Product Design in Selection Markets*

André Veiga[†] E. Glen Weyl[‡]

December 2013

Abstract

Insurers choose plan characteristics to selectively attract cheap consumers. In a model with multidimensional heterogeneity, this sorting incentive is proportional to the covariance, among marginal consumers, between marginal willingness-to-pay and cost to the insurer. Standard forms of cost-sharing attract high cost consumers, but lowering the comprehensiveness of a plan repels them. In competitive equilibrium, this covariance over the full population must vanish. A competitive equilibrium with positive insurance is possible when insurance value is sufficiently negatively correlated with cost, unlike in Handel, Hendel and Whinston (2013)'s data, where Rothschild and Stiglitz (1976)'s non-existence result still holds.

Keywords: selection markets, cream-skimming, insurance markets, multidimensional heterogeneity, product design
JEL Classification Codes: D41, D42, D43, D86, I13

*This paper replaces earlier working papers "Multidimensional Heterogeneity and Platform Design" and "Multidimensional Product Design". We thank Eduardo Azevedo, Bruno Jullien, Renato Gomes, Jacques Crémer, Matthew Gentzkow, Jon Levin, Jesse Shapiro, Mike Whinston and many workshop and conference participants for helpful comments as well as Jonathan Baker, Liran Einav, Jean Tirole and Preston McAfee for excellent formal discussions of the paper. We acknowledge the financial support of the NET Institute, the research program on intellectual property at the Institut d'Economie Industrielle (IDEI), the Marion Ewing Kauffman Foundation and the Fundação para a Ciência e a Tecnologia which funded the research assistance of Kevin Qian. We are especially grateful to Ben Handel, Igal Hendel and Mike Whinston for providing us the summary statistics that allow our calibration.

[†]Department of Economics and Nuffield College, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK; andre.veiga@economics.ox.ac.uk.

[‡]Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago IL, 60637 USA and Toulouse School of Economics, 21 allée de Brienne, Toulouse France 31000; weyl@uchicago.edu.

1 Introduction

Insurers choose non-price characteristics of their products, such as co-insurance rates and deductibles, to selectively attract the most valuable customers. For instance, in Rothschild and Stiglitz (1976) (henceforth RS) highlight the incentive of insurers to "skim the cream" from a rival by lowering both insurance coverage and price. The least risky consumers value low prices more than high coverage, so they are disproportionately attracted by such a contract, leading to a "death spiral" towards zero insurance. However, such a strategy is less attractive if risk-averse consumers, who Finkelstein and McGarry (2006) and Fang, Keane and Silverman (2008) find are typically less costly, are also repelled. Despite this, existing analyses of product design in selection markets rule out rich multidimensional heterogeneity, thereby either ruling out sorting incentives entirely under monopoly (Stiglitz, 1977) or ensuring cream-skimming is so extreme as to destroy competitive equilibrium (RS; Riley, 1979). To fill this gap, in this paper we provide a sharp characterization of sorting incentives under a range of market structures in the presence of continuous multidimensional heterogeneity. This richer setting makes possible positive insurance (pooling) competitive equilibria, quantifies how market power can benefit product design and can be calibrated based on simple reduced-form statistics.

We gain this tractability by assuming that the set of contracts offered by an insurer can be described by a vector. In particular, we do not allow for fully non-linear pricing schemes described by a function as in Stiglitz (1977) and Rochet and Choné (1998). Our analysis is based on the characterization of the marginal incentive of an insurer to use a non-price product characteristic to sort in favor of low-cost consumers, which is the product of two components. The first component is the density of marginal consumers and captures how *many* buyers change their purchase decision when a product characteristic changes. The second component captures *which* buyers adopt, and which drop, the product. This second term equals the covariance, within the set of marginal buyers, between the marginal willingness-to-pay (WTP) for the non-price product characteristic and the cost of consumers to the firm. This covariance is zero in one-dimensional models because the set of marginal types is a singleton.

Although this covariance is endogenous, it can often be characterized to yield several economically important results. For instance, it is signed when marginal WTP and cost can be written as monotonic functions of an (endogenous) common index. This occurs naturally in environments with non-linear pricing or when heterogeneity is two-dimensional. We illustrate our results using a simple parameterized model where individuals differ in their absolute risk aversion and face normally distributed wealth shocks with different means. We show that increasing the generosity of a linear actuarial rate attracts the worse risks because insurance motives are concave in the actuarial rate while risk-transfer motives are linear. We also show (computationally) that more generous deductibles and indemnity caps similarly attracts the worse risks, but increasing the probability that a shock is covered by insurance (plan comprehensiveness) attracts the best risks. These results may help explain empirically-observed contract terms.

We then extend the model to a simple competitive setting à la Hotelling (1929). We assume firms are symmetrically differentiated, offer one contract each, that the market is covered and that individual heterogeneity is orthogonal to the Hotelling preferences over insurers. Moreover, we focus on local deviations from a symmetric (that is, pooling) equilibrium. This approach allows us to show that, in the limit of undifferentiated Bertrand competition, equilibrium insurance has a simple structure: the covariance term must vanish because the density of switching consumers becomes arbitrarily large.

When consumers differ only in their risk type (RS; Riley, 1979), the covariance is always positive at any positive insurance level and thus the incentive to "skim the cream" undermines any positive level of insurance. Multidimensional heterogeneity re-introduces the possibility of positive insurance if value for insurance is sufficiently negatively correlated with mean risk, as found empirically by Finkelstein and McGarry (2006). In particular if β is the OLS regression coefficient of insurance value on mean risk in the entire population, a competitive equilibrium with positive coverage requires $\beta < -1$. When β is sufficiently posistive, on the other hand, the unique equilibrium features zero insurance. For an intermediate range of this correlation, no equilibrium exists for the same reasons as in RS.

The incentive to reduce coverage near full insurance is always positive because marginal insurance value vanishes near perfect insurance. Thus, full insurance is never an equilibrium despite it being socially optimal in the absence of moral hazard. By contrast, a monopolist has a socially optimal incentive to provide insurance (albeit at an excessive price) because it internalizes the socially harmful effects of cream-skimming. Therefore, as market power increases, the equilibrium level of insurance approaches the socially optimal level. This appears to contradict existing results (Armstrong and Vickers, 2001; Rochet and Stole, 2002) that competition drives out second-degree price discrimination leading to efficiency. The reason is that these models consider ex-post contracting where *only moral hazard* (ex-post efficient consumption) is relevant while we consider an ex-ante situation where *only insurance* is relevant. Thus, the effect of market power depends on whether making individuals bear their social costs is desirable or undesirable.

We calibrate our analysis with summary statistics from Handel, Hendel and Whinston (2013), who adopt a similar CARA-Normal framework to measure the joint distribution of types.[1] In their data, $\beta \approx 13$, implying no insurance is impossible at a pooling equilibrium. Moreover, investigating second-order conditions shows that zero insurance is not an equilibrium either and thus the RS non-existence result holds in this setting. Our approach thus uses a simple empirical summary statistic to explain rigorously the extreme adverse selection and market collapse found in Handel, Hendel and Whinston (2013)'s computational structural analysis. While market power could restore positive insurance in equilibrium, a mark-up of more than 50 times cost would be necessary to achieve an actuarial rate of 80% and a mark-up of almost 90% is necessary for an equilibrium to even exist.

We then consider, as robustness checks, including moral hazard and allowing for an imperfectly covered market. Both factors reduce the socially optimal level of insurance, but neither qualitatively changes any conclusions of our calibrated example.

The remainder of the paper is organized as follows. Section 2 presents our results in a simple CARA-normal setting. The emphasis is on the discussion of our substantive assumptions, the economic content of the model and sketches of the proofs. Section 3 contains the empirical calibration. Section 4 contains rigorous proofs and generalizations. Section 5 contains robustness checks. We conclude in Section 6, where we discuss the applicability of our approach beyond insurance markets and policy implications of our analysis. Less instructive and longer proofs are collected into the appendices following the main text.

2 A Simple Model of Sorting

This section develops the main results of the paper in a simple CARA-Normal setting where insurance is characterized by a linear actuarial rate and consumers differ in risk and risk aversion. We describe the ingredients of the model, consider a monopoly insurer, extend the model to a competitive setting and finally discuss the effect of market power on welfare.

2.1 Setup

A unit mass of individuals face an uncertain insurable wealth shock. A monopoly commits to absorbing a share $x \in [0, 1]$ of the shock for a price $p \in \mathbb{R}_+$. Thus, the actuarial rate x captures the quality of coverage. Consumers have constant absolute risk aversion (CARA) preferences, so maximize the expected value (over realizations of the shock) of $-e^{-aw}$, where w is final wealth and $a > 0$ is absolute risk aversion. Wealth shocks are normally distributed with mean $\mu > 0$ and variance $\sigma^2 > 0$. The vector (μ, a) constitutes individual heterogeneity and is not contractible.

[1] We thank these authors for generously sharing with us these statistics.

Initial wealth is zero. Expected utility is $-\exp\left\{a\mu + \frac{a^2\sigma^2}{2}\right\}$ without insurance, while with insurance it is $-\exp\left\{a(1-x)\mu + ap + \frac{(1-x)^2 a^2\sigma^2}{2}\right\}$. The willingness-to-pay (WTP) of a consumer of type (μ, a) for coverage x is the price at which the two expected utilities are equal. It is convenient to define an individual's value for insurance as $v \equiv a\sigma^2 > 0$, and to consider heterogeneity as captured by (μ, v). Then, WTP is $u(x, \mu, v) \equiv x\mu + \frac{1-(1-x)^2}{2}v$.

Buyers are those for whom WTP exceeds price: $u(x, \mu, v) > p$. For $x > 0$, WTP is strictly increasing in μ, so we can equivalently define the set of buyers as those for whom $\mu > \mu^\star(p, x, v) \equiv \frac{1}{x}\left(p - \frac{1-(1-x)^2}{2}v\right)$. Similarly, the set of marginal buyers are those for whom $u(x, \mu, v) = p \Leftrightarrow \mu = \mu^\star(p, x, v)$.

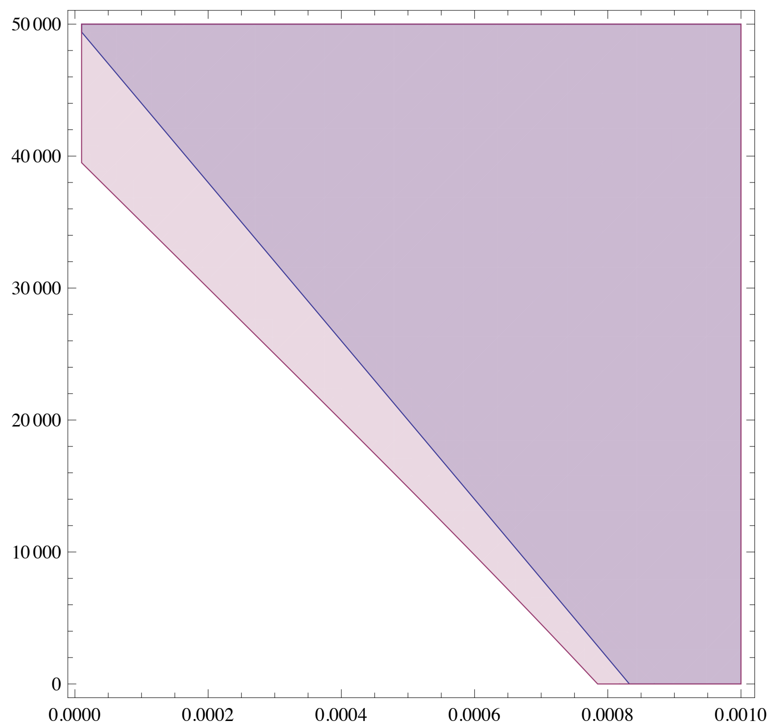

Figure 1: Set of buyers for $\mu \in [0, 10^4]$ (vertical axis), $a \in [10^{-5}, 10^{-3}]$ (horizontal axis), and $\sigma^2 = 10^4$. Price is $p = 40000$. Coverage is $x = 0.8$ (blue) and $x = 1$ (red).

The insurer knows (μ, v) is distributed according to the atomless and full support probability density function $f(\mu, v) : [\underline{\mu}, \overline{\mu}] \times [\underline{v}, \overline{v}] \to \mathbb{R}_{++}$. The expected cost to the risk-neutral insurer of providing coverage x to an individual of type μ is $c(x, \mu) \equiv x\mu$. The usefulness of defining $\mu^\star(p, x, v)$ is that we can express the quantity of buyers as the standard iterated integral: $Q(p, x) \equiv \int_{\underline{v}}^{\overline{v}} \int_{\mu^\star(p,x,v)}^{\overline{\mu}} f(\mu, v)\, d\mu dv$. It is intuitive (and we show below) that $\frac{\partial Q(p,x)}{\partial p} < 0$, so there exists a price $P(x, q)$ which solves $Q(P(x, q), x) \equiv q$ for any quantity q. This allows us to follow Spence (1975) in defining the firm's profits as a function of quantity q and coverage/quality x. As we argue below, it is useful to consider changes in x while the total number of buyers is held fixed, which is straightforward in this setup. Profit is

$$\Pi(q,x) \equiv qP(q,x) - \int_{\underline{v}}^{\overline{v}} \int_{\mu^\star(P(q,x),x,v)}^{\overline{\mu}} c(x,\mu) f(v,\mu) \, d\mu dv.$$

While the set of of buyers is two-dimensional, the set of marginal consumers is a one-dimensional curve in $\mathbb{R}^2$, as in Figure 1. Our results require us to define the notion of an expectation conditional on this set of marginal consumers.[2] Here, we define the measure of the marginal set as the response of quantity to price: $M \equiv -\frac{\partial Q(p,x)}{\partial p}$. Moreover, letting $\mu^\star = \mu^\star(p,x,v)$, for an integrable function $\zeta(x,\mu,v)$, expectation conditional on the margin is $\mathbb{E}_u[\zeta(x,\mu,v) \mid \mu = \mu^\star] = \frac{\int_{\underline{v}}^{\overline{v}} \zeta(x,\mu^\star,v) f(\mu^\star,v) dv}{\int_{\underline{v}}^{\overline{v}} f(\mu^\star,v) dv}$. The conditional covariance $\mathrm{Cov}_u[\cdot \mid \mu = \mu^\star]$ is defined similarly.

2.2 Monopoly

Given this setup, a profit-maximizing monopoly's First Order Condition (FOC) is captured by the following result, where functional arguments are omitted for simplicity and the symbol " $'$ " denotes partial derivative with respect to x.

Proposition 1. *A necessary FOC for the profit maximizing choice of x is*

$$-\underbrace{q\mathbb{E}[c' \mid \mu > \mu^\star]}_{\textit{Cost}} + \underbrace{q\mathbb{E}_u[u' \mid \mu = \mu^\star]}_{\textit{Spence term}} = \underbrace{M\mathrm{Cov}_u[u', c \mid \mu = \mu^\star]}_{\textit{Sorting}},$$

Proof. $\frac{\partial \Pi}{\partial x} = q\frac{\partial P(x,q)}{\partial x} - \int_{\underline{v}}^{\overline{v}} \int_{\mu^\star}^{\overline{\mu}} c'(x,\mu) f(v,\mu) \, d\mu dv - \int_{\underline{v}}^{\overline{v}} \frac{\partial \mu^\star}{\partial x} c(x,\mu^\star) f(v,\mu^\star) \, dv$ by the Leibniz Rule for differentiating under the integral sign. We obtain $\frac{\partial P}{\partial x} = -\frac{\partial Q/\partial x}{\partial Q/\partial p}$ by applying the Implicit Function Theorem (IFT) to $Q(P(x,q),x) = q$. We obtain $\frac{\partial \mu^\star}{\partial x} = -\frac{\partial u/\partial x}{\partial u/\partial \mu} = -\frac{1}{x}\frac{\partial u}{\partial x}$ by applying the IFT to $u(x,v,\mu^\star(x,v,p)) \equiv p$. We then use $\int_{\underline{v}}^{\overline{v}} \int_{\mu^\star}^{\overline{\mu}} c'(x,\mu) f(v,\mu) \, d\mu dv = q\mathbb{E}[c'(x,\mu) \mid \mu > \mu^\star]$, $M = -\frac{\partial Q(p,x)}{\partial p} = \frac{1}{x}\int_{\underline{v}}^{\overline{v}} f(\mu^\star,v) \, dv$, and the definition of $\mathbb{E}_u[\cdot \mid \mu = \mu^\star]$. For details, see Proposition 1. □

This characterization decomposes the insurer's marginal incentive to raise coverage (x) into three components. The first two are familiar. First, the monopolist loses the average increase in the cost of buyers ($\mathbb{E}[c' \mid \mu > \mu^\star]$) multiplied by the number of buyers (q), which follows mechanically from the additional share of the shock absorbed by the insurer. Second, increasing x causes the number of buyers to increase. In order to keep q fixed, price

[2] Typically, defining this operator requires some care to define an economically useful measure on this set. However, in this simple example, such complications do not arise and therefore we delay discussing them until Section 4, where we follow the approach in Veiga and Weyl (2013).

implicitly increases to all buyers (q) by the average marginal WTP of marginal consumers ($\mathbb{E}_u [u' \mid \mu = \mu^\star]$), as first observed by Spence (1975).[3]

The third effect is the focus of our analysis. Increasing x not only changes the *number* of buyers, but also their *composition* since x attracts some individuals (those with large u') more than others. This alters the composition of buyers sorting in favor of marginal buyers with high u'. All buyers pay the same price p, so the impact that sorting has on profit depends on whether the marginal consumers most strongly attracted by x tend to be also those with particularly high cost (c). If this is the case, sorting increases the insurer's costs. It is therefore natural that the effect of sorting on profit is captured by the covariance, among marginal consumers, between the cost of providing them with the product and their WTP for an increase in x: $\text{Cov}_u [u', c \mid \mu = \mu^\star]$. If $\text{Cov}_u [u', c \mid \mu = \mu^\star] > 0$, then additional coverage tends to decrease profit, which we call "adverse sorting." If $\text{Cov}_u [u', c \mid \mu = \mu^\star] < 0$, there is "advantageous sorting."

It is useful to distinguish sorting from what is commonly referred in the literature as selection, which captures the response of average cost to a change in the quantity of buyers. For instance, "adverse selection" occurs when the buyers with higher *level of WTP* for insurance are most costly so, as the number of buyers increases (as price falls), average cost also falls as in Akerlof (1970) and Einav, Finkelstein and Cullen (2010). Sorting refers to changes in average cost holding fixed the number of buyers, through changes in the composition of buyers. This depends on whether buyers with higher *marginal* (not *total*) *WTP* for insurance are most costly *conditional on having a given total WTP* (not *unconditional*). As we will see in the next subsection, this implies that sorting may behave quite differently from selection.

Importantly, the sorting term vanishes when consumers are homogeneous in their marginal utility for x or in their cost. It also vanishes when there is a unique type of marginal consumer, as in almost all models with unidimensional types. Finally, notice that this effect is scaled by $M = -\frac{\partial Q}{\partial p}$ which captures the average responsiveness of demand. Loosely speaking, in Figure 1, M captures translations of the line that defines the marginal set, while $\text{Cov}_u [u', c \mid \mu = \mu^\star]$ captures rotations of that line.

2.3 Signing the sorting incentive

The sorting term may seem an unlikely object to focus on since it is endogenously determined by q and x. However in this subsection we show that its sign can be ascertained directly and exogenously in a variety of contexts to yield results of economic importance. We then

[3] The effect was independently and nearly simultaneously identified by Sheshinski (1976), although we follow the convention of associating it with Spence.

use these results to determine the sorting effect of other product design instruments such as a deductible or an indemnity cap. In this section, we do not use the notation $v \equiv a\sigma^2$, since it is not useful when considering instruments other than a linear actuarial rate.[4]

Part of the reason why we focus on the simple setting of two-dimensional types is that, in this case, the set of marginal consumers is a one-dimensional curve. Then, conditional on this set, the functions $u'(x, \mu, a)$ and $c(x, \mu)$ are effectively univariate since one dimension of type can be expressed as a function of the other (along that curve). This is done by the function $\mu^\star(p, x, a)$. Then, if two univariate functions are co- or anti-monotone, the covariance between them is signed.[5] This is stated in the following result which, importantly, applies for instruments other than a linear actuarial rate.

Proposition 2. *Let* $S = S(p, x, a, \mu^\star(p, x, a)) \equiv \frac{\partial^2 u/\partial x \partial a}{\partial u/\partial a} - \frac{\partial^2 u/\partial x \partial \mu}{\partial u/\partial \mu} \mid_{\mu=\mu^\star(p,x,a)}$. *If* S *is signed, it has the opposite sign of* $\mathrm{Cov}_u[u', c \mid \mu = \mu^\star]$.

Proof. Since $\frac{\partial u}{\partial \mu} > 0$ and $\frac{\partial u}{\partial a} > 0$, within a fixed level of WTP we must have $\frac{d\mu^\star}{da} < 0$. Thus $\frac{dc(x,\mu^\star(p,x,a))}{da} = x\frac{\partial \mu^\star}{\partial a} < 0$. Since c is monotonic within the margin, the sorting incentive depends only on $\frac{d}{da}\left(\frac{\partial u(x,a,\mu^\star(p,x,a))}{\partial x}\right) = \frac{\partial^2 u}{\partial x \partial a} + \frac{\partial^2 u}{\partial x \partial \mu}\frac{\partial \mu^\star}{\partial a}$. Since $\frac{\partial \mu^\star}{\partial a} = -\frac{\partial u/\partial a}{\partial u/\partial \mu}$ and $\frac{\partial u}{\partial a} > 0$, this has the same sign as S. If these two terms have the same sign, the covariance is positive, and vice-versa. For details, see Theorem 3 □

All marginal buyers are indifferent so, within this set, higher risk μ corresponds to lower risk aversion a. Moreover, WTP is constant within this set. Therefore, sorting is adverse when risk (μ) raises marginal WTP relative to total WTP more than risk aversion (a) raises marginal WTP relative to total WTP. Thus sorting is, intuitively, in favor of whichever dimension of heterogeneity more strongly impacts *marginal* compared to *average* utility for insurance. Since the curvature of a function captures the difference between its margin and its average, sorting favors the dimension of type which generates a more convex demand for x. Theorem 3 below shows that this logic holds more generally: if θ_i is the type dimension that increases marginal WTP of x most rapidly relative to the rate it increases WTP and θ_i is also the dimension that increases cost most rapidly relative to the rate at which it increases WTP, then x sorts adversely.

The following corollary is immediate.

Corollary 1. *An actuarial rate sorts adversely (*$\mathrm{Cov}_u[u', c \mid \mu = \mu^\star] \geq 0$*).*

Proof. $\frac{\partial^2 u/\partial x \partial a}{\partial u/\partial a} - \frac{\partial^2 u/\partial x \partial \mu}{\partial u/\partial \mu} = -\sigma^2 \frac{x}{2} \leq 0$. □

[4]However, in the case of the linear actuarial rate, all results stated in this section remain true if a is replaced by $v = a\sigma^2$ even if $\sigma^2$is allowed to be heterogeneous across individuals.

[5]For a simple proof, see Schmidt (2003).

When x is an actuarial rate, we have $u(x,\mu,a) \equiv x\mu + \frac{1-(1-x)^2}{2}a\sigma^2$. The demand induced by μ is linear, while that induced by a is concave in x: there is a constant incentive to transfer mean risk, but insurance value has decreasing returns. Recall that, among marginal buyers, high μ must correspond to low a and vice-versa. For profitable marginal individuals (high a, low μ) additional x has rapidly decreasing marginal benefit, but for unprofitable marginal individuals (high μ, low a) each unit of x has only slowly declining marginal benefit. This is why, in Figure 1, raising x attracts bad risks (high μ, low a) more intensely than good risk (high a, low μ).

It is also useful to consider the sign of the sorting incentive for instruments other than an actuarial rate. Since these cases are less analytically tractable, we proceed computationally, providing details in Appendix A. For realism, we consider cases where reimbursements occur only for negative wealth shocks and study three instruments: standard actuarial rate (a share x of the shock are covered, again only for negative shocks), a deductible (value of shock above a threshold x are covered), and an indemnity cap (shock is covered up to a value x). Finally, we consider a measure of comprehensiveness, where there is a probability x that a shock is covered.

Let $f(l,\mu,\sigma^2)$ denote the Gaussian density of shocks of an individual with type μ. The insurer's policy prescribes a payment $G(l,x)$ when the individual incurs a loss l. Consumers have CARA utility as above. The (homogeneous) initial wealth is w_0. Without insurance, final wealth is $w_N = w_0 - l$. With insurance, it is $w_I = w_0 - l + G(l,x) - p$. Expected surplus from insurance is $U(x,a,\mu,p) = \int \left[e^{-aw_N} - e^{-aw_I}\right] f(l,\mu,\sigma^2)\,dl$. Our goal is to compute $S = \frac{\partial^2 u/\partial x \partial a}{\partial u/\partial a} - \frac{\partial^2 u/\partial x \partial \mu}{\partial u/\partial \mu}\,|_{\mu=\mu^\star(p,x,a)}$. Appendix A shows how to express S (and thus WTP) in terms of the function $U(x,a,\mu,p)$. Computationally, we generate draws of (p,x,a), then compute the value of $\mu^\star(p,x,a)$ that makes an individual marginal to purchasing insurance, and then evaluate $S(p,x,a,\mu^\star(p,x,a))$. We calibrate our analysis to the setting of Handel, Hendel and Whinston (2013): σ^2 is set to the order of magnitude of its mean value (10^8) in their data; $a \in [10^{-5}, 10^{-3}]$ and $\mu \in [0, 5\times 10^4]$, since these are approximately the ranges of risk aversion and risk in their data; and we focus on the relatively high levels of insurance seen in their data. In each case we assume insurance is perfect along every dimension other than the one we consider. The computational output justifying the following claims can be found in Appendix A.

Claim 1. Additional insurance sorts adversely ($\mathrm{Cov}_u\left[u', c \mid \mu = \mu^\star\right] > 0$) when:

x is an actuarial rate, so $G(l,x) = \max\{0, xl\}$.

x is a deductible, so $G(l,x) = \max\{0, l-x\}$.

x is an indemnity cap, so $G(x,l) = \min\{\max\{0,l\}, x\}$.

We find that, for all these instruments, additional insurance sorts adversely (in the case of a deductible, increasing x corresponds to less generous insurance). While we present results that focus on relatively generous levels of insurance, this seems true for any level of these instruments. This may help explain why insurers have an incentive to reduce coverage from full insurance, since this would sort advantageously. Notice that this does not predict that insurance would be reduced to zero in these dimensions, since a monopoly must balance the incentive to sort with the incentive (from the two terms on the left hand side in Proposition 1) to exploit the gains from trade of more insurance.[6]

Finally, we consider a case when the insurer covers the (full) loss l with probability x, which we refer to as "comprehensiveness." Final wealth without insurance is still w_N. With insurance, with probability x final wealth is $w_{I+} = w_0 - p$, and with probability $1 - x$ it is $w_{I-} = w_0 - l - p$. Then, surplus from insurance is $U(x, a, \mu, p) = \int \left[e^{-aw_N} - xe^{-aw_{I+}} - (1-x) e^{-aw_{I-}}\right] f(l, \mu, \sigma^2)\, dl$. We refer to x as the *comprehensiveness* of the plan in this case.[7]

Claim 2. Comprehensiveness sorts advantageously ($\mathrm{Cov}_u [u', c \mid \mu = \mu^\star] < 0$).

In this case, additional insurance (increasing x) sorts advantageously even at full insurance ($x = 1$). Given that gains from trade would also tend to raise coverage towards full insurance, this helps explain why insurers rarely offer contracts where the probability of reimbursement is random (where some conditions are not covered), since increasing this probability always favors the risk averse and therefore always sorts advantageously.

2.4 Competition

RS and a large literature building on their work discuss the existence and characterization of equilibria in competitive insurance markets with endogenous product design. The general message of this literature is that equilibrium may not exist or features very low levels of insurance. In their conclusion, RS suggest that frictions, such as multidimensional heterogeneity or imperfect competition, might mitigate these extreme results. Consistent with this hypothesis, Einav, Finkelstein and Cullen (2010) and others (see Einav and Finkelstein (2011) for an excelent summary of the literature) have shown how multidimensional heterogeneity can in practice mitigate or even reverse the welfare effects of selection on the number of individuals in the market (à la Akerlof (1970), that is, without endogenous product design) .

[6]With additional information on the distribution of heterogeneity these incentives could be balanced and an optimum derived. However this would require detailed estimation that is beyond the scope of our qualitative analysis here.

[7]This is equivalent to: 1) a condition is drawn from a $\mathcal{U}[0,1]$ distribution; 2) an associated loss drawn from a $\mathcal{N}(\mu, \sigma^2)$. Then x captures the share of conditions covered. Notice that all conditions have the same likelihood and loss distribution.

Despite this, to our knowledge no paper has quantified the effect of such frictions on the equilibrium outcomes in the RS environment where non-price product characteristics are endogenous.[8] In this subsection, we show that our characterization of the sorting incentive allows us to shed light on the interaction of both these frictions with product design. In particular we confirm RS's intuitions that these frictions can reintroduce the possibility of perfectly competitive pooling equilibria with positive insurance, though in the next section we show calibrationally that these frictions have to be extremely large to achieve this.

We extend the framework above to a simple Hotelling environment. We consider two insurers, indexed by $i \in \{0, 1\}$, where i captures location on a Hotelling line. Insurer i chooses a linear actuarial rate x_i and a price p_i. We return to using the notation $v \equiv a\sigma^2$. We assume the two insurers are identical apart from their Hotelling location, so cost is $c(x, \mu) = x\mu$ for either insurer. Consumers have an additional dimension of heterogeneity, $b \in [0, 1]$, which captures preferences over insurers. An individual with type b incurs a cost tb by purchasing from firm 0 and a cost $t(1 - b)$ from purchasing from firm 1, and this cost is fungible with price. Thus, t captures market power. WTP for coverage x is as in Sub-Section 2.2.

We assume that the market is covered: every consumer purchases from one of the insurers. Doing so allows us to focus on the effect of competition on quality (x), abstracting from its effect on the number of consumers covered. Moreover, we are particularly interested in the competitive limit where $t \to 0$, in which case the results of a covered and uncovered market are qualitatively similar, as we argue in Subsection 5.2. Finally, it is often the case that a government mandate implies a covered market, and indeed this is a common assumption in the literature.[9]

[8] In fact, few papers have studied models of competitive product design that include either of these features. While two recent papers have considered purchasers who are heterogeneous in two dimensions (Wambach, 2000; Smart, 2000), both assume a small number (four) of discrete types and provide very partial characterizations of equilibrium in specific settings. To our knowledge no work has considered equilibrium product design with imperfectly competitive firms, at least in the standard industrial organization sense (e.g. quantity competition or differentiated products).

This contrasts sharply with extensive empirical evidence that multidimensional heterogeneity and market power are crucial to understanding the functioning of insurance markets. Dozens of recent papers surveyed by Einav, Finkelstein and Levin (2010) find that multidimensional heterogeneity is crucial to explain observed behavior in insurance markets.Chiappori et al. (2006) provide empirical evidence that they argue is hard to explain unless firms have market power, writing "(O)ur findings...suggest that more attention should be devoted to the interaction between imperfect competition and adverse selection on risk aversion...there is a crying need for such models." The limitations of the theoretical treatment of market power and multidimensional heterogeneity has thus become widely perceived as an important check on progress in this empirical literature. For example, Einav and Finkelstein (2011) write, "On the theoretical front, we currently lack clear characterizations of the equilibrium in a market in which firms compete over contract dimensions as well as price, and in which consumers may have multiple dimensions of private information (like expected cost and risk preferences)." A major goal of this paper it to provide such a characterization.

[9] For instance, a covered market is mandated under the recent Affordable Care and Patient Protection

Individuals purchase from the insurer that offers the highest utility net of transportation costs. We define $b^{\star} \equiv \frac{1}{2t}\left(u(x_0,\mu,v) - p_0 - (u(x_1,\mu,v) - p_1)\right) + \frac{1}{2}$. Then buyers of insurer 0 are those for whom $b < b^{\star}$, buyers of insurer 1 have $b > b^{\star}$, and marginal individuals have $b = b^{\star}$.

We assume that b is distributed uniformly on the interval $[0,1]$, independent of (μ, v), so the joint distribution of heterogeneity is $f(\mu, v)$. This is a natural benchmark, since there is no obvious relationship between these variables in a market with symmetrically differentiated firms. Then $\pi_0 = \int_{\underline{v}}^{\overline{v}} \int_{\underline{\mu}}^{\overline{\mu}} \int_0^{b^{\star}} (p_0 - x_0\mu) f(\mu, v)\, db d\mu dv$ is the profit of insurer 0. Let Q_i be the number of buyers from insurer i.

We will focus on local deviations pooling equilibria (LDPE). Intuitively, we will assume both insurers choose $x_0 = x_1 = x$ and $p_0 = p_1 = p$, and we consider as an equilibrium a point (x,p) where both firms' First and Second Order Conditions (SOC) for profit maximization are satisfied. This concept is defined rigorously in Subsection 4.4.

It is useful to compare this approach to that of RS. In that paper, there cannot be a competitive pooling equilibrium, whereas this is possible in our setting. Moreover, in RS with a continuum of types, there is always a local deviation from all candidate positive insurance competitive equilibria, and a non-local deviation from the candidate equilibrium with zero insurance (Riley, 1979). While the correct notion of equilibrium in insurance markets is still a topic of debate, many of the intuitions present in the literature (including non-existence) still arise with the more tractable concept of LDPE. Additionally, while an insurer could plausibly identify a non-local deviation in the simpler environment of RS, this is less plausible in a setting with continuous multidimensional heterogeneity, where the profitability of a given deviation will depend on the joint distribution of multidimensional types in parts of the type space that are not revealed by behavior local to the LDPE.[10]

Focusing on LDPE allows for the following significant simplifications. In a symmetric equilibrium, the market is split evenly, so $b^{\star} = Q_0 = Q_1 = \frac{1}{2}$ and $M = -\frac{\partial Q_0}{\partial p_0} = -\frac{\partial Q_1}{\partial p_1} = \frac{1}{2t}$. The set of marginal consumers has the same composition as the set of all consumers, which dispenses with the need to consider conditional expectations and covariances. Moreover, under these assumptions, a profit maximizer catering to its marginal consumers is simulta-

Act in the United States. The assumption is mirrored in several recent papers such as Rochet and Stole (2002) and Handel, Hendel and Whinston (2013) .

[10] Following on the non-existence results of RS, several papers have introduced modified concepts of equilibrium. Riley (1979) showed that non-existence is generally the case with continuous heterogeneity although, from Dasgupta and Maskin (1986), equilibrium still exists in mixed strategies. Miyazaki (1977) and Wilson (1977) suggest notions of equilibrium which allow for more complex reactions on the part of other firms. Like this literature, our paper considers an alternative solution concept, though it does so for tractibility not existence purposes and largely maintains RS's non-existence conclusions in the circumstances they arise in their work.

neously catering to its average consumers, so we abstract from the Spence (1975) distortion mentioned in Section 4, thereby focusing on how x is determined by the sorting incentive.

It is convenient to define $\phi = \phi(x, v) \equiv u(x, \mu, v) - c(x, \mu) = \frac{1-(1-x)^2}{2}v$, the social surplus from insurance or the value a consumer places on insurance beyond the off-loading of mean risk to the insurer. We can now determine the competitive FOC and SOC.

Proposition 3. *The competitive FOC is* $\mathbb{E}[\phi'] = \frac{1}{t}\mathrm{Cov}[u', c]$ *and holds at a unique* $x \in (0, 1)$. *The SOC is* $2t\mathbb{E}[\phi''] + \mathbb{E}[\phi']^2 < 2\mathrm{Cov}[u'', c] + 4\mathrm{Cov}[u', c']$.

Proof. The FOC can be derived from Proposition 1, using $Q = \frac{1}{2}$, $M = \frac{1}{2t}$, making the expectation and covariance unconditional: $(1 - x)\mathbb{E}[v] = \frac{1}{t}\mathrm{Cov}[\mu + v(1 - x), x\mu]$. Defining, $\alpha \equiv \frac{\mathbb{E}[v]}{\mathbb{V}[\mu]} > 0$ and $\beta \equiv \frac{\mathrm{Cov}[v,\mu]}{\mathbb{V}[\mu]}$, the FOC can be written as $\frac{\alpha t}{x} - \frac{1}{1-x} = \beta$. The function $\frac{\alpha t}{x} - \frac{1}{1-x}$ has range $\mathbb{R}$ for domain $x \in (0, 1)$ and is continuous strictly decreasing in x. Therefore, for any $\beta \in \mathbb{R}$, there exists a unique $x \in (0, 1)$ where the FOC holds. For details, see Proposition 4. The SOC is derived in Appendix C. □

On the one hand, increasing x generates gains from insurance $\mathbb{E}[\phi']$, which the firm can perfectly capture because of the absence of a Spence distortion. On the other hand, firms use x to sort for the most valuable consumers thereby skimming the cream from its rival. The relative weight on these two forces is determined by market power, since the sorting term is multiplied by $\frac{1}{t}$. When competition is intense (in the sense of undifferentiated Bertrand, t low), a large weight is placed on selection, as in the perfectly competitive world of RS where a small change in x creates a large amount of cream-skimming.

Regarding the SOC, notice that $\mathbb{E}[\phi''] = -\mathbb{E}[v] < 0$. Because the other terms are bounded, the SOC is satisfied for sufficiently large x. Therefore, we focus below on a discussion of the SOC in the competitive limit as $t \to 0$, where this condition is less likely to hold, given that, as we show in the next subsection, the unique candidate pooling equilibrium rises monotonically in t.

Proposition 4. *Let* $\beta \equiv \frac{\mathrm{Cov}[v,\mu]}{\mathbb{V}[\mu]}$ *and* $\gamma \equiv \frac{1}{4}\frac{\mathbb{E}[v]^2}{\mathbb{V}[\mu]} \geq 0$. *In the limit where* $t \to 0$*:*

An LDPE requires $\mathrm{Cov}[u', c] = 0$ *and* $(1 - x)^2\gamma < \left(1 - \frac{3}{2}x\right)\beta + 1$.

If $-1 < \gamma - 1 < \beta$, *the unique LDPE has* $x = 0$.

If $2\gamma < -(\beta^3 + \beta^2)$, *which implies* $\beta < -1$, *the unique LDPE has* $x = 1 + \frac{1}{\beta}$.

For any $\gamma > 0$, *there is a range of* β *for which there is no LDPE, and this range increases in* γ. *However, for* $\gamma > 0$, *there is a unique* $\tilde{\beta} < -1$ *such that there exists an LDPE if* $\beta < \tilde{\beta} < -1$.

Proof. From the FOC, as $t \to 0$, we must have $\mathrm{Cov}[u', c] \to 0$ since $\mathbb{E}[\phi'] = (1 - x)\mathbb{E}[v]$ is bounded. This is equivalent to $x\mathbb{V}[\mu](1 + (1 - x)\beta) = 0$. The SOC is derived in Appendix

C and, when $t \to 0$, the SOC requires $(1-x)^2 \frac{\mathbb{E}[v]^2}{4\mathbb{V}[\mu]} < -\frac{x\mathrm{Cov}[v,\mu]}{2\mathbb{V}[\mu]} + (1-x)\frac{\mathrm{Cov}[v,\mu]}{\mathbb{V}[\mu]} + 1$. The FOC always holds at $x = 0$, but for the SOC to hold at $x = 0$ we require $-1 < \gamma - 1 < \beta$. The FOC holds at $x = 1 + \frac{1}{\beta}$ if $\beta < -1$, in which case the SOC requires $2\gamma < -\beta^3 - \beta^2$. By the IFT, the threshold defined by $\beta = (\gamma - 1)$ increases with γ ($\frac{\partial \beta}{\partial \gamma} = 1$), while the threshold defined by $2\gamma = -(\beta^3 + \beta^2)$ decreases with γ ($\frac{\partial \beta}{\partial \gamma} = -\frac{1}{2}(3\beta^2 + 2\beta) < 0$ and $\beta < -1$). For $\beta < -1$, we have $-\beta^3 - \beta^2$ positive, with range $\mathbb{R}_+$ and strictly decreasing. □

A competitive LDPE must have zero sorting incentive, which is a generalization of the logic of Rothschild and Stiglitz (1976) to a context of multidimensional types. Cream skimming incentives are proportional to M and in an LDPE demand is infinitely responsive ($M = \frac{1}{2t} \to \infty$) since, loosely speaking, all consumers are marginal. Thus, any incentive to sort will cause firms to deviate from the LDPE.

The set of actuarial rates possible at competitive equilibria can be computed solely on the basis of a simple moment of the distribution of types in the population. Indeed, the unique possible positive rate is $x = 1 + \frac{1}{\beta}$ where $\beta = \frac{\mathrm{Cov}(\mu,v)}{\mathbb{V}(\mu)}$ is the uncontrolled OLS regression coefficient of μ on v in the entire population. This provides a simple condition to take to data in order to calibrate how much insurance would be provided at a competitive equilibrium, which we do in Section 3.

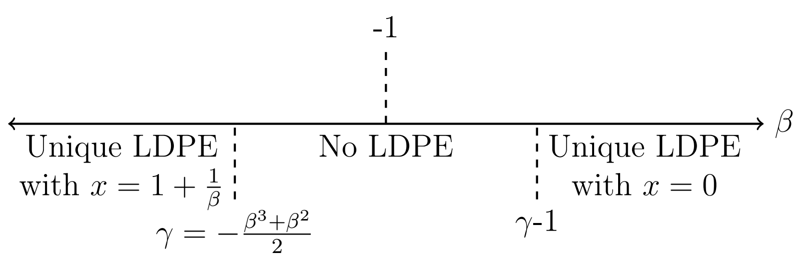

Figure 2: Full characterization of LDPE for various values of β; increasing values of β to the right.

A positive insurance competitive equilibrium is possible only if $\beta < -1$. This occurs when risk aversion is sufficiently negatively correlated with risk as in the data of Finkelstein and McGarry (2006), so that sorting is advantageous at $x = 0$. Importantly, this results requires a setting with multidimensional heterogeneity. With a single dimension of type μ, we have $\mathrm{Cov}[v, \mu] = \beta = 0$ so the only candidate LDPE is $x = 0$, as in RS.

However, the FOCs above do not guarantee the existence of equilibrium, again as in RS. If $\gamma = 0$, then $x = 1 + \frac{1}{\beta}$ is the LDPE when $\beta < -1$, and $x = 0$ is the LDPE if $\beta > -1$. In this case, an equilibrium always exists although there is no demand for insurance since $\mathbb{E}[v] = 0 \Rightarrow v \equiv 0$. The intuition again mirrors that of RS, extended to a multidimensional context. Here, γ is a force pushing for second-order deviations away from a low-insurance equilibrium, arising from the gains from trade from insurance, $\mathbb{E}[v]$. The smaller is γ, the smaller is the range over which no equilibrium exists. If $\beta > \gamma - 1$, a no-insurance LDPE exists because sorting is so adverse that it overcomes the pressure of γ to deviate from zero insurance. If β is negative enough, sorting is sufficiently advantageous that there is a first-order gain from raising insurance above $x = 0$. In the intermediate region, adverse sorting local to zero-insurance eliminates the possibility of a positive insurance LDPE, but insurance demand makes a zero-insurance equilibrium vulnerable to second-order deviations, so no LDPE exists. The results is expressed in Figure 2 above.

2.5 Welfare

A natural question following our analysis above is the socially optimal pooling contract. That is, constrained to choosing $x_0 = x_1 = x$ and $p_0 = p_1 = p$, what is the welfare maximizing value of (x, p)? To proceed with this analysis, we equate WTP with utility as in the standard Kaldor (1939)-Hicks (1939) analysis. If the market is covered, there is no optimal value of p because prices are purely distributive and do not affect behavior. Thus, we focus on the optimal level of x and on the effect of market power on welfare through the equilibrium level of x. As in the previous subsection, competition increases the importance of sorting (attracting the right kind of consumers) relative to the importance of gains from trade. However this can increase or decrease welfare. To show this we first discuss the competitive model presented above, where we continue to assume away the Spence distortion, and find that market power is beneficial.[11] Then, we present a model where a Spence distortion is present and where sorting mitigates this distortion, so competition is beneficial.

We maintain the model presented above, and define welfare as $W = \mathbb{E}[u - c] = \mathbb{E}[\phi] = \mathbb{E}\left[\frac{1-(1-x)^2}{2} v\right]$. The following result characterizes the inefficiency in the equilibrium level of x this market.

Proposition 5. *Full insurance is socially optimal, but it is never an LDPE. However, as t rises, x does as well and* $\lim_{t\to\infty} x = 1$.

Proof. We have $\frac{\partial W}{\partial x} = \mathbb{E}[\phi'] = (1-x)\mathbb{E}[v] \geq 0$, so x always increases welfare, which is maximized at full insurance ($x = 1$). However, at full insurance we have $\frac{\partial \Pi}{\partial x} = (1-x)\mathbb{E}[v] - x\mathbb{V}[\mu](1 + (1-x)\beta) = -\mathbb{V}[\mu] < 0$, so it cannot be an LDPE since profit always increases by lowering coverage away from $x = 1$. From Proposition 4, an LDPE must satisfy the FOC $(1-x)\mathbb{E}[v]t = x\mathbb{V}[\mu] + x(1-x)\mathrm{Cov}[\mu, v]$. Applying the IFT to this expression yields

[11] This builds on the general insight of Yin (2004) that the Spence distortion of non-linear pricing vanishes when average anad marginal consumers have the same preferences.

$\frac{\partial x}{\partial t} = \frac{(1-x)\mathbb{E}[v]}{\mathbb{E}[v]t+\mathbb{V}[\mu]+(1-2x)\text{Cov}[\mu,v]}$. The numerator is positive. Using the FOC to substitute for $\mathbb{E}[v]t$ in the denominator yields

$$\frac{\mathbb{V}[\mu]}{1-x} + (1-x)\text{Cov}[\mu, v] > x\mathbb{V}[\mu] + x(1-x)\text{Cov}(\mu, v) = (1-x)\mathbb{E}[v]t > 0$$

The inequality follows from $x \in (0, 1)$. The equality follows from the FOC. □

An increase in x increases the expected cost (c) to the risk-neutral insurer, but reduces both expected cost (c) and risk (ϕ) to the risk-averse consumers, so it is intuitive that full insurance is the social optimum. Notice that we are assuming the absence of moral hazard (the distribution of shocks does not depend on x), which would further lower the socially optimal level of x.

However, equilibrium insurance coverage is always insufficient because under any degree of competition as it is always optimal to lower x below 1. Insurance zero value vanishes as perfect insurance is reached, (viz. $\phi(1, v) = 0$), which has two implications. First, there are no gains from trade at full insurance, so even a small incentive to sort is enough to reduce coverage away from $x = 1$. Second, at full insurance, sorting must be adverse ($\text{Cov}[u'(1, \mu, v), c(x, \mu)] = x\mathbb{V}[\mu] \geq 0$). When coverage is low, additional coverage may lead to advantageous sorting by attracting individuals with high insurance value and low risk types, as found by Finkelstein and McGarry (2006). However, at full insurance, the marginal utility for x of individuals with large v is low, whereas marginal utility driven by μ is undiminished. Thus, sorting incentives may incentivize higher generosities of coverage at low insurance levels (if β is negative enough), but sorting always incentivizes lower insurance generosity near full insurance.

Finally, sorting must always be adverse in equilibrium because, in regions where sorting is advantageous, insurers who raise x both increase the number of buyers and improve their composition. Therefore reducing the weight placed on sorting (by increasing market power t) always raises x, which increases welfare. To see this differently, notice that there is no Spence distortion, so all distortions in x are due to sorting. Competing firms do not internalize the cost they impose on rivals when they cream-skim, whereas a monopolist would do so and thus would have the same incentive to sort as a social planner (there are no externalities on rivals in that case).[12] Therefore, approaching monopoly reduces the sorting distortion. The result establishes formally for the first time the intuition of RS that market power may improve the quality of products in selection markets.

[12]This is trivially true here as a monopolist would have no margin of consumers buying products that she does not sell, but in Subsection 5.2 we show this is true more broadly, even when the monopolist does engage in some sorting.

This result contrasts sharply with the findings of Rochet and Stole (2002) who show that similar cream-skimming drives out price discrimination and restores optimal pricing-at-cost in a related model. In the environment we have considered thus far, only insurance matters while ex-post incentives are unimportant due to the absence of moral hazard. We have therefore taken an *ex-ante* perspective, in the sense that we evaluated welfare before consumers incurred their shocks. Conversely, Rochet and Stole adopt a purely *ex-post* perspective, since there is no uncertainty and the only efficiency concern is the individual's choice of consumption (moral hazard).[13] We now present an alternative model, related to the one above, but where welfare is maximized under perfect competition.

We consider ex-post provision of healthcare by two firms that compete in two-part tariffs: p is the entry fee and x is the cost per unit of medical services used. Consumers realize some health shock $L \in \mathbb{R}_+$, which has some distribution in the population of $f(L) > 0$, then purchase healthcare. We assume demand $q^\star = q^\star(x, L) = (A - x) L > 0$ is linear and scales upwards with L, for $A > 0$ homogeneous and sufficiently large. The link to the previous model is that a lower x implies more generous insurance, since it is easier for consumers to satisfy their healthcare needs ex-post. Consumer utilization changes with per-unit price x, so there is now a moral hazard concern that is absent from the model of Proposition 5. Surplus is the area under the demand curve, $CS(x, L) = \frac{1}{2}(A - x) q^\star = \frac{1}{2}(A - x)^2 L$. We normalize the wealth value of the shock to H is such that, when $x = 0$, all consumers end up with the same surplus, so $CS(0, L) = H = \frac{1}{2}A^2 L$. Final wealth with provider i is $w_i(x_i, L, p_i) = CS(x_i, L) - H - p_i = \frac{x_i}{2}(x_i - 2A) L - p_i$. Individuals have utility $U(w)$, where w is final wealth, $U' \equiv \frac{dU}{dw} > 0$ and $U'' \equiv \frac{\partial^2 U}{\partial w^2} < 0$.

The market is covered. Consumers differ in a Hotelling location parameter b, which is distributed $b \sim \mathcal{U}[0, 1]$. The distribution of L and b are independent. A consumer purchasing from insurer 0 incurs a utility loss of tb, while purchasing from provider 1 implies a utility loss of $t(1 - b)$. A consumer of type (L, b) purchases from provider 0 if $b \leq b^\star = \frac{1}{2t}(U(w_0) - U(w_1)) + \frac{1}{2}$. Profit of provider 0 is $\pi_0 = \int_0^\infty \int_0^{b^\star} [p_0 + (x_0 - k) q^\star(L, x_0)] f(L)\, db dL$. Let Q_i be the quantity of individuals purchasing from provider i. In this setting, $M = \int_0^\infty \frac{1}{2t} U' f(L)\, dL$ and $\mathbb{E}_u[\zeta(L, b) \mid b = b^\star] = \mathbb{E}_u[\zeta(L, b)] = \frac{1}{M}\int_0^\infty U' \zeta(L, b) f(L)\, dL$. For details, see Appendix B.

Traveling costs are now fungible with utility, not price. Individuals with larger medical

[13] Rochet and Stole (2002) assume marginal consumers purchase less, not more, than infra-marginal consumers, as in Mussa and Rosen (1978). In many contexts this is an appropriate assumption, but quite broadly, as discussed by Fabinger and Weyl (2014), price discrimination is in favor of those with elastic demand that results from a high marginal utility of wealth. Thus it is common more broadly that the price discrimination driven out by cream-skimming offers ex-post insurance.Rochet and Stole (2002) also consider fully non-linear tariffs but restrict heterogeneity to two dimensions with independent marginal distributions or focus on a case with two types.

expenditures are, by those expenditures, made poorer and thus more price-sensitive and willing to switch between products for a small change in price. This provides an incentive for firms to reduce per-unit prices below marginal cost ($x < k$) in order to discriminate in favor of these marginal consumers. However, welfare is maximized by pricing at marginal cost ($x = k$). This Spence distortion is inefficient because it promotes excessive utilization by infra-marginal consumers. However, it is mitigated by the distortion associated with the sorting incentive and, since competition emphasizes sorting, it also improves welfare. This is stated formally as follows.

Proposition 6. *In the two-part tariff model described above, any LDPE requires*

$$Q\left(\mathbb{E}\left[q^{\star}\right] - \mathbb{E}_u\left[q^{\star}\right]\right) = (x - k)\left(Q\mathbb{E}\left[L\right] + M\mathbb{V}_u\left[q^{\star}\right]\right).$$

For any $t > 0$, an LDPE implies $x < k$. We have $\lim_{t\to 0} x = k$, which is socially optimal.

Proof. Welfare is clearly maximized at $x = k$ since consumers make optimal consumption decisions in this case. The condition above is obtained from Proposition 1, where $u' = -q^{\star}$ and $c = p + (x - k)\, q^{\star}$. At the LDPE, we have $b^{\star} = Q = \frac{1}{2}$. Moreover, $\mathbb{E}_u\left[\cdot \mid b = b^{\star}\right] = \mathbb{E}_u\left[\cdot\right]$ in this setting, and $\mathrm{Cov}_u\left[-q^{\star}, p + (x - k)\, q^{\star}\right] = \mathbb{V}_u\left[q^{\star}\right]$. We can then show that $\mathbb{E}\left[q^{\star}\right] - \mathbb{E}_u\left[q^{\star}\right] < 0$, since $\mathbb{E}_u\left[q^{\star}\right]$ is weighted by $\frac{\partial U}{\partial w}$ which places larger weights on users with large shocks (L) and thus higher $q^{\star}$. Since $Q\mathbb{E}\left[L\right] + M\mathbb{V}_u\left[q^{\star}\right] > 0$, we must have $x < k$ at the LDPE. By Proposition 2, as $t \to 0$, we must have the sorting term vanish, which implies $x \to k$ (since $\mathbb{V}_u\left[q^{\star}\right] > 0$). For details, see Appendix B. $\square$

While the LDPE features distortionary cream skimming ($x < k$) when firms have market power, as competition increases, sorting incentives drive per-unit price to cost ($x = k$), thus approaching the social optimum. This is in clear contrast to the result of Proposition 5, where market power ($t \to \infty$) brought x close to the social optimum. The common theme is that competition approximates $\mathrm{Cov}_u\left[u', c\right]$ to zero, thus causing prices (which are proportional to marginal utility) to approximate costs in the population overall. Proposition 5 takes an ex-ante perspective of welfare: insurance (and therefore selection) is crucial whereas moral hazard is absent (ex post consumption is always optimal). Welfare is maximized when consumers do not bear their ex-post costs, so forcing consumers to bear these costs eliminates insurance and reduces welfare. Proposition 6 takes a purely ex-post perspective: making consumers bear their ex-post cost leads them to consume the ex-post social optimal amount of healthcare. The heterogeneity of wealth ex-post, driven by health shocks, makes it attractive to provide (ex-post inefficient) insurance to cater to the more price sensitive customers, but competition eliminates this distortion through its magnification of the sorting incentive.

In Subsection 5.1 below we consider a calibrated model including both moral hazard and an insurance motive. This calibration suggests that the insurance motive is far stronger than the moral hazard concern, which is why we chose the focus we did for the bulk of this paper. But before doing so, we turn to a calibration of our basic model.

3 Empirical Calibration

We now use summary statistics from Handel, Hendel and Whinston (2013) to perform a simple calibration of the competitive model described in Section 2.4.[14] These authors use proprietary claims data from a large employer. The data does not arise from a competitive market, instead stemming from a firm which uses cross-subsidies to achieve a variety of objectives other than single-plan profit maximization, such as employee retention and productivity. Assuming that a widely-used proprietary risk estimation package represents the information set of individuals, the authors are able to recover the joint distribution of μ and v for the entire population from the joint distribution of claims, risk estimates based on the package and plan choice by individuals.[15] We use moments of this distribution to determine the properties of the unique candidate competitive LDPE in a CARA-normal framework, computed in Section 2.4. For the purposes of the calibration, we follow Handel, Hendel and Whinston in allowing three dimensions of heterogeneity: a, μ and σ^2. In determining WTP and marginal WTP with respect to co-insurance, a and σ^2 enter only through v. Thus all of our formulae from the previous section remain valid here.

To the first two significant digits, Handel, Hendel and Whinston find (in dollars and dollars squared) $\mathbb{E}[\mu] = 6.6 \times 10^3$, $\mathbb{E}[v] = 68 \times 10^3$, $\mathbb{V}(\mu) = 50 \times 10^6$ and $\text{Cov}\,[\mu, v] = 630 \times 10^6$ within their sample. The last of these two of these numbers implies that $\beta \approx 13 > 0$ and thus the unique candidate for an LDPE is $x = 0$. However we can also compute that $\gamma \approx 23 > 13 + 1$. Thus this no-insurance candidate is not actually an equilibrium because the value of insurance is too great. While these statistics are taken for the full population, the same qualitative features apply to the market in every five-year age bucket. Thus, in their data, no equilibrium exist. Moreover the multidimensionality of private information exacerbates rather than mitigates adverse selection because insurance value is positively correlated with mean risk but (as in Cohen and Einav (2007) but unlike in Finkelstein and McGarry (2006)). Interestingly, Handel, Hendel and Whinston find that a is negatively

[14]We thanks these authors for their generosity.

[15]They follow an approach analogous to that of Cohen and Einav (2007), using individuals' choices among available plans coupled with an assumption that a is distributed normally conditional on $\mu, \sigma^2$and other covariates to estimate the joint distribution of μ, σ^2 and a.

correlated with μ, as in Finkelstein and McGarry.[16] However, there is an even stronger positive correlation between σ^2 and μ, presumably driven by the fact that sick individuals have both high and highly variable expenditures, which makes $\text{Cov}[\mu, v] > 0$. This illustrates how seemingly innocuous simplifications of the nature of consumer heterogeneity can have important effects on equilibrium predictions.

Our results are highly consistent in spirit with those reported by Handel, Hendel and Whinston. They analyze two-plan equilibria between plans with exogenously specified (and approximately linear) actuarial rates. They find severe adverse selection leading almost always to complete collapse of the plan with the higher actuarial rate. While our two equilibrium concepts are clearly different, in a related paper (Veiga and Weyl, 2013) we establish a formal link between our equilibrium level of insurance and selection in a two-plan (linear actuarial rate) equilibrium. In particular, under somewhat stronger statistical assumptions, the direction of selection (adverse or advantageous) into the high actuarial rate plan is determined by whether the average of their actuarial rates is above or below our unique candidate LDPE actuarial rate. Thus the fact that this rate is 0 in this data strongly suggests any two plans, at least one of which has a positive actuarial rate, will have strong adverse selection into the more generous plan. Thus the value of β is a reduced-form statistic that helps provide intuition for their structural results.

We now use the summary statistics to calibrate the tendency of market power to increase insurance provision (Proposition 5). To do so we use the the expression from the proof of Proposition 5 for the unique candidate interior LDPE. Since insurer cost is per consumer is $x\mathbb{E}[\mu]$, we measure market power using the proportional-to-cost mark-up $\frac{t}{x\mathbb{E}[\mu]}$, which monotonically increases with t. We can therefore solve for $\frac{t}{x\mathbb{E}[\mu]} = \frac{1}{1-x}\frac{\mathbb{V}[\mu]+\text{Cov}[\mu,v]}{\mathbb{E}[v]\mathbb{E}[\mu]}$. Figure 3 shows the unique candidate LDPE co-insurance x as a function of this relative-to-cost mark-up. The most striking feature is that a large amount of market power is necessary to sustain any significant insurance. In fact, calibration of the second-order condition using Proposition 3 shows that a mark-up of 90% ($\frac{t}{x\mathbb{E}[\mu]} = 0.9$) is necessary for an LDPE to exist at all. At a mark-up of 100%, the equilibrium actuarial rate is only 40%, and to attain rates similar to the typical rate in Handel, Hendel and Whinston's data (86% actuarial rates) mark-ups of around 50 times cost are necessary. It thus seems implausible, in this calibration, that market power could offer a solution to the distortions from competition.[17] Beyond its distributive consequences, such extreme market power is likely to significantly reduce the fraction of the population that is covered, as at such high prices a mandate for individual coverage is likely

[16]However, even if σ^2 is treated as homogeneous, β, while negative, is always above -1 in their data so that no positive insurance candidate LDPE exists.

[17]This result is also consistent with the very low actuarial rates for non-group markets prior to the Affordable Care Act in the US.

to collapse. Mahoney and Weyl (2013) provide a detailed discussion of the effect of market power on population coverage.

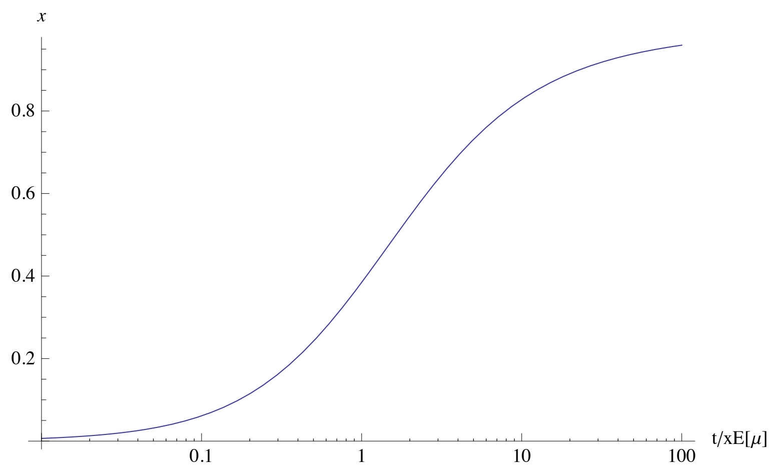

Figure 3: Unique candidate LDPE actuarial rate x as a function of $t/x\mathbb{E}[\mu]$, shown on a logarithmic scale.

4 A General Model of Sorting

In this section we show how the techniques and results of Section 2 generalize to richer settings and provides rigorous proofs of the results above. We will focus on relating these more general concepts to those of the previous sections, rather than discussion of the results or assumptions.

4.1 Setup

A monopoly offers an insurance contract characterized by a price $p \in \mathbb{R}$ and a non-price characteristic $x \in \mathbb{R}$. The scalar x can trivially be extended to be a vector (see Veiga and Weyl (2013)). There is a unit mass of consumers characterized by their type, a T-dimensional vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_T) \in (\underline{\theta_1}, \overline{\theta_1}) \times \cdots \times (\underline{\theta_T}, \overline{\theta_T}) \subseteq \mathbb{R}^T$, which is not contractible. It is common knowledge that $\boldsymbol{\theta}$ is distributed according to the atomless and full support probability density function $f(\boldsymbol{\theta})$.

Consumers face a wealth shock $l \in \mathbb{R}$, distributed according to the density $g(l, \boldsymbol{\theta})$. Consumers have a utility function $U = U(w, \boldsymbol{\theta})$, where w is final wealth and $\frac{\partial U}{\partial w} \equiv U' > 0$ and $\frac{\partial^2 U}{\partial w^2} \equiv U'' < 0$. An insurer reimburses $G = G(l, x)$ if the consumers incurs a loss l. For instance, $G = l$ is full insurance and $G = 0$ is no insurance. Initial wealth is w_0 (potentially heterogeneous). Let $\mathbb{E}_l[\cdot]$ denote expectation over l. The WTP of a consumer of type $\boldsymbol{\theta}$ is $u(x, \boldsymbol{\theta})$, the level of p that equates expected utility with insurance to that without insurance, or $\mathbb{E}_l\left[U\left(w_0 - l + G\left(l,x\right) - u\left(x,\boldsymbol{\theta}\right),\boldsymbol{\theta}\right) \mid \boldsymbol{\theta}\right] = \mathbb{E}_l\left[U\left(w_0 - l,\boldsymbol{\theta}\right) \mid \boldsymbol{\theta}\right]$. We assume that $u\left(x,\boldsymbol{\theta}\right)$ is continuously differentiable in all arguments and strictly increasing in all arguments. We define $u'\left(x,\boldsymbol{\theta}\right) \equiv \frac{\partial u(x,\boldsymbol{\theta})}{\partial x} > 0$.

Expected cost of type $\boldsymbol{\theta}$ to the risk-neutral insurer is $c\left(x,\boldsymbol{\theta}\right) = \mathbb{E}_l\left[G \mid \boldsymbol{\theta}\right]$, so $c'\left(x,\boldsymbol{\theta}\right) \equiv \frac{\partial c(x,\boldsymbol{\theta})}{\partial x} = \mathbb{E}_l\left[\frac{\partial G}{\partial x} \mid \boldsymbol{\theta}\right]$. We assume $c\left(x,\boldsymbol{\theta}\right)$ is twice continuously differentiable in all arguments.

Implicitly differentiating the equality that defines $u\left(x,\boldsymbol{\theta}\right)$ (and omitting functional arguments) yields $u'\left(x,\boldsymbol{\theta}\right) - c'\left(x,\boldsymbol{\theta}\right) \equiv \phi'\left(x,\boldsymbol{\theta}\right) = \frac{\mathrm{Cov}_l\left[U',\frac{\partial G}{\partial x}|\boldsymbol{\theta}\right]}{\mathbb{E}_l\left[U'|\boldsymbol{\theta}\right]}$. For details, see Appendix **??**. The amount by which type $\boldsymbol{\theta}$ values additional insurance ($u'\left(x,\boldsymbol{\theta}\right)$), is the transfer of mean risk to the insurer $c'\left(x,\boldsymbol{\theta}\right) = \mathbb{E}_l\left[\frac{\partial G}{\partial x} \mid \boldsymbol{\theta}\right]$ in addition to the marginal social surplus insurance $\phi'\left(x,\boldsymbol{\theta}\right)$. In turn, surplus from insurance captures the extent to which coverage increases for those states with highest marginal utility (i.e., high realizations of l and high U'). Under risk neutrality, U' is constant so $\phi'\left(x,\boldsymbol{\theta}\right) = 0$. In Section 2, we had $\phi'\left(x,\boldsymbol{\theta}\right) = \left(1-x\right)v$.

The set of consumers purchasing the product is $\Theta \equiv \{\boldsymbol{\theta} : u\left(x,\boldsymbol{\theta}\right) \geq p\}$, while the set of marginal consumers indifferent to purchasing is $\partial\Theta \equiv \{\boldsymbol{\theta} : u\left(x,\boldsymbol{\theta}\right) = p\}$. We define $\boldsymbol{\theta}_{-T} = \left(\theta_1, ..., \theta_{T-1}\right)$ and follow Veiga and Weyl (2013) in making the following assumption

Assumption 1. *There exists a function* $\theta_T^\star\left(p,x,\boldsymbol{\theta}_{-T}\right)$ *such that* $u\left(x,\boldsymbol{\theta}\right) \geq p \Leftrightarrow \theta_T \geq \theta_T^\star\left(p,x,\boldsymbol{\theta}_{-T}\right)$.

In Section 2, this was the function $\mu^\star\left(p,x,v\right)$. Defining $\mathbf{t} \equiv \left(\boldsymbol{\theta}_{-T}, \theta_T^\star\left(p,x,\boldsymbol{\theta}_{-T}\right)\right)$, an integral over the set of marginal buyers $\partial\Theta$ can be expressed as an iterated integral over the entire range of $\boldsymbol{\theta}_{-T}$. For an integrable function $\zeta\left(x,\boldsymbol{\theta}\right)$, Assumption 1 allows us define

$$\int_{\Theta} \zeta\left(x,\boldsymbol{\theta}\right) d\boldsymbol{\theta} \equiv \int_{\underline{\theta_1}}^{\overline{\theta_1}} ... \int_{\theta_T^\star(p,x,\boldsymbol{\theta}_{-T})}^{\overline{\theta_T}} \zeta\left(x,\boldsymbol{\theta}\right) d\boldsymbol{\theta},$$

$$\int_{\partial\Theta} \zeta\left(x,\mathbf{t}\right) d\boldsymbol{\theta}_{-T} \equiv \int_{\underline{\theta_1}}^{\overline{\theta_1}} ... \int_{\underline{\theta_{T-1}}}^{\overline{\theta_{T-1}}} \zeta\left(x,\boldsymbol{\theta}_{-T},\theta_T^\star\left(p,x,\boldsymbol{\theta}_{-T}\right)\right) d\boldsymbol{\theta}_{-T}$$

The quantity of buyers is $Q = Q\left(x,p\right) \equiv \int_{\Theta(x,p)} f(\boldsymbol{\theta})d\boldsymbol{\theta}$. It is shown in the proof of Proposition 1 below that $\frac{\partial Q}{\partial p} < 0$ for all $Q > 0$. We can therefore invert $Q\left(x,p\right)$ with respect to p to recover the function $P\left(x,q\right)$, which solves $Q\left(P\left(x,q\right),x\right) = q$. We can then define profit as $\Pi \equiv qP\left(x,q\right) - C$, where $C \equiv \int_{\Theta(P(x,q),x)} c\left(x,\boldsymbol{\theta}\right) d\boldsymbol{\theta}$ is total cost.

We define $M \equiv -\frac{\partial Q}{\partial p} = \int_{\partial\Theta} \frac{f(\mathbf{t})}{\partial u(x,\mathbf{t})/\partial\theta_T} d\boldsymbol{\theta}_{-T}$ (see proof of Proposition 1 for details on $\frac{\partial Q}{\partial p}$). For a function $\zeta\left(x,\boldsymbol{\theta}\right)$, we define expectation conditional on the set of marginal consumers as

$$\mathbb{E}_u\left[\zeta\left(x,\boldsymbol{\theta}\right) \mid \partial\Theta\right] \equiv \frac{1}{M}\int_{\partial\Theta} \frac{\zeta\left(x,\mathbf{t}\right) f\left(\mathbf{t}\right)}{\partial u\left(x,\mathbf{t}\right)/\partial\theta_T} d\boldsymbol{\theta}_{-T}.$$

For functions $\zeta_1(x, \boldsymbol{\theta})$ and $\zeta_2(x, \boldsymbol{\theta})$, we similarly define $\mathrm{Cov}_u[\zeta_1, \zeta_2 \mid \partial\Theta] \equiv \mathbb{E}_u[\zeta_1\zeta_2 \mid \partial\Theta] - \mathbb{E}_u[\zeta_1 \mid \partial\Theta]\,\mathbb{E}_u[\zeta_2 \mid \partial\Theta]$. Notice that, in Section 2, we had $\frac{\partial u}{\partial \theta_T} = \frac{\partial u}{\partial \mu} = \frac{1}{x}$, so this term vanishes in $\mathbb{E}_u[\zeta(x, \boldsymbol{\theta}) \mid \partial\Theta]$ since it is present in the numerator and denominator. We will omit most functional arguments for simplicity henceforth.

4.2 Monopoly

The firm chooses q and x to maximize profit. The FOC with respect to q is familiar ($P - \frac{q}{M} = \mathbb{E}_u[c \mid \partial\Theta]$, or marginal revenue equals marginal cost) so we do not discuss it more extensively.[18] Instead, our analysis will focus on the the non-price product dimension x.

Theorem 1. *A necessary FOC for the profit maximizing choice of x is*

$$-\underbrace{q\mathbb{E}[c' \mid \Theta]}_{cost} + \underbrace{q\mathbb{E}_u[u' \mid \partial\Theta]}_{Spence\ term} = \underbrace{M\mathrm{Cov}_u[u', c \mid \partial\Theta]}_{sorting\ effect}.$$

Proof. Differentiating $\Pi(x, q)$ with respect to x yields $\frac{\partial \Pi}{\partial x} = q\frac{\partial P}{\partial x} - \frac{\partial C}{\partial x} - \frac{\partial C}{\partial p}\frac{\partial P}{\partial x}$. We compute each of these terms below, following Veiga and Weyl (2013).

We apply the IFT to $u(x, \mathbf{t}) - p = 0$ to obtain $\frac{\partial \theta_T^\star}{\partial p} = -\frac{1}{\partial u(x,\mathbf{t})/\partial \theta_T}$ and $\frac{\partial \theta_T^\star}{\partial x} = -\frac{\partial u(x,\mathbf{t})/\partial x}{\partial u(x,\mathbf{t})/\partial \theta_T}$. Using the Leibniz Rule to differentiate $Q(p, x)$ and $C \equiv \int_\Theta cf(\boldsymbol{\theta})d\boldsymbol{\theta}$ yields

$$\frac{\partial Q}{\partial p} = \int_{\partial\Theta} \frac{\partial \theta_T^\star}{\partial p} f(\mathbf{t})d\boldsymbol{\theta}_{-T} = \int_{\partial\Theta} -\frac{f(\mathbf{t})}{\partial u(x,\mathbf{t})/\partial \theta_T} d\boldsymbol{\theta}_{-T} = -M$$

$$\frac{\partial Q}{\partial x} = \int_{\partial\Theta} \frac{\partial \theta_T^\star}{\partial x} f(\mathbf{t})d\boldsymbol{\theta}_{-T} = \frac{M}{M}\int_{\partial\Theta} \frac{u'(x,\mathbf{t})f(\mathbf{t})}{\partial u(x,\mathbf{t})/\partial \theta_T} d\boldsymbol{\theta}_{-T} = M\mathbb{E}_u[u' \mid \partial\Theta]$$

$$\frac{\partial C}{\partial p} = \int_{\partial\Theta} -\frac{c(x,\mathbf{t})f(\mathbf{t})}{\partial u(x,\mathbf{t})/\partial \theta_T} d\boldsymbol{\theta}_{-T} = -M\mathbb{E}_u[c \mid \partial\Theta]$$

$$\frac{\partial C}{\partial x} = \int_{\partial\Theta} \frac{\partial \theta_T^\star}{\partial x} c(x,\mathbf{t})f(\mathbf{t})d\boldsymbol{\theta} + \int_\Theta c' f(\boldsymbol{\theta})d\boldsymbol{\theta} = M\mathbb{E}_u[u'c \mid \partial\Theta] + q\mathbb{E}[c' \mid \Theta]$$

Applying the IFT to $Q(P, x) = q$, and using the results above for $\frac{\partial Q}{\partial x}$ and $\frac{\partial Q}{\partial p}$ we obtain

$$\frac{\partial P(x, q)}{\partial x} = -\frac{\partial Q(P, x)/\partial x}{\partial Q(P, x)/\partial p} = \mathbb{E}_u[u' \mid \partial\Theta].$$

Plugging these results into $\frac{\partial \Pi}{\partial x}$ and using the definition of $\mathrm{Cov}_u[u', c \mid \partial\Theta]$ yields the result. □

[18] Its more sophisticated component, marginal cost $\mathbb{E}_u[c \mid \partial\Theta]$, is discussed in Veiga and Weyl (2013), Einav, Finkelstein and Cullen (2010) and Einav and Finkelstein (2011).

4.3 Signing the sorting incentive

The following two propositions present commonly-satisfied conditions under which the sorting incentive can be signed. Essentially, this is the case when u' and c, conditional on $\boldsymbol{\theta} \in \partial\Theta$, can be expressed as monotone univariate functions.

Theorem 2. *Suppose there exists a function $g(x, \boldsymbol{\theta}) : \mathbb{R} \times (\underline{\theta_1}, \overline{\theta_1}) \times \cdots \times (\underline{\theta_T}, \overline{\theta_T}) \mapsto \mathbb{R}$ such that $c \equiv \hat{c}(x, g(x, \boldsymbol{\theta}))$ and $u' \equiv \hat{u}(x, g(x, \boldsymbol{\theta}))$. If $\hat{c}(x, g(x, \boldsymbol{\theta}))$ and $\hat{u}(x, g(x, \boldsymbol{\theta}))$ are monotone in $g(x, \boldsymbol{\theta})$, $\frac{\partial \hat{c}(x, g(x, \boldsymbol{\theta}))}{\partial g} \frac{\partial \hat{u}(x, g(x, \boldsymbol{\theta}))}{\partial g}$ has the same sign as $\mathrm{Cov}_u[u', c \mid \partial\Theta]$.*

Proof. $\mathrm{Cov}_u(u', c \mid \boldsymbol{\theta} \in X) = \mathrm{Cov}_u(\hat{c}(x; g(x, \boldsymbol{\theta})), \hat{u}(x; g(x, \boldsymbol{\theta})) \mid \boldsymbol{\theta} \in X)$ for any sub-set of consumers X. If $\hat{c}(x; g(x, \boldsymbol{\theta}))$ and $\hat{u}(x; g(x, \boldsymbol{\theta}))$ are monotone increasing in their second argument, from Schmidt (2003), the covariance of two monotone increasing functions of a single variable is positive. The other cases are obtained by changing the direction of monotonicity of $\hat{c}(x, g(x, \boldsymbol{\theta}))$ and $\hat{u}(x, g(x, \boldsymbol{\theta}))$. □

The assumptions of Theorem 2, while seemingly special, apply naturally in contexts such as non-linear pricing, where the scalar amount purchased by each individual determines both her cost to the firm and her marginal utility from a change in the per-unit price, as in Proposition 6.

The covariance term can be signed more generally when types are bi-dimensional, say (θ_1, θ_2), so that $\partial\Theta$ is a one-dimensional curve. Since WTP is is increasing in all arguments, if (θ_1, θ_2) and $(\hat{\theta_1}, \hat{\theta_2})$ are both indifferent, $\hat{\theta_1} > \theta_1$ must imply $\hat{\theta_2} < \theta_2$ so, within $\partial\Theta$, θ_2 is a decreasing function of θ_1. Thus an index can be constructed and the logic of Theorem 2 can be applied. The result is formalized below.

Theorem 3. *Suppose that $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$. If $\forall \boldsymbol{\theta} \in \partial\Theta$, we have*

$$\left(\frac{\partial u'/\partial\theta_1}{\partial u/\partial\theta_1} - \frac{\partial u'/\partial\theta_2}{\partial u/\partial\theta_2}\right)\left(\frac{\partial c/\partial\theta_1}{\partial u/\partial\theta_1} - \frac{\partial c/\partial\theta_2}{\partial u/\partial\theta_2}\right) > 0,$$

then $\mathrm{Cov}_u[u', c \mid \partial\Theta] > 0$. The statement remains true if the two inequalities are reversed.

Proof. Given Assumption 1, $\partial\Theta$ is defined by a function $\theta_2^\star(p, x, \theta_1)$ such that $u(x, \theta_1, \theta_2^\star(p, x, \theta_1)) = p$. Thus, conditional on $\partial\Theta$, $u' = u'(x, \theta_1, \theta_2^\star(p, x, \theta_1)) = \hat{u}(x, \theta_1)$ and $c = c(x, \theta_1, \theta_2^\star(p, x, \theta_1)) = \hat{c}(x, \theta_1)$. We may thus apply the logic of the proof of Theorem 2 as θ_1 is a unidimensional index.

It thus remains only to determine conditions for the monotonicity of $\hat{u}$ and $\hat{c}$ in θ_1. First, we apply the IFT to $u(x, \theta_1, \theta_2^\star(p, x, \theta_1)) = p$, yielding $\frac{\partial \theta_2^\star(x, p, \theta_1)}{\partial \theta_1} = -\frac{\partial u\left(x, \theta_1, \theta_2^\star(p, x, \theta_1)\right)/\partial\theta_1}{\partial u\left(x, \theta_1, \theta_2^\star(p, x, \theta_1)\right)/\partial\theta_2}$. Then, we compute

$$\frac{\partial \hat{u}(x,\theta_1)}{\partial \theta_1} = \frac{\partial u'\left(x,\theta_1,\theta_2^{\star}\left(p,x,\theta_1\right)\right)}{\partial \theta_1} - \frac{\partial u'\left(x,\theta_1,\theta_2^{\star}\left(p,x,\theta_1\right)\right)}{\partial \theta_2}\frac{\partial u\left(x,\theta_1,\theta_2^{\star}\left(p,x,\theta_1\right)\right)/\partial \theta_1}{\partial u\left(x,\theta_1,\theta_2^{\star}\left(p,x,\theta_1\right)\right)/\partial \theta_2}$$

Since u increasing in all arguments, $\frac{\partial \hat{u}(x;\theta_1)}{\partial \theta_1}$ has the sign of $\frac{\partial^2 u/\partial x \partial \theta_1}{\partial u/\partial \theta_1} - \frac{\partial^2 u/\partial x \partial \theta_2}{\partial u/\partial \theta_2}$. The analogous formula applies for $\frac{\partial \hat{c}}{\partial \theta_1}$ by the same logic. □

Whether the relevant terms in the statement of Theorem 3 can be signed depends on the details of the specific model, although we have yet to find a case where this approach cannot be usefully applied, as in the computational exercise of Subsection 2.3. Unlike Theorem 2, Theorem 3 does not require the covariance within $\partial\Theta$ to have the same sign as the covariance within the set of all buyers. For example, when $\mathrm{Cov}[\mu, v] < 0$ it is still the case that raising the actuarial rate sorts adversely (Corollary 1).

4.4 Competition

We extend our analysis to a simple Hotelling (1929) competitive environment. We consider two insurers, indexed by $i \in \{0,1\}$, where i captures the insurer's location on a Hotelling line and $-i$ refers i's competitor. Each insurer i offers a single product characterized by a non-price characteristic x_i and a price p_i. The two insurers are identical apart from their Hotelling location. Consumers heterogeneity $\boldsymbol{\theta} \in \mathbb{R}^T$ is augmented by the scalar $b \in [0,1]$, which captures preferences between insurers. An individual with type b incurs a cost (fungible with price) of tb by purchasing from firm 0 and by $t(1-b)$ from purchasing from firm 1, where t captures market power. Individuals have a WTP u as above and choose the product that offers them highest utility net of price and transportation costs.

We make the two following simplifying assumptions.

Assumption 2. *The market is covered.*

Assumption 3. *The joint distribution of heterogeneity is $f(\boldsymbol{\theta})$.*

Given these assumptions, individuals purchasing product 0 are those for whom $b \leq b^{\star} \equiv \frac{1}{2t}\left(u(x_0,\boldsymbol{\theta}) - p_0 - u(x_1,\boldsymbol{\theta}) + p_1\right) + \frac{1}{2}$ and those purchasing from insurer 1 are those for whom $b > b^{\star}$. The set of consumer for whom $b = b^{\star}$ is marginal to both insurers.[19]

We will focus on *local deviation pooling equilibri*a (LDPE), which we define as follows.

Definition 1. A *local deviation pooling equilibrium* (LDPE) is a pair (x,p) such that, if firm i plays $(x_i,p_i) = (x,p)$ then there exists an $\epsilon > 0$ such that (x,p) maximizes the profit of

[19] Assumption 3 is relaxed in an earlier version of this paper, Veiga and Weyl (2013).

firm $-i$ in the set $(x-\epsilon, x+\epsilon) \times (p-\epsilon, p+\epsilon)$. If the FOC and SOC are satisfied at (x,p) then (x,p) is an LDPE.

If (x,p) is an LDPE, the market is split evenly, so $b^\star = \frac{1}{2}$. Defining $Q_0 \equiv \int_{\boldsymbol{\theta}} \int_0^{b^\star} f(\boldsymbol{\theta})\, db d\boldsymbol{\theta}$ as in Section 4, we obtain that, at an LDPE, $x_0 = x_1 = x$, $Q_0 = Q_1 = \frac{1}{2}$ and $\frac{\partial Q_0}{dp_0} = -\frac{1}{2t} \equiv -M$. Otherwise the model is as in Subsection 4.1, with subscripts denoting the relevant firm. Notice that, under these assumptions, the composition of $\partial\Theta$ is the same as the composition of buyers overall.

We begin by considering that firms have some degree of market power ($t > 0$). Pricing at an LDPE in this Hotelling context is well-known: $p + t = \mathbb{E}[c]$, where marginal cost $\mathbb{E}[c]$ is the average cost in the full population by Assumptions 3 and 2. As in Section 4, we focus on the firm's choice of x. Recall that $u' - c' \equiv \phi'$.

Theorem 4. *If $t > 0$, a necessary FOC for the optimization of x is*

$$\mathbb{E}[\phi'] = \frac{1}{t}\mathrm{Cov}[u', c]$$

and there always exists an interior point where this condition holds. The SOC is

$$t\mathbb{E}[\phi''] + \frac{1}{2}\mathbb{E}[\phi']^2 < \mathrm{Cov}[u'', c] + 2\mathrm{Cov}[u', c'].$$

Proof. The competitive FOC and SOC are derived in Appendix C. The derivative of profit evaluated at no insurance is $\frac{\partial \Pi}{\partial x} = \frac{1}{2}\mathbb{E}[\phi'] + 0 > 0$. The proof of Proposition 7 shows that the derivative of profit evaluated at full insurance is $\frac{\partial \Pi}{\partial x} < 0$. Therefore, there is some interior point at which $\frac{\partial \Pi}{\partial x} = 0$ by continuity. $\square$

We can now consider the competitive limit as $t \to 0$, which is described in the following result. Recall $G = l$ corresponds to full insurance while $G = 0$ corresponds to zero insurance. Recall $G = l$ corresponds to full insurance while $G = 0$ corresponds to zero insurance. In this limit, firms make zero profit as in RS, since we have $p - \mathbb{E}[c] = t \to 0$.

Corollary 2. *Asssume $\mathbb{E}[\phi]$ is bounded and consider the limit as $t \to 0$:*

An LDPE must satisfy $\mathrm{Cov}[u', c] = 0$.

This condition always holds at no insurance ($G = 0$).

No insurance is an LDPE if $\frac{1}{4}\frac{\mathbb{E}[\phi']^2}{\mathbb{V}[c']}\mid_{G=0} -1 < \frac{\mathrm{Cov}[\phi', c']}{\mathbb{V}[c']}\mid_{G=0}$.

The condition holds at some interior point if $\frac{Cov(\phi', c')}{\mathbb{V}[c']}\mid_{G=0} < -1$.

Proof. For an interior LDPE, if $\mathbb{E}[\phi]$ is bounded, as $t \to 0$, satisfying Equation 4 requires $\mathrm{Cov}[u', c] \to 0$.

At no insurnace, c is constant, so sorting vanishes. Evaluating the SOC at $t = G = 0$, and noticing $\text{Cov}[u'', c] = 0$ in this case, yields the inequality result.

Proposition 7 shows that sorting is adverse at full insurance. Then, by continuity, the sorting term vanishes at some interior point if $\frac{\partial \text{Cov}(u',c)}{\partial x}\mid_{G=0} < 0$. Then, $\frac{\partial Cov(u',c)}{\partial x} = Cov(u'', c) + Cov(u', c')$. However, $G = 0 \Rightarrow Cov(u'', c) = 0$ so, at no insurance, $\frac{\partial Cov(u',c)}{\partial x} < 0 \Leftrightarrow \mathbb{V}[c'] + Cov(\phi, c') < 0$. □

In Section 2, we had $\frac{Cov(\phi,c')}{\mathbb{V}[c']}\mid_{G=0} = \frac{(1-x)Cov(v,\mu)}{\mathbb{V}[\mu]}\mid_{x=0} = \beta$, so the condition for existence of an interior candidate LDPE is a generalization of Proposition 2. Similarly, the SOC at no insurance presented above implies that, in Section 2, the SOC holds for $x = 0$ if $\gamma - 1 < \beta$, as in Proposition 4.

4.5 Welfare

A consequence of Proposition 4 is that, under imperfect competition, complete insurance cannot be an equilibrium, even in the absence of moral hazard.[20] To define a welfare optimum, we equate WTP with utility and define welfare as $W = \mathbb{E}[u - c] = \mathbb{E}[\phi]$.

Proposition 7. *Full insurance is socially optimal.*

If cost is correlated with marginal cost ($\text{Cov}[c, c'] > 0$), full insurance is not an LDPE.

If $\text{Cov}[u', c]$ is bounded, then as $t \to \infty$, x approaches its welfare maximizing level.

Proof. We have $\phi = \frac{\text{Cov}_l\left[U', \frac{\partial G}{\partial x}|\boldsymbol{\theta}\right]}{\mathbb{E}_l[U'|\boldsymbol{\theta}]}$. At full insurance, U' does not vary, so this term is zero, so the FOC for the maximization of welfare is satisfied. The derivative of profit at full insurance is $\frac{\partial \Pi}{\partial x} = \mathbb{E}[\phi'] - \text{Cov}[c' + \phi', c]$. At full insurance, $\phi' = 0$, so this derivative is negative at full insurance under the assumptions above. From Proposition 4, an LDPE satisfies $\mathbb{E}[\phi]\, t = \text{Cov}[u', c]$. If the SOC holds, then $\frac{\partial \Pi}{\partial x^2} < 0$. Applying the IFT to the FOC yields $\frac{\partial x}{\partial t} = -\frac{\mathbb{E}[\phi']}{\partial \Pi / \partial x^2} > 0$. So market power increases x if and only if welfare ϕ' is increasing in x. □

The condition $\text{Cov}[c, c'] > 0$ is quite natural. When x captures a linear coinsurance rate, cost is $xc(\boldsymbol{\theta})$, so this condition amount to there existing some heterogeneity in expected losses at full insurance: $\mathbb{V}[c'] > 0$. In Section 2, this condition is $\mathbb{V}[\mu] > 0$.

5 Robustness Checks

In this section we discuss the robustness of our conclusions to two factors that could dampen the incentives for cream-skimming and the social benefit of high actuarial rates: moral hazard

[20] We discuss the robustness of this assumption in Section 5.

and market expansion.

5.1 Moral hazard

Our calibration in Section 3 relied on the assumption that mean risk μ was invariant to the generosity of insurance coverage x, which implies that full insurance was optimal. In reality, there is substantial evidence of moral hazard in insurance markets, which would imply that the social optimum prescribes less than full insurance (Aron-Dine, Einav and Finkelstein, 2013). In this subsection, we consider the effect of introducing a simple form of moral hazard.

In the model of Subsection 2.4, suppose that mean health expenditures μ respond to out-of-pocket expenses according to a constant elasticity function of the kind used in Saez (2001). Then, mean risk at an actuarial rate of x is $\mu(x) = \mu\left(\hat{x}\right)\left(\frac{1-\hat{x}}{1-x}\right)^{\epsilon}$, where ϵ is the constant elasticity of utilization and $\hat{x}$ is a reference level of insurance against which $\mu(x)$ is measured. We calibrate this model to the Handel, Hendel and Whinston (2013) data using the canonical estimate of $\epsilon = 0.2$ for the elasticity of medical expenditures from the RAND experiment, and use as common reference level for all individuals 85%, which is the mean actuarial rate.[21] Then, the optimal actuarial rate is approximately 87%. Even a clearly-unrealistically high elasticity of 1 yields optimal insurance of about 78%. Therefore moral hazard reduces our estimate of socially optimal insurance rates, though not significantly. Given that in our calibration model, insurance is driven entirely out of the market (or at leaset no positive insurance equilibrium exists), our basic welfare conclusions based on assuming away moral hazard in Subsection 2.5 and Section 3 see likely to be robust to plausible degrees of moral hazard.

5.2 Market expansion

The model of competition in Section 4.4 assumes a covered market. At the cost of some additional notation, it is simple to relax this assumption while maintaining the simple structure of the model. If the market is not covered, there are two margins to which competing firms attend: the "exiting" consumers indifferent between a firm and no insurance, and the "switching" consumers indifferent between the two firms. We present here a summary of the main conclusions of this extension as details can be found in Veiga and Weyl (2012).

We maintain Assumption 3 of Section 4.4 and focus on LDPE. We then assume, following White and Weyl (2012) that the Hotelling transportation cost to the nearest insurer is zero, while to the furthest insurer is $t\,(1-2b)$. This preserves the difference in transportation costs

[21] On the RAND experiment see Manning et al. (1987) and for a more recent interpretation see Aron-Dine, Einav and Finkelstein (2013).

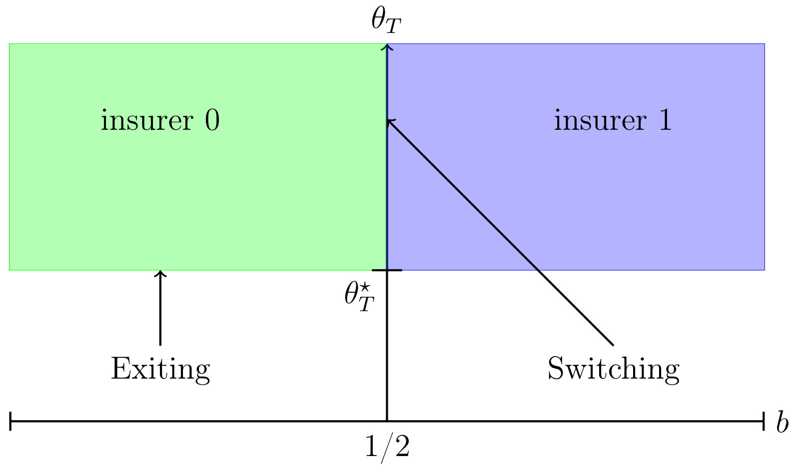

Figure 4: The $b - \theta_T$ plane of a competitive market with an expansion margin.

between the insurers and therefore the marginal incentives of consumers to choose between insurers. In turns this means that the switching margin behaves in the same way as in does in Section 4.4. However, at an LDPE no consumer pays a transportation cost and therefore, at each $b \in [0, 1]$, the set of types $\boldsymbol{\theta}$ purchasing insurance is the same. We can then follow Section 4 in assuming that, at an LDPE and for each value of $b \in [0, 1]$, consumers purchase insurance from the insurer "closest" to them if $\theta_T \geq \theta_T^\star$. Under these assumptions, consumers on the switching margin ($\left\{(\boldsymbol{\theta}, b) : b = \frac{1}{2}\right\}$) are representative of the set of all buyers, whereas consumers on the exiting margin ($\{(\boldsymbol{\theta}, b) : \theta_T = \theta_T^\star\}$) are not.

The presence of the exiting margin has several implications. Expectations can no longer be unconditional, but must distinguish between the set of switching consumers (which is identical to all participating consumers), exiting consumers, and all potential consumers. While this is not the focus of our analysis, equilibrium in this setting will feature a distorted quantity as a result of market power and the selection distortions of quantity first described by Akerlof (1970). Mahoney and Weyl (2013) provide a detailed analysis of this interaction between selection, market power and quantity provided, holding fixed quality.

Regarding the optimal choice of x, when competition is imperfect, a firm will maximize profit taking into consideration all its marginal consumers, and will thus consider a weighted average of switching consumers (representative of all buyers) and exiting consumers. Therefore we no longer assume away the Spence (1975) distortion. However, as competition increases, more weight placed on the switching margin and, in the limit as $t \to 0$, the firm caters exclusively to those consumers. Since they are representative of buyers overall the

Spence distortion vanishes in the competitive limit.[22] Furthermore, our characterization of x in the competitive limit in Corollary 2 must continue to hold as the number of switching consumers, and thus the incentive for cream-skimming, still explode in this limit, except that the expectations would be conditioned on the set of individuals buying insurance. Thus given our calibration in Section 3, zero insurance would still result unless the set of insurance purchasers has a radically different value of β or γ than does the whole population.[23] Therefore, in the limit, any potentially beneficial effect of competition is swamped by its harm in this case. It is possible, however, that for a different calibration coverage can be sustained in the perfectly competitive limit and in this case the potential benefits of competition in eliminating Spence and population coverage distortions would have to be traded off against the harms of competition through cream-skimming.

Socially optimal insurance is also impacted by the presence of the exiting margin. In the environment of Section 4.4, profit-maximizing sorting incentives were always socially harmful, since there was no sorting term in the FOC for welfare maximization in Proposition 7. When there is an exiting margin, even a social planner engages in an optimal amount of sorting. In fact, the monopolistic incentive for sorting (Theorem 1) is identical to the incentive faced by a social planner (Theorem 7). The reason is that, since marginal consumers are indifferent about buying the product and not, their contribution to social welfare is only their contribution to profit, so their value to a profit maximizing monopolist is the same as their value to a welfare maximizer. However, the incentive to sort of a duopolist is not welfare-maximizing because the firm does not internalize how cream-skimming affects its rivals profit. Thus, distortions in sorting incentives still increase with competition in this environment as they do in Section 4.4.

6 Conclusion

This paper makes three primary contributions. First, we provide a characterization of a firm's incentive to use non-price instruments in a selection market to attract the most profitable consumers. Second, we apply this characterization to a simple model of imperfect competition, deriving reduced-form statistics that characterize the equilibrium level of insurance. Third, we calibrate these characterizations to show the direction and size of sorting effects and their impact on social welfare.

From a policy perspective, our results emphasize that selection markets are those in

[22] See also White and Weyl (2012) on the interaction between the Spence distortion and competition.

[23] Given that Handel, Hendel and Whinston (2013)'s data is drawn from a pool of individuals all of whom are insured, it arguably is more representative of the set of insured individuals than the whole population anyway.

which it is socially optimal (but perhaps not physically necessary) to design products so that consumers differ in their value to the firm. If cost is observable ex-post, it can always be transferred back to the consumer, eliminating selection.[24] However, this would eliminate insurance opportunities through reclassification risk (Handel, Hendel and Whinston, 2013). It is therefore often optimal to tolerate selection in equilibrium for reasons like insurance, redistribution, administrative costs of ex-post verification or limited liability.[25] As Subsection 2.5 shows, analyzing an industry as a selection market instead of a price discrimination market therefore often implies "normative baggage" about the desirability of competition that is typically not driven by empirical measurements. The sorting incentive of Proposition 1 provides a simple measure of when a product characteristic will become the subject of cream-skimming. Moreover, since a monopoly's sorting incentive is the same as the social planner's, the former can be used to calibrate the level of socially optimal policy. However, before judging such cream-skimming outcome as distorted or efficient, it is important to determine an appropriate normative frame external to the dynamics of selection itself to decide whether cream-skimming is efficiently restoring incentives or inefficiently eroding insurance. In the latter case, when competition can be harmful, there is a case for restricting competition in the harmful non-price product characteristic.

However, blanket structural attempts to limit competition are unlikely to be optimal policy because competition on price is often beneficial to population coverage, particularly under adverse selection (Mahoney and Weyl, 2013). Despite this, such remedies may be superior to allowing cream-skimming to destroy a market altogether, if better regulation is infeasible. In such a second-best case, estimating reduced-form statistics such as the one we develop may help determine whether it is better to allow competition with its other benefits (efficient quantities, cost reductions or innovations) or whether competition is likely to be so destructive that it should be strictly avoided, as suggested by the calibration in Section 3. Such estimation may also help determine ways in which markets can be re-organized through various forms of pooling so that dimensions of heterogeneity offset one another within each pooled group thereby preventing harmful cream-skimming. In particular, employment relationships typically bundle together safety measures, multiple tasks, several types of insurance, etc. These packages may help to mitigate cream-skimming in sub-elements of the employment package, thereby allowing valued services to be provided at

[24] Often much selection can be eliminated by incorporating ex-ante information, such as detailed pricing on the basis of medical history, pre-existing conditions and demographics as in Finkelstein and Poterba (2006). In Akerlof (1970), selection disappears if damages can be costlessly awarded to the buyer for any negative surprises regarding the car's condition.

[25] Bénabou and Tirole (2013) give the example of incentive pay in an multi-task environment in the spirit of Holmström and Milgrom (1991) where it may be optimal to blunt pay-for-performance to avoid excessive risk-taking in a way that cream-skimming makes impossible.

competitive equilibria.

While we focused throughout the paper on applications to insurance markets, selection is of interest in many other contexts, including employment relationships, matching markets with non-transferable utility, platform markets, etc.[26] Research on these markets has, in recent years, increasingly focused on the possibility of rich, multidimensional heterogeneity to which our analysis is adapted.

Additionally, a number of theoretical issues remain unaddressed in our analysis. We do not consider the possibility of asymmetric "separating" equilibria, of more than two firms, or non-local deviations by firms. It would also be interesting to study alternatives to our assumption that switching consumers are representative of the full population of consumers, as Bonatti (2011) does in the Rochet and Stole (2002) context.

[26] An earlier version of this paper, Veiga and Weyl (2012) contains several examples

References

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

Armstrong, Mark, and John Vickers. 2001. "Competitive Price Discrimination." *RAND Journal of Economics*, 32(4): 579–605.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein. 2013. "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives*, 27(1): 197–222.

Bénabou, Roland, and Jean Tirole. 2013. "Bonus Culture: Competitive Pay, Screening, and Multitasking." http://www.princeton.edu/∼rbenabou/papers/Multitasking%20March%2019%20paper.pdf.

Bonatti, Alessandro. 2011. "Brand-Specific Tastes for Quality." *International Journal of Industrial Organization*, 29(5): 562–575.

Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié, and Francois Salanie. 2006. "Asymmetric Information in Insurance: General Testable Implications." *The RAND Journal of Economics*, 37(4): 783–798.

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review*, 97(3): 745–788.

Dasgupta, Partha, and Eric Maskin. 1986. "The Existence of Equilibrium in Discontinuous Economic Games, I: Theory." *The Review of Economic Studies*, 53(1): 1–26.

Einav, Liran, Amy Finkelstein, and Jonathan Levin. 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics*, 2(1): 311–336.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *Quarterly Journal of Economics*, 125(3): 877–921.

Einav, Liran, and Amy Finkelstein. 2011. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives*, 25(1): 115–138.

Fabinger, Michal, and E. Glen Weyl. 2014. "Price Discrimination is Typically Welfare-Enhancing." This paper is under preparation. Contact Glen Weyl at weyl@uchicago.edu for notes.

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

Finkelstein, Amy, and James Poterba. 2006. "Testing for Adverse Selection with "Unused Observables"." http://economics.mit.edu/files/699.

Finkelstein, Amy, and Kathleen McGarry. 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review*, 96(4): 938–958.

Handel, Ben, Igal Hendel, and Michael D. Whinston. 2013. "Equilibria in Health Exchanges: Adverse Selection vs. Re-Classification Risk." https://www.econ.berkeley.edu/%7Ebhandel/wp/ASRC_2013.pdf.

Hicks, J. R. 1939. "The Foundations of Welfare Economics." *Economic Journal*, 49(196): 696–712.

Holmström, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics & Organization*, 7(Special Issue): 25–52.

Hotelling, Harold. 1929. "Stability in Competition." *Economic Journal*, 39(153): 41–57.

Kaldor, Nicholas. 1939. "Welfare Propositions in Economics and Interpersonal Comparisons of Utility." *Economic Journal*, 49(145): 549–552.

Mahoney, Neale, and E. Glen Weyl. 2013. "Imperfect Competition in Selection Markets." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2372661.

Manning, Willard G, Joseph P Newhouse, Naihua Duan, Emmett B Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *The American Economic Review*, 77(3): 251–277.

Miyazaki, Hajime. 1977. "The Rat Race and Internal Labor Markets." *The Bell Journal of Economics*, 8(2): 394–418.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and Product Quality." *Journal of Economic Theory*, 18(2): 301–317.

Riley, John G. 1979. "Informational Equilibrium." *Econometrica*, 47(2): 331–359.

Rochet, Jean-Charles, and Lars A. Stole. 2002. "Nonlinear Pricing with Random Participation." *Review of Economic Studies*, 69(1): 277–311.

Rochet, Jean-Charles, and Philippe Choné. 1998. "Ironing, Sweeping, and Multidimensional Screening." *Econometrica*, 66(4): 783–826.

Rothschild, Michael, and Joseph E. Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*, 90(4): 629–649.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68(1): 205.

Schmidt, Klaus D. 2003. "On the Covariance of Monotone Functions of a Random Variable." http://www.math.tu-dresden.de/sto/schmidt/dsvm/dsvm2003-4.pdf.

Sheshinski, Eytan. 1976. "Price, Quality and Quantity Regulation in Monopoly Situations." *Economica*, 43(170): 127–137.

Smart, Michael. 2000. "Competitive Insurance Markets with Two Unobservables." *International Economic Review*, 41(1): 153–170.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *Bell Journal of Economics*, 6(2): 417–429.

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

Veiga, André, and E. Glen Weyl. 2012. "Multidimensional Production Design." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1935912.

Veiga, Andre, and E Glen Weyl. 2013. "The Leibniz Rule for Multidimensional Heterogeneity." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2344812.

Wambach, Achim. 2000. "Introducing Heterogeneity in the Rothschild-Stiglitz Model." *Journal of Risk and Insurance*, 67(4): 579–591.

White, Alexander, and E. Glen Weyl. 2012. "Insulated Platform Competition." http://alex-white.net/Home/Research_files/WWIPC.pdf.

Wilson, Charles. 1977. "A Model of Insurance Markets with Incomplete Information." *Journal of Economic Theory*, 16(2): 167–207.

Yin, Xiangkang. 2004. "Two-Part Tariff Competition in Duopoly." *International Journal of Industrial Organization*, 22(6): 799–820.

A Computational signing of monopoly sorting incentive

We consider that each consumer faces a normally distributed wealth loss $l \sim \mathcal{N}(\mu, \sigma^2)$, where μ is heterogeneous between consumers. Let $f(l, \mu\sigma^2)$ denote the Gaussian density of shocks of an individual with type μ. The insurer's policy prescribes a payment $G(l, x)$ when the insurer incurs a loss l and the insurer's instrument is x. Consumers have CARA utility $-e^{-aw}$, where w is final wealth and a is the (heterogeneous between consumers) absolute risk aversion. Let the (homogeneous) initial wealth be w_0.

Without insurance, final wealth is $w_N = w_0 - l$. With insurance, it is $w_I = w_0 - l + G(l, x) - p$. Then, expected surplus from insurance in

$$U(x, a, \mu, p) = \int \left[-e^{-aw_I} - \left(-e^{-aw_N}\right)\right] f\left(l, \mu\sigma^2\right) dl.$$

Then, we express the elements of

$$S(x, p, \mu) = \frac{\partial^2 u/\partial x \partial a}{\partial u/\partial a} - \frac{\partial^2 u/\partial x \partial \mu}{\partial u/\partial \mu}$$

in terms of the function $U(\mu, a, x, p)$, by applying the Implicit Function Theorem to the expression $U(x, a, \mu, u(x, a, \mu)) \equiv 0$. We obtain $\frac{\partial u}{\partial x} = -\dfrac{\partial U/\partial x}{\partial U/\partial p}$, $\frac{\partial u}{\partial v} = -\dfrac{\partial U/\partial v}{\partial U/\partial p}$ and $\frac{\partial u}{\partial \mu} = -\dfrac{\partial U/\partial \mu}{\partial U/\partial p}$. Moreover, we have

$$\frac{\partial^2 u}{\partial x \partial v} = \frac{\partial}{\partial v}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial v}\left(-\frac{\partial U/\partial x}{\partial U/\partial p}\right) = -\frac{\left(\frac{\partial U}{\partial x \partial v} + \frac{\partial U}{\partial x \partial p}\frac{\partial u}{\partial v}\right)\frac{\partial U}{\partial p} - \left(\frac{\partial U}{\partial p \partial v} + \frac{\partial U}{\partial p \partial p}\frac{\partial u}{\partial v}\right)\frac{\partial U}{\partial x}}{\left(\partial U/\partial p\right)^2}$$

$$\frac{\partial^2 u}{\partial x d\mu} = \frac{\partial}{\partial \mu}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial \mu}\left(-\frac{\partial U/\partial x}{\partial U/\partial p}\right) = -\frac{\left(\frac{\partial U}{\partial x \partial \mu} + \frac{\partial U}{\partial x \partial p}\frac{\partial u}{\partial \mu}\right)\frac{\partial U}{\partial p} - \left(\frac{\partial U}{\partial p \partial \mu} + \frac{\partial U}{\partial p \partial p}\frac{\partial u}{\partial \mu}\right)\frac{\partial U}{\partial x}}{\left(\partial U/\partial p\right)^2}.$$

Following Handel, Hendel and Whinston (2013), we set $w_0 = 10^6$ and $\sigma^2 = 10^8$. We consider $a \in [10^{-5}, 10^{-3}]$ and $\mu \in [0, 5 \times 10^4]$. We evaluate $S = \frac{\partial^2 u/\partial x \partial a}{\partial u/\partial a} - \frac{\partial^2 u/\partial x \partial \mu}{\partial u/\partial \mu}$ as a function of a, p and x, knowing that there exists some function $\mu^\star(x, p, a)$ that determines the boundary, which can be characterized by the equations above. In practice, we consider several discrete values for p and x. Then, we make draws of a from the range on which where is some type $\mu^{\star}(p,x,a) > 0$ that is marginal. For each such value of a, we compute the $\mu^{\star}(p,x,a)$ that makes a consumer indifferent and then compute $S(p,x,\mu = \mu^{\star}(p,x,a))$.

We consider $p \in \{16000, 20000, 24000\}$. We focus on relatively high levels of insurance, since these are the ones with greater economic relevance. In particular, other than x, insurance is complete. In each case, we consider 3 values of x equally spaced within a given range, described for each of the instruments below. In the graphs presented, the top row corresponds to $p = 16000$, while the bottom row has $p = 24000$. Each column corresponds to a level of x, with the left column capturing the lowest value of x. Notice that, in the case of a deductible, a low x captures more generous insurance). At times, the estimated value of $\mu^{\star}$ converges to a negative value, which should be ignored.

First, we consider the case where x is a regular actuarial rate, then $G(l,x) = \max\{0, xl\}$. The range of the instrument is $x \in \left[\frac{1}{2}, 1\right]$. It is clear from the graphs that $S < 0$. Since higher values of x correspond to better insurance, insurance sorts adversely.

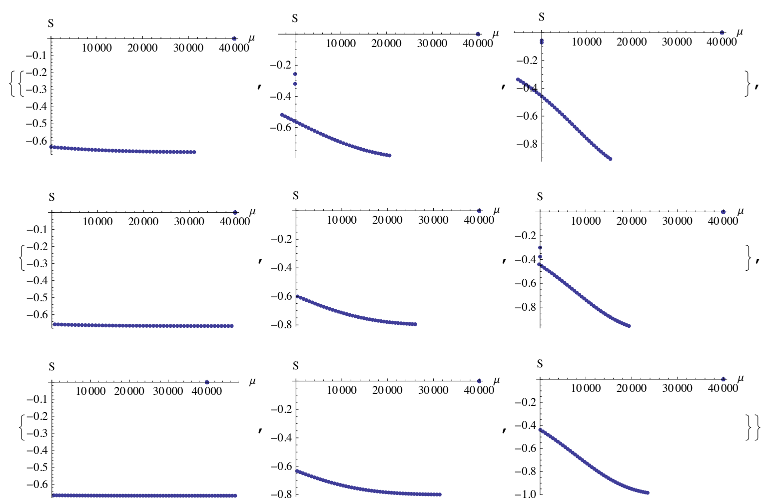

Figure 5: The sorting sign S with an actuarial rate on losses only for various values of that rate (corresponding to columns) and various prices (corresponding to rows) as described in the text. Each graph is S as a function of μ, where a responds to keep the individual marginal to purchasing.

Second, we consider the case where x is a deductible, then $G(l,x) = \max\{0, l - x\}$. The considered range is $x \in [0, 2000]$. It is clear from the graphs that $S > 0$. Since higher values of x correspond to worse insurance in the case of a deductible, more insurance sorts adversely in this case also.

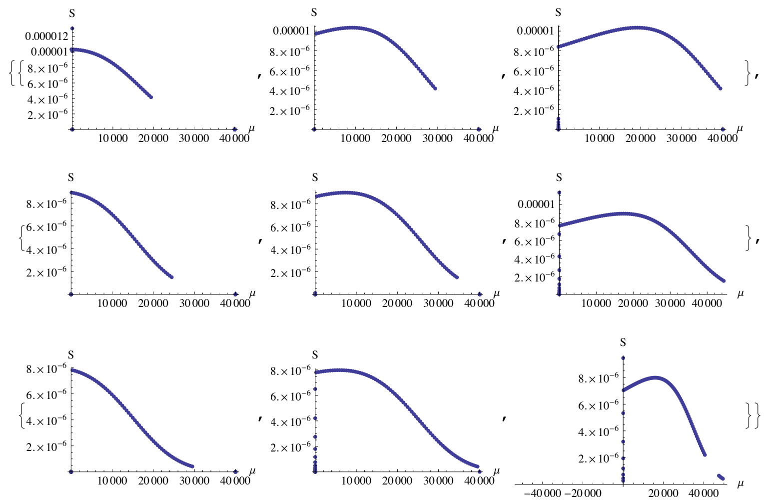

Figure 6: The sorting sign S with a deductible rate for various values of the deductible (corresponding to columns) and various prices (corresponding to rows) as described in the text. Each graph is S as a function of μ, where a responds to keep the individual marginal to purchasing. Because increasing x *decreases* the amount of insurance, the sign is reversed relative to the other cases.

Third, we consider that the insurer's instrument is a cap on the total amount covered. We have $G(x,l) = \min\{\max\{0,l\}, x\}$ and the considered range is $x \in [30000, 70000]$. It is clear from the graphs that $S < 0$. In this case, higher values of x correspond to better insurance, so more generous insurance sorts adversely in this case also.

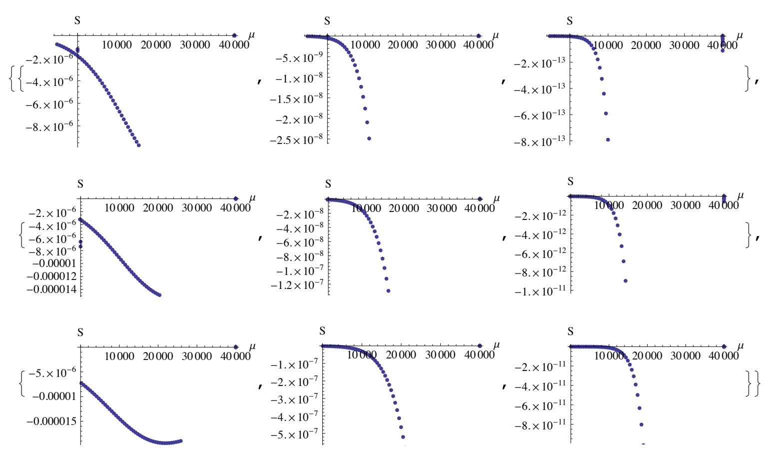

Figure 7: The sorting sign S with an indemnity cap for various values of that cap (corresponding to columns) and various prices (corresponding to rows) as described in the text. Each graph is S as a function of μ, where a responds to keep the individual marginal to purchasing.

Finally, we consider the case of an instrument which sorts advantageously at near full insurance. We consider that the insurer covers the loss l with probability x, but with probability $1-x$, the consumer incurs the full wealth shock and pays the insurer the insurance premium. Final wealth without insurance is $w_N = w_0 - l$. With insurance, with probability x final wealth is $w_{I+} = w_0 - p$, and with probability $1-x$ final wealth is $w_{I-} = w_0 - l - p$. Then, surplus from insurance in

$$U(x, a, \mu, p) = \int \left[x\left(-e^{-aw_{I+}}\right) + (1-x)\left(-e^{-aw_{I-}}\right) - \left(-e^{-aw_N}\right) \right] f\left(l, \mu\sigma^2\right) dl.$$

The range of x considered is $x \in [0.9, 1]$.

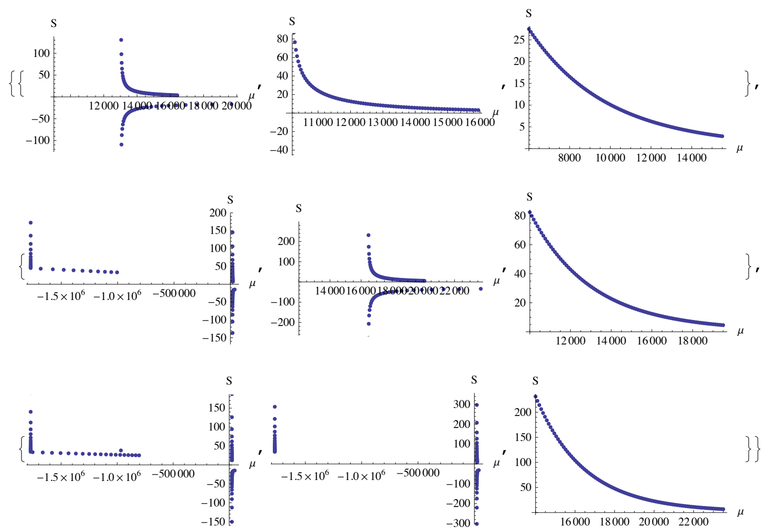

Figure 8: The sorting sign S with comprehensiveness for various values of that comprehensiveness (corresponding to columns) and various prices (corresponding to rows) as described in the text. Each graph is S as a function of μ, where a responds to keep the individual marginal to purchasing.

While the algorithm struggles to converge consistently for the smaller values of x, it is clear that for $x = 1$ we have $S > 0$. In this case, higher values of x correspond to more insurance, so insurance sorts advantageously in this case.

B Two-part tariff model (Proposition 6)

The setup is as described in Subsection 2.4. We have

$$\frac{\partial b^\star}{\partial p_0} = \frac{1}{2t}\frac{\partial U(w_0)}{\partial w}\frac{\partial w}{\partial p_0} = -\frac{1}{2t}\frac{\partial U(w_0)}{\partial w} < 0$$

and

$$\frac{\partial b^\star}{\partial x_0} = \frac{1}{2t}\frac{\partial U(w_0)}{\partial w}\frac{\partial w}{\partial x_0} = -\frac{1}{2t}\frac{\partial U(w_0)}{\partial w}(-q^\star) = -\frac{\partial b^\star}{\partial p_0}q^\star > 0.$$

We have $\frac{\partial w}{\partial L} = x\frac{1}{2}(x-2A) < 0$ because consumers only purchase medical services if $x < A$, so $x^\star - 2A < 0$. Moreover, $\frac{\partial w}{\partial x^\star} = \frac{1}{2}(x-A)L = -q^\star(x,L) < 0$. Healthcare utilization has constant marginal cost k. Then profit of provider 0 is $\pi_0 = \int_L \int_0^{b^\star} [p_0 + (x_0 - k) q^\star(L, x_0)] f(L)\, dbdL$. Define the vectors $\mathbf{p} = (p_0, p_1)$ and $\mathbf{x} = (x_0, x_1)$. Let

$$M = -\frac{\partial Q_0}{\partial p_0} = -\int_L \frac{\partial b^\star}{\partial p_0} f(L)\, dL.$$

$$\mathbb{E}_u[g(\mathbf{x},\mathbf{p},L) \mid b = b^\star] \;=\; \frac{1}{M}\int_L \frac{\partial b^\star}{\partial p_0} g(\mathbf{x},\mathbf{p},L) f(L)\, dL = \mathbb{E}_u[g(\mathbf{x},\mathbf{p},L)].$$

Notice that $\mathbb{E}_u[g(\mathbf{x},\mathbf{p},L)]$ is different from a standard expectation because it is weighted by $\frac{\partial U}{\partial w}$, but it is independent of both t and b.

At an LDPE we have $x_1 = x_0 = x$ and $p_1 = p_0 = p$. Let $q^\star = q^\star(L, x)$. Then we have

$$\frac{\partial \pi_0}{\partial p_0} \;=\; Q + \int_L \frac{\partial b^\star}{\partial p_0}[p + (x-k) q^\star(L,x)] f(L)\, dL = Q - M\mathbb{E}_u[p + (x-k) q^\star] = 0.$$

Recall $\frac{\partial q^\star}{\partial x} = -L$. Using this condition and the FOC for p, we obtain

$$\begin{aligned}
\frac{\partial \pi_0}{\partial x_0} &= \int_L \int_0^{b^\star} \left[q^\star + (x_0 - k)\frac{\partial q^\star}{\partial x}\right] f(L)\, dbdL + \int_L \frac{\partial b^\star}{\partial x_0}[p_0 + (x_0-k) q^\star] f(L)\, dL \\
&= Q\mathbb{E}[q^\star - (x-k)L \mid b < b^\star] + M\mathbb{E}_u[-q^\star[p + (x-k)q^\star] \mid b = b^\star] \\
&= Q\mathbb{E}[q^\star - (x-k)L] + M\mathbb{E}_u[-q^\star]\mathbb{E}_u[p + (x-k)q^\star] + M\mathrm{Cov}_u[-q^\star, p + (x-k)q^\star] \\
&= Q(\mathbb{E}[q^\star] - \mathbb{E}_u[q^\star]) - Q(x-k)(\mathbb{E}[L] + M\mathbb{V}_u[q^\star]).
\end{aligned}$$

Let $U_w = \frac{\partial U}{\partial w}$. We have

$$\mathbb{E}[q^\star] - \mathbb{E}_u[q^\star] = \mathbb{E}[q^\star] - \frac{\int U_w q^\star f dL}{\int U_w f dL} = \mathbb{E}[q^\star] - \frac{\mathbb{E}[U_w]\mathbb{E}[q^\star] + \mathrm{Cov}[U_w, q]}{\mathbb{E}[U_w]} = -\frac{\mathrm{Cov}[U_w, q^\star]}{\mathbb{E}[U_w]} < 0.$$

We know $\mathrm{Cov}[U_w, q^\star] > 0$ because, when the shock L is large, wealth w is small, then marginal utility U_w is large by concavity of U. Moreover, when the sock L is large, the quantity of healthcare consumer $q^\star$ is also large. Clearly, $\mathbb{E}[U_w] > 0$ also. Similarly, $\mathbb{E}[L] > 0$

and $\mathbb{V}_u[q^\star] > 0$ since it is a variance. If the FOC for x holds ($\frac{\partial \pi_0}{\partial x_0} = 0$), we must have $x - k < 0$.

C LDPE FOC and SOC (Proposition 4, Theorem 4 and Corollary 2)

We have $b^\star \equiv \frac{u(x_0,\boldsymbol{\theta})-p_0-(u(x_1,\boldsymbol{\theta})-p_1)}{2t} + \frac{1}{2}$. Then $\frac{\partial b^\star}{\partial p} = -\frac{1}{2t}$ and $\frac{\partial b^\star}{\partial x} = \frac{u'}{2t}$. Profit is $\pi = \int_{\boldsymbol{\theta}} \int_0^{b^\star} (p - c(x,\boldsymbol{\theta})) f(\boldsymbol{\theta}) \, db d\boldsymbol{\theta}$. We obtain

$$\frac{\partial \pi}{\partial p} = \int_{\boldsymbol{\theta}} \int_0^{b^\star} f(\boldsymbol{\theta}) \, db d\boldsymbol{\theta} + \int_{\boldsymbol{\theta}} -\frac{1}{2t}[p - c(x,\boldsymbol{\theta})] f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} = Q - M\mathbb{E}[p-c].$$

$$\frac{\partial \pi}{\partial p^2} = \int_{\boldsymbol{\theta}} -\frac{1}{2t} f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} + \int_{\boldsymbol{\theta}} -\frac{1}{2t} f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} = -2M < 0.$$

$$\frac{\partial \pi}{\partial p \partial x} = \int_{\boldsymbol{\theta}} \frac{u'}{2t} f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} + \int_{\boldsymbol{\theta}} -\frac{1}{2t}(-c') f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} = M\mathbb{E}[u' + c'] > 0.$$

$$\frac{\partial \pi}{\partial x} = \int_{\boldsymbol{\theta}} \int_0^{b^\star} -c' f(\boldsymbol{\theta}) \, db d\boldsymbol{\theta} + \int_{\boldsymbol{\theta}} \frac{u'}{2t}(p-c) f(\boldsymbol{\theta}) \, d\boldsymbol{\theta}.$$

$$\begin{aligned} \frac{\partial \pi}{\partial x^2} &= \int_{\boldsymbol{\theta}} \frac{u'}{2t}(-c') f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} + \int_{\boldsymbol{\theta}} \int_0^{b^\star} (-c'') f(\boldsymbol{\theta}) \, db d\boldsymbol{\theta} + \frac{1}{2t} \int_{\boldsymbol{\theta}} \{u''(p-c) + u'(-c')\} f(\boldsymbol{\theta}) \, d\boldsymbol{\theta} \\ &= M\{t\mathbb{E}[u'' - c''] - \mathrm{Cov}[u'', c] - 2\mathbb{E}[u']\mathbb{E}[c'] - 2\mathrm{Cov}[u', c']\}. \end{aligned}$$

Then the Hessian determinant is

$$\begin{aligned} H &= \frac{\partial \pi}{\partial p^2}\frac{\partial \pi}{\partial x^2} - \left(\frac{\partial \pi}{\partial p \partial x}\right)^2 = \\ &= -2MM\{t\mathbb{E}[u'' - c''] - \mathrm{Cov}[u'', c] - 2\mathbb{E}[u']\mathbb{E}[c'] - 2\mathrm{Cov}[u', c']\} - (M\mathbb{E}[u' + c'])^2 \\ &= -M^2\left\{2t\mathbb{E}[u'' - c''] - 2\mathrm{Cov}[u'', c] - 4\mathrm{Cov}[u', c'] + \mathbb{E}[u' - c']^2\right\}. \end{aligned}$$

Then H is positive if

$$t\mathbb{E}[\phi''] + \frac{1}{2}\mathbb{E}[\phi']^2 < \mathrm{Cov}[u'', c] + 2\mathrm{Cov}[u', c'].$$

Evaluating this SOC at $t = G = 0$ yields $\frac{1}{4}\frac{\mathbb{E}[\phi']^2}{\mathbb{V}[c']} - 1 < \frac{\mathrm{Cov}[\phi', c']}{\mathbb{V}[c']}$.