alpha_H)p_L c_L^L(\chi) + (1-\alpha_H)(1-p_L)c_L^{NL}(\chi) + \alpha_H(l(w-p_H)+\chi) \tag{B.54}$$

$$\chi = p_H c_H^L(\chi) + (1-p_H)c_H^{NL}(\chi) - (w - lp_H). \tag{B.55}$$

Differentiating each with respect to χ and solving yields

$$(c_H^{NL})'(\chi) = \frac{u''(c_H^L(\chi))u'(c_H^{NL}(\chi))}{p_H u'(c_H^L(\chi))u''(c_H^{NL}(\chi)) - (p_H-1)u''(c_H^L(\chi))u'(c_H^{NL}(\chi))} \tag{B.56}$$

$$(c_H^L)'(\chi) = \frac{u'(c_H^L(\chi))u''(c_H^{NL}(\chi))}{p_H u'(c_H^L(\chi))u''(c_H^{NL}(\chi)) - (p_H-1)u''(c_H^L(\chi))u'(c_H^{NL}(\chi))} \tag{B.57}$$

$$(c_L^L)'(\chi) = \Big[u'(c_H^L(\chi))u''(c_H^{NL}(\chi))\big(q_H(\alpha_H(-p_L)+\alpha_H+p_L-1)u'(c_H^L(\chi)) + \alpha_H p_H(q_H-1)u'(c_L^{NL}(\chi))\big) \tag{B.58}$$

$$+\,\alpha_H(p_H(-q_H)+p_H+q_H-1)u''(c_H^L(\chi))u'(c_H^{NL}(\chi))u'(c_L^{NL}(\chi)) \tag{B.59}$$

$$+\,(\alpha_H-1)(p_L-1)(q_H-1)u''(c_H^L(\chi))u'(c_H^{NL}(\chi))^2\Big] \tag{B.60}$$

$$/\Big[(\alpha_H-1)\big((p_H-1)u''(c_H^L(\chi))u'(c_H^{NL}(\chi)) - p_H u'(c_H^L(\chi))u''(c_H^{NL}(\chi))\big) \tag{B.61}$$

$$\times\big((p_L-1)q_H u'(c_L^L(\chi)) - p_L(q_H-1)u'(c_L^{NL}(\chi))\big)\Big] \tag{B.62}$$

$$(c_L^{NL})'(\chi) = \Big[u''(c_H^L(\chi))u'(c_H^{NL}(\chi))\big((\alpha_H-1)p_L(q_H-1)u'(c_H^{NL}(\chi)) - \alpha_H(p_H-1)q_H u'(c_L^L(\chi))\big) \tag{B.63}$$

$$+\,q_H u'(c_H^L(\chi))u''(c_H^{NL}(\chi))\big((p_L-\alpha_H p_L)u'(c_H^L(\chi)) + \alpha_H p_H u'(c_L^L(\chi))\big)\Big] \tag{B.64}$$

$$= /\Big[(\alpha_H-1)\big((p_H-1)u''(c_H^L(\chi))u'(c_H^{NL}(\chi)) - p_H u'(c_H^L(\chi))u''(c_H^{NL}(\chi))\big) \tag{B.65}$$

$$\times\big((p_L-1)q_H u'(c_L^L(\chi)) - p_L(q_H-1)u'(c_L^{NL}(\chi))\big)\Big]. \tag{B.66}$$

Being overly explicit,the objective can be written:

$$V_L(q_L,\chi) = q_L u(c_L^L(\chi(q_H),q_H)) + (1-q_L)u(c_L^{NL}(\chi(q_H),q_H))$$

By the implicit function theorem, the sign of $\chi'(q_H)$ is the same as the sign of $\partial^2 V_L(q_L,\chi)/\partial\chi\partial q_L$. So computing:

$$\begin{aligned}\frac{\partial^2 V_L(q_L,\chi)}{\partial\chi\partial q_H} &= q_L\frac{\partial c_L^L\chi}{\partial q_H}u'(c_L^L(\chi(q_H),q_H)) + (1-q_L)\frac{\partial c_L^{NL}(\chi(q_H),q_H)}{\partial q_H}u'(c_L^{NL}(\chi(q_H),q_H)) \qquad \text{(B.67)}\\ &+ q_L(c_L^L)'(q_H)(c_L^L)'(\chi)u''(c_L^L(\chi(q_H),q_H)) + (1-q_L)(c_L^{NL})'(q_H)(c_L^{NL})'(\chi)u''(c_L^{NL}(\chi(q_H),q_H)) \qquad \text{(B.68)}\end{aligned}$$

where $(c_L^{NL})'(q_H)$ is shorthand for the partial derivative with respect to q_H when χ is fixed (i.e. only the explicit dependence is taken into account, not the implicit dependence through $\chi(q_H)$. Such quantities can be found by considering each consumption quantity to be a function of q_H, not χ in (B.52) - (B.55), differentiating them all and solving. The solution to this is, suppressing dependence on q_H for brevity:

$$\begin{aligned}(c_H^{NL})'(q_H) &= \frac{(p_H-1)u'(c_H^L(q_H))^2}{(q_H-1)^2\big((p_H-1)u''(c_H^L(q_H))u'(c_H^{NL}) - p_Hu'(c_H^L)u''(c_H^{NL})\big)}\\ (c_H^L)'(q_H) &= \frac{(p_H-1)^2u'(c_H^L)^2}{p_H(q_H-1)^2\big((p_H-1)u''(c_H^L)u'(c_H^{NL}) - p_Hu'(c_H^L)u''(c_H^{NL})\big)}\\ (c_L^L)'(q_H) &= -\Big[(p_L-1)(p_H^2(q_H-1)^2u'(c_H^L)u''(c_H^{NL})(u(c_H^L)-u(c_H^{NL})-u(c_L^L)+u(c_L^{NL}))\\ &\quad -(p_H-1)p_H(q_H-1)^2u''(c_H^L)u'(c_H^{NL})(u(c_H^L)-u(c_H^{NL})-u(c_L^L)+u(c_L^{NL}))\\ &\quad +(p_H-1)p_H(q_H-1)u'(c_H^L)^2u'(c_H^{NL})-(p_H-1)^2q_Hu'(c_H^L)^3)\Big]\\ &\quad /\Big[p_H(q_H-1)^2((p_L-1)q_Hu'(c_L^L)-p_L(q_H-1)u'(c_L^{NL}))\big((p_H-1)u''(c_H^L)u'(c_H^{NL})-p_Hu'(c_H^L)u''(c_H^{NL})\big)\Big]\\ (c_L^{NL})'(q_H) &= \Big[p_L(-p_H^2(q_H-1)^2u'(c_H^L)u''(c_H^{NL})(u(c_H^L)-u(c_H^{NL})-u(c_L^L)+u(c_L^{NL}))\\ &\quad +(p_H-1)p_H(q_H-1)^2u''(c_H^L)u'(c_H^{NL})(u(c_H^L)-u(c_H^{NL})-u(c_L^L)+u(c_L^{NL}))\\ &\quad +p_H(p_H(-q_H)+p_H+q_H-1)u'(c_H^L)^2u'(c_H^{NL})+(p_H-1)^2q_Hu'(c_H^L)^3)\Big]\\ &\quad /\Big[p_H(q_H-1)^2((p_L-1)q_Hu'(c_L^L)-p_L(q_H-1)u'(c_L^{NL}))((p_H-1)u''(c_H^L)u'(c_H^{NL})-p_Hu'(c_H^L)u''(c_H^{NL}))\Big]\end{aligned}$$

The appropriate second order partials with respect to χ can be found by differientiating (B.56)-(B.63) with respect to χ once more. Substituting these all in, evaluating at $\boldsymbol{q}=\boldsymbol{p}$ and hence at $c_H^{NL}=c_H^L=c$ yields

$$\begin{aligned}&\frac{\partial}{\partial q_H}\chi'(q_H)\mid_{\boldsymbol{q}=\boldsymbol{p}} = (p_L-1)p_L((u(c_L^L)-u(c_L^{NL}))\\ &\times\Bigg[\frac{(-(\alpha_H-1)u'(c)((p_L-1)u''(c_L^L)-p_Lu''(c_L^{NL}))+\alpha_H(p_H-1)u''(c_L^L)u'(c_L^{NL})-\alpha_Hp_Hu'(c_L^L)u''(c_L^{NL}))}{(\alpha_H-1)(p_H(p_L-1)u'(c_L^L)-(p_H-1)p_Lu'(c_L^{NL}))^2}\\ &+\frac{-(u'(c_L^L)-u'(c_L^{NL}))((\alpha_H-1)u'(c)((p_L-1)u'(c_L^L)-p_Lu'(c_L^{NL}))+\alpha_Hu'(c_L^L)u'(c_L^{NL})))}{(\alpha_H-1)(p_H(p_L-1)u'(c_L^L)-(p_H-1)p_Lu'(c_L^{NL}))^2}\Bigg]\\ &< 0.\end{aligned}$$

The denominator is negative and the numerator is positive as $c_L^L < c_L^{NL}$, hence $u(c_L^L) < u(c_L^{NL})$ and $u'(c_L^L) > u'(c_L^{NL})$. This shows $\frac{\partial}{\partial q_H}\chi'(q_H) \mid_{\boldsymbol{q}=\boldsymbol{p}} < 0$. Similar working shows that for a small error by the low type:

$$\frac{\partial}{\partial q_L}\chi'(q_L) \mid_{\boldsymbol{q}=\boldsymbol{p}} = \frac{p_L\left((\alpha_H-1)u'(c)\left((p_L-1)u'(c_L^L)-p_Lu'(c_L^{NL})\right)+\alpha_Hu'(c_L^L)u'(c_L^{NL})\right)}{(\alpha_H-1)\left(p_H(p_L-1)u'(c_L^L)-(p_H-1)p_Lu'(c_L^{NL})\right)} > 0.$$

once one notices that $p_H > p_L, 1-p_L > 1-p_H$ and $u'(c_L^L) > u'(c_L^{NL})$ and hence the denominator is positive.

Lastly it remains to show that welfare increases in χ. Recall welfare is given by

$$\alpha_H(p_Hu(c_H^L(\chi))+(1-p_H)u(c_H^{NL}(\chi)))+(1-\alpha_H)(p_Lu(c_L^L(\chi))+(1-p_L)u(c_L^{NL}(\chi))).$$

Differentiating with respect to χ and substituting in $W = (B.58)$ and $(B.63)$, and then evaluating at $\boldsymbol{p} = \boldsymbol{q}$, yields

$$\frac{\partial W}{\partial \chi} = \frac{u'(c)\left((p_L-1)(\alpha_Hp_H-\alpha_Hp_L+p_L)u'(c_L^L)+p_L(-\alpha_Hp_H+(\alpha_H-1)p_L+1)u'(c_L^{NL})\right)}{p_H(p_L-1)u'(c_L^L)-(p_H-1)p_Lu'(c_L^{NL})} \tag{B.69}$$

$$+\frac{\alpha_H(p_H-p_L)u'(c_L^L)u'(c_L^{NL})}{p_H(p_L-1)u'(c_L^L)-(p_H-1)p_Lu'(c_L^{NL})}. \tag{B.70}$$

Note that the first order condition here reads:

$$p_L\frac{\partial c_L^L}{\partial \chi}+(1-p_L)\frac{\partial c_L^{NL}}{\partial \chi}=0,$$

and substituting this in and simplifying yields such that the change in welfare is just given by the effect on the high types utility:

$$\frac{\partial W}{\partial \chi} = \alpha_Hu'(c) > 0.$$

■

C Dataset Construction

C.1 Full set of mortality prediction covariates

The set of variables that appear in the long logit specification, and that were the starting point for the variable selection in the lasso logit specification, are as follows:

1. The respondents subjective elicitation.

2. Sex and age(nearest year)

3. Dummies for diabetes, cancer, lung disease, heart disease, stroke, arthritis, and the first time difference for all of these.

4. Dummies for whether the respondent has ever had any of these diseases.

5. The full set of interactions of 2. and 3.

6. BMI, dummies for being married, seperated, divorced, never marraid. The length of the current marriage.
7. A dummy for whether the respodnents mother and father are alive, and their current age (or age of death).
8. The subjective self health assessment and the first time difference.
9. The number of overnight hospital visits, nursing home visits, doctor visits and episodes of home care since the previous interview. Out of pocket medical expenditures and the first time difference.
10. Indices for activities of daily living (and a time difference), mobility, large muscle, gross motor skills, fine motor skills, and instrumental activities of daily living.
11. Indicators for whether depression was experienced, whether the respondent was happy, the number of days each week in which alchohol was drunk and the first time difference, whether the respondent now smokes or ever smoked, an indicator for high blood pressure.
12. Indicators for past or current memory problems, whether the repondent has public or private health insurance (if the latter how many plans), has life insurance.
13. The number of children, an indicator for whether the respondent is retired, current income and the first time difference in income.

D Logit prediction routine and results - Mortality Risk

The main analysis uses the preferred lasso logit specification. Nevertheless, while the visual evidence is compelling, the shrinkage estimator might introduce bias into the inference, even if it does reduce variance. In this section I use a standard logit prediction to check that, in the case of mortality risk, this bias is of no quantitative importance. The moral of this appendix is that the prediction quality and the qualitative takeaways from the inference are essentially unchanged no matter which algorithm is used.

The LASSO adds to the logit's log-likelihood function a 'penalty' or 'regularization' term of the form $\lambda |\beta|$ where λ is a tuning parameter to be chosen by cross-validation and β is the vector of coefficients to be estimated. The LASSO 'zeroes out' coefficients of variables with minimal predictive power which leads to a less noisy and better prediction.

Figure 10 plots the receiver-operator curves for the long logit model from the main text, the lasso logit under study here, the age and sex differentiated life tables from the Social Security Administration, and the self-reports. The lasso logit makes a minute improvement over the long logit, both of which are far superior predictions to the self-reports or the life tables. On this dimension then, the use of the regular logit model is unlikely to dramatically change prediction quality.

I now compare the results of the inference when the lasso logit predicted $\hat{p}_i$ is used in place of the long logit. For reference I report the latter once more. I estimate the following four OLS specifications:

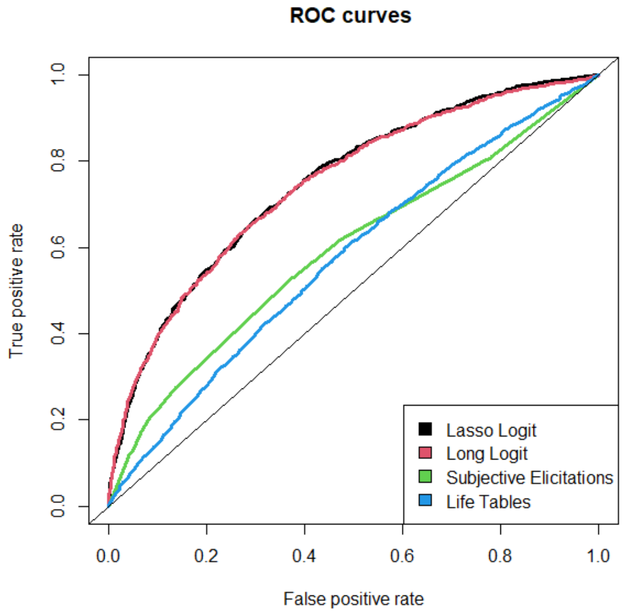


Figure 10: Receiver-operator curves for lasso logit, long logit, subjective elicitations (self-reports) and life tables

$$
\begin{aligned}
q_i &= \beta \hat{p}_i^{LassoLogit} + \epsilon_i, \\
q_i &= \beta \hat{p}_i^{LongLogit} + \epsilon_i, \\
\tilde{q}_i &= \beta \hat{p}_i^{LassoLogit} + \epsilon_i, \\
\tilde{q}_i &= \beta \hat{p}_i^{LongLogit} + \epsilon_i.
\end{aligned}
$$

As before, the quantites of interest are $E(q_i \mid p_i = 0)$, $E(q_i \mid p_i = 1)$ and $\beta \approx \frac{\partial q_i}{\partial p_i}$, the average error at the top of the risk distribution, the average error at the bottom of the risk distribution, and the slope of the conditional mean, interpretable approximately as the partial effect on the subjective elicitation of a change in the predicted true risk. The estimates for these three quantities for each of the four specifications are in table 5 below. Standard errors are clustered by respondent throughout and only the test set is included, never the training set.

The long logit results are marginally further to unbiasedness, but still in each of the three key tests the hypothesis of perfect private information can be resoundingly rejected.

Replicating the analysis using data and methods the first wave data yields similarly unchanged results. Those results are in the table below, along with their lasso logit analogues from the main text. This similarity gives confidence that the lasso logit model that might featrure bias introduced to reduce variance does not in fact bias the inference in any material way.

	Lasso Logit		Long Logit	
Quantity	Raw q_i	Derounded $\tilde{q}_i$	Raw q_i	Derounded $\tilde{q}_i$
	0.23***	0.13***	0.25***	0.16***
$E(q_i \mid p_i = 0)$	(0.01)	(0.005)	(0.01)	(0.005)
	0.46***	0.70***	0.38***	0.58***
β	(0.003)	(0.02)	(0.03)	(0.02)
	0.69***	0.83***	0.63***	0.74***
$E(q_i \mid p_i = 1)$	(0.02)	(0.01)	(0.02)	(0.01)
Observations (Ind × wave)	5,305	5,305	4,736	4,736
Individuals	2,227	2,227	2,090	2,090

Table 5: Results. *** means significant at the 1% level against 0 (first row) or 1 (second and third rows).

	Lasso Logit		Long Logit	
Quantity	Raw q_i	Derounded $\tilde{q}_i$	Raw q_i	Derounded $\tilde{q}_i$
	0.23***	0.19***	0.26***	0.23***
$E(q_i \mid p_i = 0)$	(0.01)	(0.01)	(0.01)	(0.01)
	0.45***	0.53***	0.31***	0.38***
β	(0.04)	(0.04)	(0.04)	(0.03)
	0.67***	0.71***	0.57***	0.61***
$E(q_i \mid p_i = 1)$	(0.03)	(0.03)	(0.03)	(0.03)
Observations (Individuals)	1,449	1,449	1,420	1,420

Table 6: Results from Wave One. *** means significant at the 1% level against 0 (first row) or 1 (second and third rows).