Constrained efficient equilibria in selection markets with continuous types*

Irina Gemmo, Christian Kubitza, and Casey Rothschild
ETH Zurich University of Bonn Wellesley College

June 26, 2020

Abstract

We prove the existence of the constrained efficient Miyazaki (1977)-Wilson (1977)-Spence (1978) equilibrium in competitive markets with adverse selection when the distribution of unobservable types is continuous. Our existence proof applies under extremely general assumptions about individual preferences. When we restrict preferences to have the widely-used-in-the-selection-markets-literature quasilinear form, we characterize the properties of this equilibrium by developing a simple and computationally efficient numerical method for constructing it. Applying this method, we show in a natural setting how one would compute the equilibrium allocation, potentially facilitating empirical work using the MWS equilibrium. We illustrate this empirical application in the context of policy interventions and show that the welfare implications of a coverage mandate critically hinge on whether the market implements a constrained efficient allocation like the MWS equilibrium or a constrained inefficient allocation like in Azevedo and Gottlieb (2017).

JEL classification: D82, G22, D41.

Keywords: Asymmetric and private information; adverse selection; insurance markets; equilibrium existence.

* We are grateful for comments and suggestions by Daniel Gottlieb, Nathaniel Hendren, Pierre Picard, Florian Scheuer, and Glen Weyl. Casey Rothschild gratefully acknowledges a travel grant from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE. Irina Gemmo and Christian Kubitza gratefully acknowledge financial support from the International Center for Insurance Regulation (ICIR) at Goethe-University Frankfurt.

1 Introduction

Economists have long understood that private information can lead to adverse-selection-driven pathologies in competitive markets. Adverse selection typically prevents first-best outcomes from being achieved. Nonetheless, depending on market dynamics, markets may still achieve second-best (i.e., constrained efficient) outcomes. Indeed, the so-called Miyazaki (1977)-Wilson (1977)-Spence (1978) (henceforth MWS) equilibrium concept, used widely in the insurance market literature in models with small type spaces, predicts a constrained efficient allocation. The first contribution of this paper is to extend the MWS equilibrium concept to richer type spaces. Specifically, we prove the existence of the MWS equilibrium in models with continuous one-dimensional type spaces under general assumptions about preferences.

The second contribution of this paper is to characterize that equilibrium. Specifically, under the assumption of quasilinear preferences,[1] we develop a simple and efficient algorithm for computing the MWS equilibrium with a continuum of types. This algorithm—which amounts to solving a sequence of straightforward differential equations—demonstrates the value of our continuous type extension of the MWS equilibrium concept: much as in the optimal income tax literature following Mirrlees (1971), it is *easier*, both theoretically and computationally, to characterize equilibrium allocations in continuous type spaces than it would be to do so in large but finite type spaces. As such, our algorithm thereby facilitates what is, to the best of our knowledge, the first qualitative (and quantitative) characterization of MWS equilibrium in rich type spaces and opens the door to empirical work comparing the applicability of this constrained-efficient concept to other concepts such as the one developed in Azevedo and Gottlieb (2017).

If the market dynamics in a particular market make the MWS a good predictor for market outcomes, then our construction is of direct interest. Independent of whether it ends up being a good predictor for market outcomes, our construction is also important for analyzing policy interventions. For example, the equilibrium concepts described in Azevedo and Gottlieb (2017) and Einav et al. (2010a) are commonly used to evaluate the welfare effects of policy interventions, and both

[1] This is a standard assumption in the modern literature (e.g., Einav et al. (2010a, 2013), Handel (2013), Weyl and Veiga (2017)) and significantly less restrictive in an insurance context than in an optimal income tax context.

of these equilibrium concepts often predict market outcomes that fail to be constrained efficient. As such, the overall welfare effects of policy interventions conflate two conceptually distinct effects: (1) welfare changes that arise because of the interventions' interactions with the information frictions inherent in the economic environment, and (2) welfare changes that arise because of the intervention's interaction with the inefficiencies caused by the market dynamics underlying the equilibrium concept. Our construction of the MWS equilibrium mutes the second source and, thus, can be used to decompose these two sources of welfare effects.

We set up and describe the MWS equilibrium in Section 2, prove that it exists in Section 3, and describe a computational algorithm for it in Section 4. We then demonstrate the policy-relevance of our results by contrasting the welfare effects of coverage mandates in insurance markets with constrained efficient MWS equilibrium outcomes with the welfare effects in the Azevedo and Gottlieb (2017) equilibrium. Specifically, in Section 5, we describe a plausibly-calibrated model of an insurance market and compute and illustrate both the MWS and AG equilibria in the absence of a mandate. Consistent with the fact that the AG equilibrium is not constrained efficient, we show that total welfare (as measured by consumer surplus) is roughly 20% higher in the MWS equilibrium. We then re-compute the equilibria in the presence of a mandate that requires insurance policies to cover at least M% of expected losses. Such a mandate has large welfare effects in AG equilibrium: it increases consumer surplus by an order of 1% for $M = 50\%$ and by over 25% as M approaches 100%. In contrast, the welfare gains from such a mandate in the MWS equilibrium are only about 10% as large. Hence, our example indicates that the majority—on the order of 90%—of the welfare gains from imposing a mandate in the AG equilibrium world are attributable to the fact that the AG equilibrium concept is not constrained efficient, and a mandate improves welfare largely by moving the equilibrium allocation closer to the constrained-efficient frontier. Section 6 offers some brief conclusions.

2 Setup and definitions

We aim to extend the previous literature employing the MWS equilibrium concept (e.g., Spence (1978), Netzer and Scheuer (2014), Picard (2014, 2019)), all of which considers models with a discrete number of types, to allow a continuum of types.

"Types" differ in their privately known cost $c \in [\underline{c}, \bar{c}]$, $0 < \underline{c} < \bar{c} < \infty$. Our interest is in the case where c has a continuous distribution F over $[\underline{c}, \bar{c}]$ with density f uniformly bounded by $0 < \underline{f} \leq f \leq \bar{f} < \infty$. Our existence proof for this continuous type case builds on sequences of finite-type approximations.

A type c's individual allocation is a quantity-price pair $(q(c), p(c)) \equiv \vec{A}(c)$; her associated utility is $V(q(c), p(c); c)$. A firm's profit from selling (q, p) to type c is $\Pi(q, p; c) = p - cq$. A market allocation (or simply *allocation*) is the function $\vec{A}(\cdot)$ mapping types c into their individual allocations $\vec{A}(c)$.

We allow for very general consumer preferences, which are, in particular, more general than the commonly-used assumption of quasi-linear preferences (employed, e.g., in Einav et al. (2010a, 2013), Handel (2013), Weyl and Veiga (2017)). Following Spence (1978), we assume that utility is twice differentiable, increasing and concave in coverage ($V_q > 0$ and $V_{qq} < 0$), decreasing in price ($V_p < 0$), and strictly quasiconcave in (q, p) for each c. We assume that $|V_c| < k < \infty$ uniformly; it is then without loss of generality to assume that $V_c < \xi < 0$; we make this assumption henceforth.[2] By strict quasiconcavity, the marginal rate of substitution, $MRS = -V_q/V_p$ is positive and decreasing to the northeast along indifference curves in (q, p) space. We additionally assume the single crossing property—that MRS is increasing in c—and that MRS is uniformly bounded by $\underline{MRS} < MRS < \overline{MRS}$.[3] Finally, we assume that for each type c there exists a finite and unique "full insurance" level $q^*(c)$ satisfying $MRS(q^*, q^*c; c) = c$.

2.1 MWS equilibrium in a two-type setting

An example of the general model is the classic two-type model of Rothschild and Stiglitz (1976), in which preferences take the expected utility form $V(q, p; c) = (1 - c)u(w - p) + cu(w - p - l + q)$, p is an insurance premium, q is the gross indemnity in the event of an accident which causes a financial loss of size l out of initial wealth w, and c is the probability of experiencing such a loss. For readers unfamiliar with the MWS equilibrium concept, we reprise here the Rothschild-Stiglitz equilibrium candidate (henceforth "RS allocation") and the MWS equilibrium allocations in

[2] If a utility function $\tilde{V}$ does not satisfy this property, the cardinally equivalent $V = \tilde{V} - (k + \xi)c$ does.

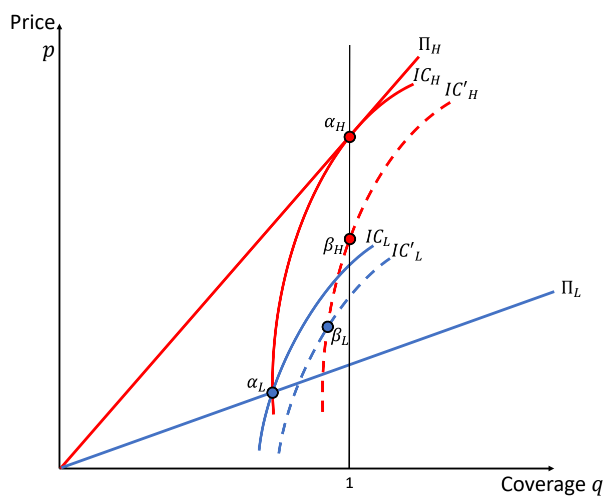
[3] Given the bounded type space, this last assumption is essentially without loss of generality.

this classic setting.[4]

The pair (α_L, α_H) in Figure 1 depicts the RS allocation. In this allocation, the high-cost type receives contract α_H which provides full insurance (here, unit coverage) at her type-specific actuarially fair price (as represented by the actuarially fair line Π_H in the diagram). The low-cost type receives the contract α_L which provides less than full coverage at her type-specific actuarially fair price. Specifically, the L type gets as much coverage as is possible without violating the incentive compatibility constraint that H-types do not prefer α_L to α_H (illustrated in the figure by the fact that α_L lies on the H-type indifference curve IC_H through α_H).

Figure 1: Illustrating the difference between the RS allocation and MWS equilibrium allocation in a canonical two-type model.

Lines Π_L and Π_H depict the type-specific actuarially fair contract pairs (q, p) for the L and H types, respectively. IC_L and IC'_L (respectively, IC_H and IC'_H) depict L-type (respectively, H-type) indifference curves. The pair (α_L, α_H) depicts the standard Rothschild-Stiglitz (RS) equilibrium allocation. The incentive compatible allocation (β_L, β_H) Pareto dominates the RS allocation, and is resource feasible when the proportion of high-cost types is sufficiently low; when such an allocation exists, the MWS equilibrium will diverge from the RS equilibrium and will involve Pareto-improving cross-subsidies from low- to high-cost types.



Alternative pairs of contracts like (β_L, β_H) in Figure 1 are incentive compatible and Pareto dominate (α_L, α_H). Since β_L lies above the low costs' individual zero profit line, such allocations (β_L, β_H) will also be weakly profitable for firms, on

[4]See Mimra and Wambach (2014) for a more detailed discussion.

average, if the proportion of high-cost types is sufficiently small. In other words, there may exist incentive compatible and resource feasible allocations which involve cross subsidies from the low- to high-cost types and that are Pareto-improving relative to the RS allocations. When such Pareto-improving contracts exist, the MWS equilibrium diverges from the RS allocation. Specifically, relative to the RS allocation, the MWS equilibrium implements any and all Pareto-improving cross-subsidies from L to H types, resulting in the best-for-L contract pair among the (information and resource) constrained efficient allocations.

There is a rich literature providing formal equilibrium microfoundations for the MWS equilibrium in the two-type case. The eponymous papers (Miyazaki (1977)-Wilson (1977)-Spence (1978)) describe a quasi-game-theoretic "anticipatory" equilibrium notion. This notion posits that firms who are considering "cream-skimming" deviations (which attract only the profitable low-cost types) will anticipate that incumbent firms will then exit the market, leaving them *also* attracting the high-cost types and becoming unprofitable. Netzer and Scheuer (2014) provide a modern game-theoretic formalization of this notion. Mimra and Wambach (2019a) formulate a formal dynamic game with both anticipation *and* reactionary contract offers and show that the MWS allocation is an equilibrium of this game. Picard (2014) introduces into the basic Rothschild and Stiglitz (1976) model the possibility that firms can endogenously choose to become mutual insurers, and shows that this leads to the MWS equilibrium allocation in a *static* game among insurers.[5]

2.2 MWS equilibrium with continuous types

This section builds up to a definition of the MWS equilibrium in the continuum of types case, as outlined above. To that end, we first provide an alternative formulation of the MWS equilibrium in the finite type case considered in Spence (1978), adapted to our notation. Consider a discrete set of types $c_1 < c_2 < \cdots < c_n$ with

[5] Intuitively, introducing mutual firms microfounds the "anticipatory" concept because mutual firms pay ex-post dividends if they are profitable and levy ex-post supplemental premiums if they end up being unprofitable. "Cream-skimming" deviations would render mutuals unprofitable, leading to the anticipation of supplemental premiums by high-cost types, making the mutual firm less appealing to high-cost types. This replicates—without a second stage where the incumbent firms make an additional move—the anticipation that unprofitable firms will exit the market.

probability masses $f(c_i) > 0$, $\sum_{i=1}^{n} f(c_i) = 1$.[6] Spence (1978)'s approach defines a set of reservation utilities $\bar{V}(c_i)$ for each i recursively. Specifically, define:

$$\bar{V}(c_i) \equiv \max_{\{\vec{A}(c_j)\}_{j\geq i}} V(\vec{A}(c_i); c_i) \tag{1}$$

subject to

$$V(\vec{A}(c_j); c_j) \geq V(\vec{A}(c_k); c_j) \quad \forall\, j,k \geq i \text{ and} \tag{2}$$

$$\sum_{j=i}^{n} \Pi(\vec{A}(c_j); c_j) f(c_j) \geq 0 \text{ and} \tag{3}$$

$$V(\vec{A}(c_j); c_j) \geq \bar{V}(c_j) \quad \forall j > i. \tag{4}$$

Inequalities (2) and (3) are, respectively, the standard incentive and (aggregate) break-even or zero-profit constraints. Constraints (4) are minimum utility constraints. The minimum utility values $\bar{V}$ are defined recursively, and, intuitively speaking, reflect the "outside option" utility available to each type.[7] The outside option $\bar{V}(c_i)$ for type i defined in program (1)-(4) is precisely the utility that i types would get if they formed a sub-economy consisting only of i- and all higher-cost types. Spence (1978) shows that this program can be rationalized by an anticipatory cream-skimming argument: if the minimum utility constraint for a particular type m were violated, then a new (or deviating) firm could offer the menu of contracts close to the one solving the program (1)-(4) for the m type that would be profitable if it attracts the m type and safe even if it attracts all $j > m$ types.

An allocation $\{\vec{A}(c_j)\}_{j=1}^{n}$ solving the lowest-cost type's sub-problem is called the *MWS equilibrium allocation*. Once reservation utilities $\bar{V}(c_i)$ are known, computing $\{\vec{A}(c_j)\}_{j=1}^{n}$ is straightforward. In the finite-type case, characterizing the reservation utilities $\bar{V}_i \equiv \bar{V}(c_i)$ is also conceptually straightforward, since it can be done recursively. Computing an allocation given a set of reservation utilities remains straightforward when we move to the continuum of types case, but, because recursion is impossible in that case, computing those reservation utilities raises new conceptual challenges.

[6] Note that indices i in Spence (1978)'s original formulation are inversely related to the cost c_i. We flip this dependence for a more consistent notation throughout the paper.

[7] Note that constraint (4) is null for $i = n$, so $\bar{V}(c_n)$ is well-defined. As such, constraint (4) is well defined for $i = n - 1$, and $\bar{V}(c_{n-1})$ is also well-defined. Similarly for $\bar{V}(c_{n-2})$, $\bar{V}(c_{n-3})$, and so forth.

Towards overcoming these challenges, it is useful to reformulate the programs (1)-(4) defining the reservation utilities in a non-recursive way. To that end, observe that the resource constraints (3) will always bind. The reservation utilities can therefore alternatively be characterized (dually) as the unique function $\bar{V}:\{c_1,\cdots,c_n\}\to\mathbb{R}$ with the property that, for all i,

$$0=\max_{\{\vec{A}(c_j)\}_{j\geq i}}\sum_{j=i}^{n}\Pi(\vec{A}(c_j);c_j)f(c_j) \tag{5}$$

$$\text{subject to}$$

$$V(\vec{A}(c_j);c_j)\geq V(\vec{A}(c_k);c_j)\quad\forall\, j,k\geq i \text{ and} \tag{6}$$

$$V(\vec{A}(c_j);c_j)\geq \bar{V}(c_j)\quad\forall j\geq i. \tag{7}$$

The dual approach generalizes naturally and as follows to the continuum of types case.[8]

Definition 1. *A MWS equilibrium is an allocation $\vec{A}^*$ that solves*

$$\max_{\{\vec{A}\}}\int_{c=\underline{c}}^{\bar{c}}\Pi(\vec{A}(c);c)dF(c) \tag{8}$$

$$\textit{subject to}$$

$$V(\vec{A}(c);c)\geq V(\vec{A}(c');c)\quad\forall\, c,c'\geq\underline{c}\textit{ and} \tag{9}$$

$$V(\vec{A}(c);c)\geq\bar{V}(c)\quad\forall c\geq\underline{c} \tag{10}$$

for some function $\bar{V}$ on the set of types for which, for every type $\hat{c}$,

$$0=\max_{\{\vec{A}\}}\int_{c=\hat{c}}^{\bar{c}}\Pi(\vec{A}(c);c)dF(c) \tag{11}$$

$$\textit{subject to}$$

$$V(\vec{A}(c);c)\geq V(\vec{A}(c');c)\quad\forall\, c,c'\geq\hat{c}\textit{ and} \tag{12}$$

$$V(\vec{A}(c);c)\geq\bar{V}(c)\quad\forall c\geq\hat{c}. \tag{13}$$

It is important to note that a MWS equilibrium is not the solution to *one* optimization problem, but to a *family* of optimization problems that interrelate to each

[8] For finite types, this definition corresponds to the standard notion used in the insurance literature (Spence (1978), Netzer and Scheuer (2014)).

other via the minimum utility constraints (13). Therefore, existence of the resulting allocation is non-trivial.

Conditional on knowing $\bar{V}$, it is straightforward to solve the optimization problem (8) - (10) and derive the equilibrium allocation. We therefore focus on the characterization of $\bar{V}$ in the existence proof and constructions that follow in the subsequent sections.[9]

3 Equilibrium existence

3.1 Discretizing the cost type distribution F

To prove the existence of an MWS equilibrium in the continuum of types case, we consider an increasingly fine sequence of finite approximations to the distribution of types F. As in Spence (1978), there is a well-defined MWS equilibrium set of allocations for each such discretization.

For each $n \in \mathbb{N}$, define the set of types C^n

$$C^n = \{c_0^n, c_1^n, \cdots, c_k^n, \cdots, c_{2^n+1}^n\} \tag{14}$$
$$= \left\{\underline{c}, \underline{c} + \frac{(\bar{c} - \underline{c})}{2^n}, \cdots, \underline{c} + k\frac{(\bar{c} - \underline{c})}{2^n}, \cdots, \bar{c}\right\} \tag{15}$$

and the corresponding cdf F^n via:

$$F^n(c) = \min_{c' \in C^n, c' \geq c} F(c).$$

The distribution F^n effectively 'collapses' all types in an interval $[c_k^n, c_{k+1}^n)$ under F onto the point c_k^n, so that the probability mass at c_k^n is (for each $0 \leq k \leq 2^n$) given by:

$$f^n(c_k^n) = F(c_{k+1}^n) - F(c_k^n).$$

[9]Note that although we refer to the allocation described above as an MWS "equilibrium", it is beyond the scope of our paper to provide explicit microfoundations (e.g., game theoretic) for this equilibrium. It is not hard to show, however, that the anticipatory logic described in Spence (1978) to the many-finite-type case can be readily extended to the continuum-of-types case, and we conjecture that the mutual-firm microfoundations of this anticipatory logic in Picard (2014) (viz footnote 5) and the many-finite-type extension in Picard (2019) extends as well.

Define

$$\mathcal{C} = \bigcup_{n \in \mathbb{N}} C^n \tag{16}$$

as the set of all types which appear in any discretization.

As in Spence (1978), for each discretization $F^n, n \in \mathbb{N}$, an MWS equilibrium exists. In particular, there is a well-defined and unique set of reservation utilities $\{\bar{V}^n(c_k^n)\}_{k=0,\cdots,2^n+1}$ that are consistent with Definition 1 of the MWS equilibrium for each $n \in \mathbb{N}$. Fixing any type $\hat{c} \in C^n$ and given the (recursively definable) reservation utilities $\bar{V}^n(c_k^n)$ for all $c_k^n \in C^n$ with $c_k^n \geq \hat{c}$, we refer to the mathematical program characterized by (11)-(13) "the MWS sub-problem for type $\hat{c}$". Define $\vec{A}^n(c;\hat{c})$ as the allocation of type c in a given solution to the MWS sub-problem for type $\hat{c}$ in discretization n (which is defined only if $c, \hat{c} \in C^n$ and $c \geq \hat{c}$). A key object in the proof that follows will be $\hat{V}^n(\hat{c}, T)$, which we define now.

Definition 2. *For $\hat{c} \in C^n$ and any $T \geq 0$, define*

$$\hat{V}^n(\hat{c}, T) \equiv \max_{\{\vec{A}^n(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]\cap C^n}} V(\vec{A}^n(c;\hat{c});c) \tag{17}$$

subject to

$$V(\vec{A}^n(c;\hat{c});c) \geq V(\vec{A}^n(c';\hat{c});c) \quad \forall c, c' \geq \hat{c} \quad \text{with} \quad c, c' \in C^n \text{ and} \tag{18}$$

$$V(\vec{A}^n(c;\hat{c});c) \geq \bar{V}^n(c) \quad \forall c \geq \hat{c} \quad \text{with} \quad c \in C^n \text{ and} \tag{19}$$

$$\sum_{c\in[\hat{c},\bar{c}]\cap C^n} \Pi(\vec{A}^n(c;\hat{c});c) f^n(c) \geq -T. \tag{20}$$

When $T = 0$, (17)-(20) coincides with the primal approach to the MWS equilibrium (i.e., (1)-(4)). Hence, $\hat{V}^n(\hat{c}, 0) = \bar{V}^n(\hat{c})$. When $T > 0$, the resource constraint is relaxed, and $\hat{V}^n(\hat{c}, T)$ can be interpreted as the maximum utility of type $\hat{c}$ in a MWS-like sub-problem involving a subsidy of size $T > 0$ to the set of types $c \in [\hat{c}, \bar{c}] \cap C^n$. We make one high-level technical assumption about $\hat{V}^n(\hat{c}, T)$:

Assumption 1. *For any $c^* < \bar{c}$, there exists an $N' > 0$ such that the family of functions $\{\hat{V}^n(\hat{c}, T)\}_{\hat{c}\in[\underline{c},c^*],n\geq N'}$ is equicontinuous in T at $T = 0$.*

With Assumption 1, we assume that, across all sub-problems and discretizations, the value of the MWS sub-program has uniformly bounded sensitivity to small subsidies. It is straightforward to show that this assumption holds in standard models.[10]

3.2 Convergence of discretized allocations

We now adapt an argument used by Hellwig (2007)[11] to show that there is a subsequence of these well-defined MWS equilibrium set of allocations which converges on a dense set of types c. We use the completion of this convergent subsequence to define a candidate MWS function $\bar{V}(c)$, which we show is continuous in c.

The following lemma shows that there exists a subsequence of discretizations that converges for all $c \in \mathcal{C}$.

Lemma 1. *There exists a subsequence $\{n_m\}_{m\in\mathbb{N}}$ of $\mathbb{N}$ such that $\lim_{m\to\infty} \vec{A}^{n_m}(c;\hat{c}) \equiv \vec{A}^{\infty}(c;\hat{c})$ exists for all $c, \hat{c} \in \mathcal{C}$.*

Define:

$$\bar{V}^*(c) \equiv V(\vec{A}^{\infty}(c;c);c) = \lim_{m\to\infty} \bar{V}^{n_m}(c) \quad \forall c \in \mathcal{C}, \tag{21}$$

where $\vec{A}^{\infty}$ is defined in Lemma 1. The following lemma shows that $\bar{V}^*(c)$ is readily extended to $c \in [\underline{c}, \bar{c}]$ and is continuous.

Lemma 2. *If Assumption 1 holds, then the limit $\lim_{c' \nearrow c, c' \in \mathcal{C}} \bar{V}^*(c') \equiv \bar{V}^*(c)$ exists, and $\bar{V}^*(c)$ is continuous in c.*

Proof. See Appendix A. □

3.3 Statement and proof of main theorem

We verify, in two steps, that this candidate MWS function $\bar{V}(c)$ is indeed an MWS equilibrium in the sense of Definition 1. The first step involves a simple continuity argument which establishes that the appropriately-taken limits of allocations in the discrete MWS problems are *feasible* in the continuous problem (i.e., satisfy

[10] Details for the many-type version of Rothschild and Stiglitz (1976)'s model are available in Online Appendix B.

[11] The proof, which we omit here, follows the same argument as for Helly (1921)'s Selection Theorem and as in Hellwig (2007)'s proof of Lemma B.3 (viz p. 812, where Hellwig cites Billingsley (1968)).

constraints (12) and (13)) and, moreover, yield zero profits at the limit. In the second step, we show that *no other feasible allocation* can yield positive profits in the limit problems. This second step is done by contradiction: if a feasible allocation did yield positive profits, then the continuity of $\bar{V}(c)$ could be used to construct a feasible allocation that would yield positive profits in some (sufficiently fine) discretization.

Theorem 1. *There exists a MWS equilibrium $\vec{A}^*$ corresponding to the reservation utility function $\bar{V}^*(c)$ defined in (21).*

Proof. Establishing the theorem requires showing that the function $\bar{V}(c) = \bar{V}^*(c)$, verifies the condition described in (11)-(13) in Definition 1. We do this in two steps. In Step 1, we use a limiting argument to construct, for each $\hat{c}$, an allocation $\{\vec{A}^*(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]}$ that satisfies (12) and (13) for which the value of (11) is 0. In Step 2, we show that there cannot exist any $\hat{c}$ and any associated allocation $\{\vec{A}^\dagger(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]}$ that satisfies (12) and (13) for which the value of (11) is strictly positive.

Step 1: Constructing $\{\vec{A}^*(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]}$.

We make extensive use of Lemma 1, which defines $\vec{A}^\infty(c;\hat{c})$ as a limit of a sequence $\{n_m\}_{m\in\mathbb{N}}$ of allocations $\vec{A}^{n_m}(c;\hat{c})$ for each $c,\hat{c}\in\mathcal{C}, c\geq\hat{c}$.

First, consider $\hat{c}\in\mathcal{C}$, and define $\vec{A}^*(c;\hat{c}) = \vec{A}^\infty(c;\hat{c})$ for all $c\in\mathcal{C}\cap[\hat{c},\bar{c}]$. Each component of $\vec{A}^\infty(c;\hat{c})$ is monotonic in c (which follows from single crossing and incentive compatibility). So $\lim_{\tilde{c}\nearrow c,\tilde{c}\in\mathcal{C}} \vec{A}^\infty(\tilde{c};\hat{c})$ and $\lim_{\tilde{c}\searrow c,\tilde{c}\in\mathcal{C}} \vec{A}^\infty(\tilde{c};\hat{c})$ are both well-defined and coincide except possibly at a countable number of points, which have measure 0 under the continuous distribution F. Extend $\vec{A}^*$ to $c\notin\mathcal{C}$ via

$$\vec{A}^*(c;\hat{c}) \equiv \lim_{c'\nearrow c, c'\in\mathcal{C}} \vec{A}^\infty(c';\hat{c}).$$

Then, $\{\vec{A}^\infty(c;\hat{c})\}_{c\in\mathcal{C}\cap[\hat{c},\bar{c}]}$ is incentive compatible for types $c\in\mathcal{C}\cap[\hat{c},\bar{c}]$, so $\{\vec{A}^*(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]}$ defined in this way is incentive compatible for all types $c\in[\hat{c},\bar{c}]$. Similarly, $V(\vec{A}^\infty(c;\hat{c});c)\geq\bar{V}^*(c)$ for all $c\in\mathcal{C}\cap[\hat{c},\bar{c}]$, so $V(\vec{A}^*(c;\hat{c});c)\geq\bar{V}^*(c)$ for all types $c\in[\hat{c},\bar{c}]$.

We will now show that $\int_{c=\hat{c}}^{\bar{c}}\Pi(\vec{A}^*(c;\hat{c});c)dF(c) = 0$ whenever $\hat{c}\in\mathcal{C}$. To that end, for each m, extend $\vec{A}^{n_m}(c;\hat{c})$ to all $c\in[\hat{c},\bar{c}]$ via

$$\vec{A}^{m*}(c;\hat{c}) = \vec{A}^{n_m}\left(\max\{\tilde{c}\in C^{n_m}\cap[\hat{c},c]\};\hat{c}\right). \tag{22}$$

That is, "assign" types c outside of C^{n_m} to the allocation of the closest lower-cost type in C^{n_m}. Exactly as in Hellwig (2007)'s Lemma B.1, the (almost everywhere) pointwise convergence of $\vec{A}^{m*}(\cdot;\hat{c})$ to $\vec{A}^{*}(\cdot;\hat{c})$ and the setwise convergence (here, weak convergence) of F^{n_m} to F implies

$$\int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^*(c;\hat{c});c)dF(c) = \lim_{m\to\infty} \int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^{m*}(c;\hat{c});c)dF^{n_m}(c) = \lim_{m\to\infty} 0 = 0. \quad (23)$$

Second, consider $\hat{c} \notin \mathcal{C}$, and take any sequence $k = 1, \cdots, \infty$ of $c_k \in \mathcal{C}$ with $c_k \nearrow \hat{c}$. Use the associated sequence of (sub) allocations $\{\vec{A}^*(c;c_k)\}_{c\in[c_k,\bar{c}]}$ to construct (via a diagonalization argument as alluded to in footnote 11) a subsequence $\{k_m\}$ that converges for each $c \in \mathcal{C} \cap [\hat{c},\bar{c}]$, and define $\vec{A}^*(c;\hat{c})$ as the limit for each such c. Complete the allocation by defining $\vec{A}^*(c;\hat{c})$ for $c \notin \mathcal{C}$ in terms of left-hand limits of $\vec{A}^*(c';\hat{c})$ for $c' \in \mathcal{C}$. The resulting allocation is incentive compatible and has $V(\vec{A}^*(c;\hat{c});c) \geq \bar{V}^*(c)$ for all $c \in [\hat{c},\bar{c}]$. Moreover, since

$$\lim_{m\to\infty} \int_{c=c_{k_m}}^{\hat{c}} \Pi(\vec{A}^*(c;c_{k_m});c)dF(c) = 0,$$

we have

$$\begin{aligned}
&\int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^*(c;\hat{c});c)dF(c) \\
&= \int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^*(c;\hat{c});c)dF(c) + \lim_{m\to\infty} \int_{c=c_{k_m}}^{\hat{c}} \Pi(\vec{A}^*(c;c_{k_m});c)dF(c) \\
&= \lim_{m\to\infty} \int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^*(c;c_{k_m});c)dF(c) + \lim_{m\to\infty} \int_{c=c_{k_m}}^{\hat{c}} \Pi(\vec{A}^*(c;c_{k_m});c)dF(c) \\
&= \lim_{m\to\infty} \int_{c=c_{k_m}}^{\bar{c}} \Pi(\vec{A}^*(c;c_{k_m});c)dF(c) = \lim_{m\to\infty} 0 = 0, \qquad (24)
\end{aligned}$$

where the last line follows from Equation (23). For each $c \in [\underline{c},\bar{c}]$, then, we have identified a feasible allocation $\{\vec{A}^*(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]}$ which satisfies (12) and (13) and yields zero profits.

Step 2: Showing that $\{\vec{A}^*(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]}$ is optimal.

Suppose, by way of contradiction, that there was some $\hat{c}$ and some other allocation $\{\vec{A}^\dagger(c)\}_{c\in[\hat{c},\bar{c}]}$ satisfying (12) and (13) with $\int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^\dagger(c);\hat{c})dF(c) = \delta > 0$. It is straightforward to show that if there were such an allocation, there would be another incentive compatible allocation $\{\vec{A}^\circ(c)\}_{c\in[\hat{c},\bar{c}]}$ with $\int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^\circ(c);\hat{c})dF(c) \geq$

$\delta/2$ and
$$V(\vec{A}^{\circ}(c);c) \geq V(\vec{A}^{\dagger}(c);c) + \varepsilon \tag{25}$$
for some $\varepsilon > 0$.[12] But, as we will now show, no such allocation can exist.

If it did, then $\{\vec{A}^{\circ}(c)\}_{c\in C^n \cap [\tilde{c},\bar{c}]}$ would be obviously incentive compatible in the nth discretization. For sufficiently high N, it would also satisfy $V(\vec{A}^{\circ}(c);c) > \bar{V}^N(c)$. To see this, recall that $\bar{V}^n(c)$ converges pointwise to $\bar{V}^*(c)$. $\bar{V}^*(c)$ is monotonic and, by Lemma 2, continuous in c, so $\bar{V}^n(c)$ in fact converges to $\bar{V}^*(c)$ uniformly in n. Hence we could find N large enough so that $\bar{V}^n(c) < \bar{V}^*(c) + \varepsilon$ for all $c \in C^n \cap [\tilde{c},\bar{c}]$ and $n > N$. Per equation (25), then, we would have $V(\vec{A}^{\circ}(c);c) > \bar{V}^n(c)$ for all $n > N$. But
$$\lim_{n\to\infty} \int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^{\circ}(c);c)dF^n(c) = \int_{c=\hat{c}}^{\bar{c}} \Pi(\vec{A}^{\circ}(c);c)dF(c) > \delta/2 > 0,$$
so, for sufficiently high $N' > N$, the allocation $\{\vec{A}^{\circ}(c)\}_{c\in C^N \cap [\hat{c},\bar{c}]}$ would be strictly profitable, incentive compatible, and it would satisfy the minimum utility constraints $V(\vec{A}^{\circ}(c);c) > \bar{V}^{N'}(c)\ \forall c \in C^{N'} \cap [\tilde{c},\bar{c}]$, contradicting the fact that $\bar{V}^{N'}(\cdot)$ are the reservation utilities associated with an MWS equilibrium in the N'th discretization. $\square$

4 Equilibrium construction

Having established existence of the MWS equilibrium for a continuous type space, we now show how to construct it. A major advantage of working in continuous type space in general is that it eases computation time. For example, assume that types are uniformly distributed and have the same optimal contract $q^* \equiv \arg\max_q [V(q, qc; c)]$. For a discretized type space with only 25 types, there are 624 constraints in the "Spence" program for the lowest cost type *alone* (24 minimum utility and $25 \cdot 24$ incentive compatibility constraints). It takes our computer roughly 2.2 minutes to solve for the equilibrium allocation recursively (starting

[12] One can construct such an allocation by letting $q^{\circ}(c) = q^{\dagger}(c)$, $p^{\circ}(\bar{c}) = p^{\dagger}(\bar{c}) - \tilde{\varepsilon}$, by integrating the incentive compatibility constraint down to find $p^{\circ}(c)$ for $c \in [\hat{c},\bar{c})$, and taking a sufficiently small $\tilde{\varepsilon}$.

with the highest cost type).[13] Solving only the final program (given reservation utilities $\bar{V}$ for higher cost types) takes roughly 40 seconds on its own. Below, we will derive a method that leverages the properties of the continuous type equilibrium and computes the equilibrium in this example roughly 100 times faster (in 0.35 seconds).

4.1 Assumptions

We only add one additional assumption to those in Section 3: we assume that preferences are linear in price, i.e., that utility is $V(q, p; c) = \nu(q; c) - p$ where $\nu(q; c)$ is continuous, increasing, and concave in q. This quasilinearity assumption is common in theoretical models for selection markets (e.g., Weyl and Veiga (2017), Levy and Veiga (2019)) and empirical applications (e.g., Einav et al. (2010a, 2013), Handel (2013), Hackmann et al. (2015), Handel et al. (2015, 2019)).

4.2 Construction

The main idea of the construction is to dissect the type space into intervals of types for which minimum utility constraints *do not* bind (i.e., are slack) and those for which they *do* bind, and to do this for a sequence of sub-problems of the form $[\hat{c}, \bar{c}]$ with successively lower values of $\hat{c}$. Figure 2 illustrates this idea. It plots, for several different $\hat{c}$'s, the difference $\Delta V(c; \hat{c}) \equiv V(\vec{A}(c; \hat{c}); c) - \bar{V}(c)$ between the *realized* utility of each type c in the $[\hat{c}, \bar{c}]$ interval and the reservation utility levels $\bar{V}(c)$ which, as in Definition 1, play the key role in the MWS construction.

The points labeled c^1 and c^2 separate values of $\hat{c}$ in conceptually distinct "steps" in our construction. Values of $\hat{c}$ in the interval $[c^1, \bar{c}]$—such as c^A in the diagram—correspond to what we call "step 1". Values of $\hat{c}$ in the interval $[c^2, c^1)$—such as c^B in the diagram—correspond to what we call "step 2". Values of $\hat{c} < c^2$, such as c^C in the diagram—correspond to what we call "step 3".

For $\hat{c}$ in step 1 (such as c^A), Figure 2 shows $\Delta V(c; c^A) > 0$ on $(c^A, \bar{c}]$: in this step (which may be empty), the entire interval of types $[\hat{c}, \bar{c}]$ have non-binding minimum utility constraints (13). Interior types in this interval are cross-subsidized by

[13] The code is implemented in MATLAB R2018b and uses MATLAB's optimization function *fmincon*.

types below them. As $\hat{c}$ is lowered within this step, the cross-subsidies grow, until c^1 is reached and we enter step 2.

For $\hat{c}$ in step 2 (such as c^B), Figure 2 shows $\Delta V(c; c^B) = 0$ on $[c^B, c^1]$ and $\Delta V(c; c^B) > 0$ on $(c^1, \bar{c}]$. Types in $[c^B, c^1]$ are "break-even" types for these subproblems, while groups $[c, \bar{c}]$ with $c > c^1$ are cross-subsidized. In other words, step 2 is characterized by an interval of break-even types at the bottom. As $\hat{c}$ is lowered within this step, the interval of break-even types expands without any effects on higher types.

For $\hat{c}$ in step 3 (such as c^C), there are two intervals of cross-subsidized types (with $\Delta V(c; c^C) > 0$), one at the top and one at the bottom, and an interval in the middle of break-even types with $\Delta V(c; c^C) = 0$. As $\hat{c}$ is lowered within this interval, the break-even interval at the bottom expands, "eating" away at the intermediate break-even interval. At some point, however, a new break-even interval at the bottom (a la step 2) may appear, and steps 2 (and then 3) can be iterated. Alternatively, as $\hat{c}$ is lowered, the whole break-even interval may be "eaten" and, as in step 1, no minimum utility constraints will bind.

Two key things will be critical as we build this sequence of steps. First, we need a way to compute the allocation of types *within* a step (which implicitly determines the reservation utilities $\bar{V}$), and a criterion for determining when a step "ends"– e.g., for finding c^1 and c^2 in the figure. In the subsections that follow, we show, intuitively, how to do both of these things in each step and explain the underlying logic. Formalizing this intuition is straightforward but tedious.[14]

We identify allocations via the coverage-utility pair (q, u), where $u \equiv v(q; c) - p$ is type c's utility from buying coverage q at price p. Then (because of single crossing and as in Mirrlees (1971)) we can replace the global incentive compatibility constraints (12) with local incentive constraints

$$u'(c) = v_c(q(c); c) \tag{26}$$

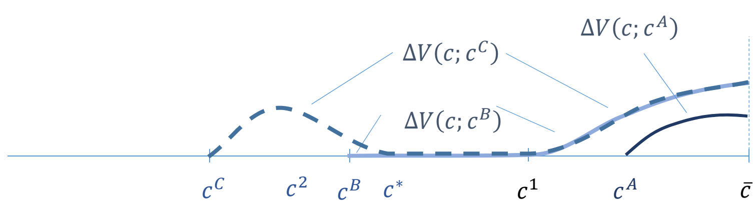
and monotonicity constraints

$$q'(c) \geq 0 \quad \forall c \in [\hat{c}, \bar{c}]. \tag{27}$$

[14] Details and pseudo-code for numerical implementation are available in Online Appendix C.

Figure 2: Qualitative illustration of equilibrium construction.

The construction illustrates the dissection of the type space into intervals of types for which minimum utility constraints *do not* bind (i.e., are slack) and those for which they *do* bind. It does this for a sequence of sub-problems for sub-economies consisting of an interval of types of $[\hat{c}, \bar{c}]$ with successively lower values of $\hat{c}$. For each $\hat{c}$, it plots the difference $\Delta V(c;\hat{c}) \equiv V(\vec{A}(c;\hat{c});c) - \bar{V}(c)$ between the *realized* utility of each type c in the $[\hat{c}, \bar{c}]$ interval and the *reservation* utility levels $\bar{V}(c)$. In sub-problem $[c^A, \bar{c}]$ (solid dark line), minimum utility constraints do not bind for any type; in sub-problem $[c^B, \bar{c}]$ (solid light line), minimum utility constraints do not bind for types $c \in [c^1, \bar{c}]$ but bind for types $c \in [c^B, c^1)$; and in sub-problem $[c^C, \bar{c}]$ (dashed line), minimum utility constraints do not bind for types $c \in [c^C, c^*]$ and types $c \in [c^1, \bar{c}]$ but do bind for types $c \in (c^*, c^1)$. See text for more details.



In this section, we follow the common-in-the-optimal-tax-literature (viz Rothschild and Scheuer (2013)) approach of dropping the monotonicity constraint. This approach is valid insofar as the resulting allocations are indeed monotonic. Otherwise, optimal bunching, wherein a range of types receive the same allocation, needs to be considered. Incorporating bunching is conceptually straightforward, but rotationally cumbersome. In the interest of expositional clarity, we relegate bunching considerations to Online Appendix C.3.

4.2.1 Step 1: Non-binding minimum utility constraints

As a first step, we solve, for each $\hat{c} \in [\underline{c}, \bar{c})$, the problem:

$$\max_{\{(q(c),u(c))\}_{c\in[\hat{c},\bar{c}]}} u(\hat{c}) \tag{28}$$

subject to the local incentive constraints (26) for all $c \in [\hat{c}, \bar{c}]$ and the resource constraint

$$\int_{\hat{c}}^{\bar{c}} (v(q(c);c) - u(c) - q(c)c)\, f(c)dc \geq 0. \tag{29}$$

This is the primal form of the $\hat{c}$-type MWS sub-program, relaxed by dropping the minimum utility constraints (13). The first order conditions for this family of problems imply that solutions $q^0(c)$ satisfy

$$\frac{v_q(q^0(c);c) - c}{v_{qc}(q^0(c);c)} = \frac{1 - F(c)}{f(c)}, \tag{30}$$

and hence are independent of $\hat{c}$. Given $q(c) = q^0(c)$, the local incentive constraint (26) defines a differential equation that determines the utility allocation

$$u^0(c;\hat{c}) = -\int_{c}^{\bar{c}} v_c(q^0(c');c')dc' + u^0(\bar{c};\hat{c})$$

up to the constant $u^0(\bar{c};\hat{c})$, which is in turn pinned down by the resource constraint (29).

Let c^1 be the smallest type for which $u^0(\bar{c};\hat{c})$ is non-increasing on the interval $\hat{c}$ on $[c^1, \bar{c}]$. For $\hat{c}$ in this interval, $u^0(c;\hat{c}) \geq u^0(c;c)$ for all $c \geq \hat{c}$. As such, $\{(q^0(c), u^0(c;c))\}_{c\in[\hat{c},\bar{c}]}$ solves (28) subject to (26), (29), *and* the minimum utility constraints $u(c) \geq u^0(c;c)$ for all $c \geq \hat{c}$. In other words, the utilities $\bar{V}(c) \equiv u^0(c;c)$ are the MWS reservation utilities (in the sense of Definition 1) for $c \in [c^1, \bar{c}]$.

4.2.2 Step 2: Binding minimum utility constraints

For $\hat{c}$ slightly *below* c^1, $u^0(c^1;\hat{c}) < u^0(c^1;c^1) \equiv \bar{V}(c^1)$. Consequently, minimum utility constraints are violated in the relaxed sub-problem for $\hat{c}$. Intuitively, this is because the $\hat{c}$ type does not want to cross-subsidize higher types. Indeed, it wants cross-subsidies *from* them. The minimum utility constraints prevent such "downward" cross-subsidies from taking place. In other words, there will be an interval below c^1 for which types break even (i.e., receive actuarially fair allocations); their allocation will satisfy:

$$u(c) = v(q(c); c) - cq(c). \tag{31}$$

This individual break-even condition (together with incentive compatibility) implies that

$$q'(c) = \frac{q}{v_q(q(c); c) - c}. \tag{32}$$

Using initial conditions from the allocation for type c^1, we solve (32) to determine the insurance coverage for break-even types, $q^1(c)$, and then use the local incentive compatibility constraints (26) to determine utility, $u^1(c)$.

We can construct $(q^1(c), u^1(c))$ in this way for all $c < c^1$. For any $\hat{c} < c^1$, we can consider the allocation which assigns $(q^1(c), u^1(c))$ to all types $c \in [\hat{c}, c^1)$ and $(q^0(c), u^0(c; c^1))$ to all types $c \in [c^1, \bar{c}]$. Such allocations may or may not be constrained efficient—i.e., they may or may not maximize profits subject to incentive compatibility and to the constraints $u(c) \geq u^1(c)$ $\forall c \in [\hat{c}, c^1)$ and $u(c) \geq u^0(c; c^1)$ $\forall c \in [c^1, \bar{c}]$. If they are constrained efficient for all $\hat{c}$ in an interval below c^1, then we can readily verify that the MWS reservation utilities are $\bar{V}(c) = u^1(c)$ in that interval.[15]

Finding the maximal interval of break-even types thus boils down to decreasing $\hat{c}$ until the resulting allocation *ceases* to be constrained efficient. Werning (2007) derives a simple test for the constrained efficiency of income tax systems. The following lemma adapts this test to the insurance context of our model.[16]

Lemma 3. *Let*

$$g(c) \equiv \frac{d}{dc}\left[\frac{(v_q(q^1(c), c) - c) f(c)}{v_{cq}(q^1(c), c)}\right] + f(c), \tag{33}$$

[15]The argument is: if we "guess" that $\bar{V}(c) = u^1(c)$, then the fact that the allocations are constrained efficient and yield zero profits together imply that they meet the criteria of Definition 1.

[16]Under some additional assumptions, Werning (2007) also derives a (weaker) test based on local "Laffer" effects: does a local reduction in marginal tax rate at one point in the type distribution increase revenue? If so, the extra revenue can be used to make other types better off and the tax schedule is constrained inefficient. The test we develop here is a qualitative obverse. It amounts to testing whether *raising* the tax on the lowest cost type and using the extra revenue to make all of the other types better off relaxes the incentive constraints by enough to make that lowest cost type better off in spite of the higher tax.

and $c^2 \equiv \sup\left\{c \in [\underline{c}, c^1) | g(c) < 0\right\}$ *(and* $c^2 = \underline{c}$ *if* $g(c) \geq 0$ *for all* $c \in [\underline{c}, c^1)$*). Then,*

$$\bar{V}(\hat{c}) = u^1(\hat{c})$$

is the MWS-reservation utility associated with the interval of types $\hat{c} \in [c^2, c^1]$*.*

Proof. See Appendix A. □

4.2.3 Step 3: Pareto-improving transfers

As we lower $\hat{c}$ to $c^2 - \varepsilon$, we (by construction) fail the constrained Pareto optimality test if we assign the break-even allocation from Step 2 to types $[c^2 - \varepsilon, c^2)$. Since the allocation for types $c \in [c^2, \bar{c}]$ is constrained Pareto efficient, this failure can only happen insofar as there is scope for Pareto-improving transfers from types $c \in [c^2 - \varepsilon, c^2)$ to the types above. As these Pareto-improving transfers are implemented, minimum utility constraints become slack for some interval containing c^2. Interval (c^C, c^*) in Figure 2 illustrates such an interval. Computing this interval is reasonably straightforward, since, as in step 1, slack minimum utility constraints (together with the fact that the interval must break even) allow one to use the first order conditions to characterize the allocation.

As we further lower $\hat{c}$ in step 3, three things can happen. First, the upper bound of the interval of slack minimum utility constraints (c^* in Figure 2) can continue to increase until it hits c^1. Second, the upper bound of the interval of slack minimum utility constraints could continue to increase, without reaching c^1, until the lower bound hits $\underline{c}$. Third, the upper bound of the interval of slack minimum utility constraints could start to *decrease*.

Let c^3 be the critical $\hat{c}$ at which one of these three things occurs. Then, for each $\hat{c} \in [c^3, c^2]$, $\bar{V}(\hat{c})$ will be the utility of the $\hat{c}$ type's allocation in the allocation computed for the interval associated with the $\hat{c}$ type. In the first case, the break-even interval disappears, and all types become cross-subsidized, and we can revert to a (slightly modified) version of step 1 to proceed to lower $\hat{c}$'s. In the second case, we are done. The third case is analogous to the end of step 1: cross-subsidies would start to shrink, indicating that at c^3 we should return to step 2 and begin to construct a new break-even interval.

4.3 Equilibrium allocation

The above procedure iterates over groups of types with binding and non-binding minimum utility constraints until it hits the lowest cost type $\underline{c}$. It does not only determine reservation utilities $\bar{V}(c)$. It also determines the equilibrium allocation, which is the solution to the lowest cost type's program. This solution features exactly the different sub-groups of types we have identified before, namely intervals with and without binding minimum utility constraints. As the construction above determines the allocation for each type in these intervals, it also determines the final equilibrium allocation. The next section provides an illustration.

Remark 1 (Cross-subsidization.)**.** *Note that the preceding construction also provides a criterion for the existence of cross-subsidies in the MWS equilibrium. Subsidies from low cost to high cost types are reflected by positive transfers* T^0*. If* $c^1 \equiv \bar{c}$*, and* $g(c) \geq 0$ *for all types* $c \in (\underline{c}, \bar{c})$ *(so that* T_0 *is locally non-increasing at* $\bar{c}$ *and the separating and individually break-even allocations are all constrained efficient), then there are no cross-subsidies in the MWS equilibrium. In this case, it coincides with Riley (1979)'s and Azevedo and Gottlieb (2017)'s equilibrium.*

Remark 2 (Numerical stability)**.** *The continuous construction described in the algorithm above is significantly faster than numerically recursively solving Spence's discrete problems. For example, with uniformly distributed types we find that the continuous construction finishes within less than a second while the time to solve Spence's discrete problems can easily take an hour or more.*

We have also found the continuous construction to be significantly more stable, numerically. This is intuitive: each type's (sub-problem) program recursively depends on the solution of the previous program. Thus, numerical errors from one recursion step propagate and multiply through to the final allocation. For example, with uniformly distributed types we find that the continuous construction quickly converges. E.g., a discretized type space with only 40 types is sufficient such that allocations do not change by more than 0.25% when refining the discretization. Instead, refining the type space beyond 40 types increases *the change in allocations when using Spence's discrete problems: it does* not *converge. In this case, an increase in numerical errors due to additional sub-problems and constraints to be numerically solved dominates the benefit of a more precise type space.*

Our results suggest that a numerical implementation of Spence's discrete problems is *neither sufficiently fast nor numerically stable to be used in (empirical) applications.*[17] *As we show in the following application, our continuous construction can be readily employed in applied contexts.*

5 Application

We illustrate our approach by considering a particularly salient question in the context of insurance and selection markets, namely: how effective are policy interventions? The quantitative welfare effects of policy interventions in markets with asymmetric information have received much attention in recent years (e.g., Finkelstein et al. (2009), Einav et al. (2010b), Hackmann et al. (2015), Azevedo and Gottlieb (2017)).

Below, we study the effect of a minimum coverage mandate, which is the canonical solution to adverse selection in insurance markets (Akerlof (1970)).[18] Our analysis closely follows Azevedo and Gottlieb (2017)'s approach in that we compute the effect of a marginal increase in the mandate. We contrast the effect of mandates in MWS equilibrium to that in Azevedo and Gottlieb (2017)/Riley (1979) (AG) equilibrium. Our results uncover substantial differences in the welfare effects of mandates—in particular much larger welfare effects in the AG framework, as measured by consumer surplus. Intuitively, the larger effects in the AG framework stem from the fact that the AG equilibrium is not constrained efficient, and the mandate moves the equilibrium "closer" to the constrained efficiency frontier.[19]

[17] We provide more detailed results on numerical stability and efficiency in Online Appendix C.5.

[18] Other possible policy interventions that address adverse selection frictions include employer-provided pooled insurance (e.g., Finkelstein (2002)), risk adjustment (e.g., Handel et al. (2015)), and government-provided basic insurance (e.g., Boone (2015)). For an overview, we refer to Geruso and Layton (2017).

[19] This result is similar to Dahlby (1981), who shows that mandating partial pooled-price insurance coverage—and allowing for optional supplemental insurance—can lead to Pareto-improvements in the Rothschild and Stiglitz (1976) model. The reason is that compulsory pooled-price insurance implements the informationally feasible Pareto-improving cross-subsidies that are ruled out in Rothschild and Stiglitz (1976). In contrast, mandates in an MWS world cannot yield Pareto improvements.

5.1 Comparative equilibrium effects of mandates

We consider a minimum-coverage mandate $M > 0$ which requires that all consumers purchase a contract with $q \geq M$. We compute the effects of this mandate by comparing the equilibria with it and without it; we do so using both the MWS equilibrium concept and the AG equilibrium concept. Computing the AG equilibrium (with or without a mandate) is straightforward.[20] So is adapting the construction of the MWS equilibrium from the preceding section to allow for a mandate.[21]

We use a calibration inspired by Azevedo and Gottlieb (2017, Section 5) and Einav et al. (2013). Consumers have utility $V(q, p; c) = (qc - \frac{\gamma}{2}q(q-2)\sigma^2) - p$, with $\gamma = 10^{-5}$ and $\sigma = 25,000$. Types' cost c is normally distributed with mean $4,340$ and standard deviation 300, truncated to $c \in [3340, 5340]$. Notice that in our calibrations each type's individually optimal (first-best) insurance coverage is $q = 1$. Figure 3 (a) contrasts expected utility in MWS and AG equilibrium allocations. The MWS equilibrium yields significantly higher welfare than the AG equilibrium: it Pareto dominates the AG equilibrium, and the MWS equilibrium yields roughly 20% higher (sum across types) consumer surplus.

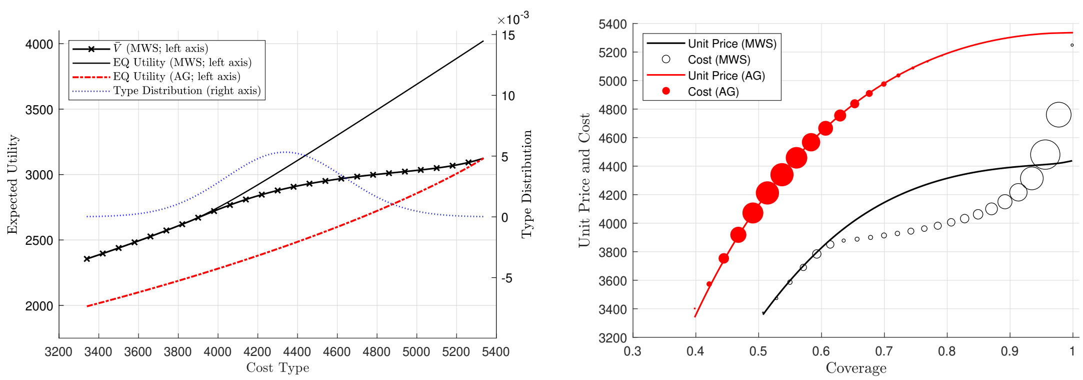
One can "see" the steps of the MWS equilibrium construction in Figure 3 (a). For the costliest type, equilibrium utility exceeds her reservation utility $\bar{V}(\bar{c})$ (which coincides with the AG equilibrium utility, i.e., her utility with her first-best break-even contract). Thus, she is cross-subsidized by lower cost types. Moving down cost types, MWS equilibrium and reservation utility converge, indicating that cross-subsidies become smaller. Roughly at cost type $c = 4,000$, $\bar{V}$ coincides with the MWS equilibrium utility, which signals the end of cross-subsidies and the beginning of a break-even interval; this is where the algorithm in Section 4 transitions from step 1 to step 2. The lower cost types' equilibrium utility and reservation utility coincide, implying that their contracts individually break even. So, in this example, step 3 is never needed.

[20] In AG equilibrium, each contract breaks even. It is computed analogously to a break-even interval in step 2 of the MWS equilibrium construction. Details are available on request.

[21] If $M \geq q^0(c^1)$, M is reached in step 1 of the equilibrium construction (with the notation from Section 4). In this case, there exists a type c' such that all types $c \in [\underline{c}, c']$ receive coverage M, $q^0(c) = M$. For these types, transfers T^0 increase when adding lower cost types in step 1. Thus, the construction will finish in step 1. If M is reached in step 2 instead, step 2 immediately reverts to step 3, since there is necessarily cross-subsidization among the types that buy M. Then, step 3 implements cross-subsidies from low cost types buying M to higher cost types. Details are available on request.

Figure 3 (b) depicts, for a given coverage level, the unit price and the average cost of types that buy the contract in equilibrium. The illustration mirrors the insights from Figure 3 (a). Unit prices and costs coincide for the AG equilibrium since each contract individually breaks even. This is similar for the break-even interval $q \leq 0.61$ in MWS equilibrium. In contrast, the unit prices for mid-sized coverage $q \in (0.61, 0.95)$ exceed the cost of supplying them. Firms make profit on these contracts, which enables them to cross-subsidize higher cost types, resulting in lower unit prices for $q \in (0.95, 1]$ than their costs.

Figure 3: Equilibrium allocations without a mandate.
Figure (a) illustrates equilibrium utility, reservation utility $\bar{V}$, and distribution of consumer types. Figure (b) depicts the equilibrium unit price (p^*/q^*; solid lines) and average cost of consumers that buy a contract with given coverage level in equilibrium ($\mathbb{E}[c \mid q^*(c) = q]$; circles). The size of the circles represents the mass of consumers buying a contract. The MWS equilibrium involves net cross subsidies from intermediate cost types to the highest cost types (black circles in Figure (b) lie below the cost curve for intermediate coverage levels and above for high coverage levels) but no cross subsidies from the lowest cost types (black circles lie on the cost line in Figure (b) and MWS equilibrium utilities coincide with reservation utilities in Figure (a)). Cross subsidies are Pareto-improving (the MWS equilibrium utilities rightmost circles lie everywhere above the AG utilities in Figure (a)). Prices and average costs (definitionally) coincide in the AG equilibrium.



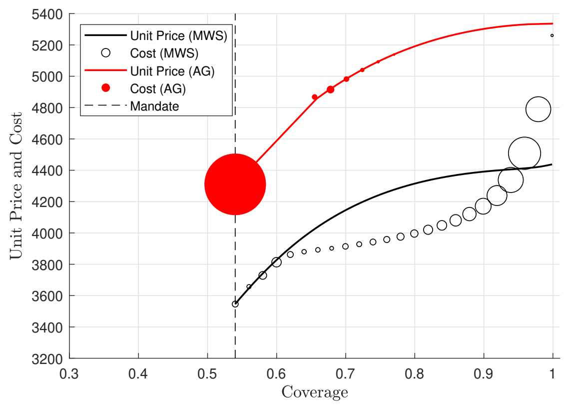
(a) Equilibrium utility.

(b) Unit prices and costs.

Figure 4 illustrates the effects of imposing a mandate $q \geq M \equiv 0.54$ in the AG and MWS equilibrium. In AG equilibrium, the mandate forces many low-cost types to pool at coverage M. This causes M to be cheap, inducing many (indeed, most) of the high-cost types to switch from their original AG contract to the minimum contract M. (In the language of Azevedo and Gottlieb (2017, p. 89), the "knock-on" effects are very strong.)

In contrast, the effect of a mandate is almost negligible in MWS equilibrium.

Figure 4: Equilibrium allocations and costs with an $M = 54\%$ coverage mandate.
The figure depicts the equilibrium per-unit price (p^*/q^*; solid lines) and average costs of consumers purchasing each coverage level in equilibrium (circles, the sizes of which reflect the mass of buyers).



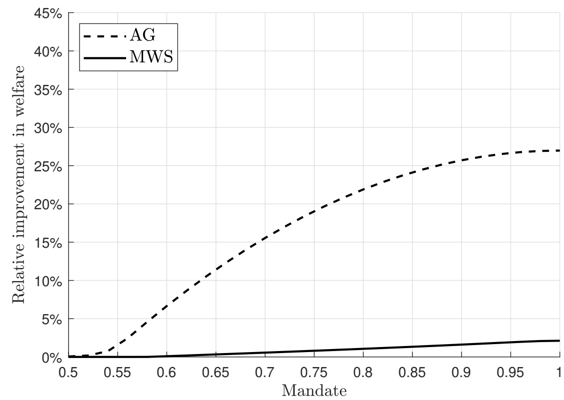
Since, in the absence of a mandate, consumers receive more coverage in MWS than in AG equilibrium due to cross-subsidies, there are less incentives for consumers to change their choices upon the implementation of a mandate in MWS compared to AG equilibrium. Moreover, MWS equilibrium "smooths" a mandate's distortionary effect on higher cost types by implementing cross-subsidies. As a result, the mandate in Figure 4 has a negligible effect in MWS equilibrium, while it has a strictly positive effect on AG equilibrium, increasing (total) consumer surplus by 0.75%. The effect differential stems from the inability of the AG equilibrium to implement Pareto-improving cross-subsidies. The mandate then partially implements such subsidies, reducing the deviation of the AG allocation from the (constrained) efficiency frontier.

Figure 5 illustrates the effect on total consumer surpluses of different mandate levels. In line with the previously developed intuition, we find that a mandate leads to a larger increase in total consumer surplus in AG than in MWS equilibrium across all mandates. The difference between the mandate's effects is large: a mandate increases consumer surplus by over ten times more in the AG equilibrium than in the MWS equilibrium.[22]

[22] In results not reported here, we also computed that the difference in a mandate's welfare effect (between AG and MWS equilibrium) is increasing with the ratio of the mass of low-cost to high-cost types. The intuition is that there are more cross-subsidies in MWS equilibrium the less expensive it is to implement them, i.e., the larger the ratio of low-cost to high-cost types. Then, the MWS equilibrium allocation "differs more" from the AG allocation in the absence of mandate—which boosts the difference in the effect of mandates.

Figure 5: Welfare effects of mandates.

The figure plots the change in consumer surplus as a function of the minimum coverage mandate, relative to the consumer surplus with unrestricted insurance coverage for the AG and MWS equilibria.



In our example, a mandate of 100% coverage maximizes consumer surplus, since it achieves a first-best allocation. Thus, the effect of a 100% mandate reflects the distance between an original (no-mandate) equilibrium allocation and the first-best equilibrium. In MWS equilibrium, a 100% mandate increases consumer surplus by 2.1% relative to the original MWS allocation. The MWS equilibrium is thus very close to the "first-best" informationally unconstrained frontier. Since the MWS allocation is second-best efficient, this result suggests that the cross-subsidies implemented in the MWS equilibrium resolve "most" of the adverse selection-driven pathologies in the market.

In AG equilibrium, a 100% mandate increases consumer surplus by 27% relative to the original AG allocation. Thus, the original AG equilibrium is far from being first-best. Intuitively, this welfare loss can be decomposed into two sources: information frictions arising from private information; and a failure of the equilibrium to achieve a constrained efficient outcome. Combining this result with the MWS result suggests (intuitively rather than formally) that only 2.1 percentage points of this distance is due to inherent information frictions. The rest arises because of the failure of the AG market dynamics to reach the constrained efficiency frontier.

5.2 Discussion of the AG and MWS equilibrium concepts

As discussed in Hendren (2014), Rothschild and Stiglitz (1976)'s equilibrium-non-existence result stems from a tension between two conceptually distinct competitive forces. On the one hand, competition should drive profits of individual contracts to zero (otherwise there is scope for cream-skimming). On the other hand, constrained inefficient outcomes can be exploited by the free entry of new firms offering an array of products that simultaneously and profitably attract the entire market. Insofar as constrained efficiency *requires* cross-subsidization across contracts sold to different types of buyers, these two distinct competitive forces are fundamentally unreconcilable.

The equilibrium concepts of Azevedo and Gottlieb (2017) and Riley (1979) describe market dynamics which resolve this tension in favor of the first competitive force, yielding market equilibria which have individually break-even contracts but are often constrained inefficient. Instead, the MWS equilibrium concept resolves the tension in favor of the second competitive force, yielding market equilibria which are constrained efficient but (may) involve cross subsidies across contracts.[23]

Concepts in both veins have been widely employed for studying competitive markets with adverse selection in markets with small, finite type spaces (including Hoy (1982), Crocker and Snow (1985), Puelz and Snow (1994), Crocker and Snow (2008), Finkelstein et al. (2009), and Mimra and Wambach (2019b) for the MWS concept, and Besanko and Thakor (1987), Landers et al. (1996), Newhouse (1996), Inderst (2005), Handel et al. (2015), Boyer and Peter (2018), and Mimra and Wambach (2019b) for the AG/Riley concept). But adjudicating which concept is appropriate for empirical applications, and in what circumstances, calls for tractable models with richer type spaces. Our paper is the first to provide such a model for the MWS concept.

The application in the preceding subsection shows that the MWS and AG con-

[23] As observed by Farinha Luz (2017), resolving the tension in favor of individually break-even contracts means that equilibrium outcomes are locally insensitive to the distribution of types. Farinha Luz extends the two-type Rothschild and Stiglitz (1976) model to allow firms to make *stochastic* contract offers, and shows that the resulting equilibrium exists, depends continuously on the underlying type distribution, and involves cross-subsidies across types. MWS equilibria—including our characterization of it with continuous types—do not involve stochastic offers, but, as in Farinha Luz (2017), yield equilibrium outcomes that are continuous in the type distribution and involve cross-subsidies.

cepts have radically different predictions about allocations. They are, therefore, empirically distinguishable. These differences also have potential policy implications. For example, insofar as one finds real-world allocations are better described by the AG equilibrium allocations in our application, then the resulting allocations lie far below the constrained efficient frontier. In this case, there is, at least in principle, scope for public policy interventions that are not just welfare improving in aggregate (e.g., in certainty equivalent terms), but which are *Pareto* improving.

6 Conclusions

We show how to extend the MWS equilibrium concept to models with a one-dimensional continuum of cost types, and we prove the existence of such an equilibrium under weak conditions. The underlying argument and conclusions readily extend to mixed-distribution models with a continuum of costs *and* a finite number of mass points.

We also describe a quasi-recursive method for characterizing and computing the MWS equilibrium with continuous types. This method relies on ordinary differential equation techniques and a test for Pareto optimality drawn from the optimal tax literature, and it is efficient enough to be of potential use for in empirical applications. Among these empirical applications is the ability to compare the constrained efficient MWS equilibrium concept with the individual-contract-break-even solution concept developed in Azevedo and Gottlieb (2017) in empirically realistic settings.

Even if the AG equilibrium concept turns out to be more empirically realistic in a broad range of settings, our algorithm for computing the MWS equilibrium is still valuable, because it provides a useful benchmark for analyzing the welfare implications of market interventions. Specifically, because the MWS equilibrium is constrained efficient, the welfare effects of a given policy intervention in an MWS world are exclusively a result of movements along the second-best Pareto frontier towards allocations which are closer to the first-best frontier. As such, comparing a (notional) MWS world to an (actual) AG world provides a useful decomposition of the welfare effects in the latter into the components due (1) to fundamentally un-solvable asymmetric-information-driven market pathologies and (2) to failures of the competitive equilibrium to address in-principle solvable problems.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500.

Azevedo, E. M. and Gottlieb, D. (2017). Perfect competition in markets with adverse selection. *Econometrica*, 85(1):67–105.

Besanko, D. and Thakor, A. V. (1987). Competitive equilibrium in the credit market under asymmetric information. *Journal of Economic Theory*, 42(1):167–182.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley, New York.

Boone, J. (2015). Basic versus supplementary health insurance: Moral hazard and adverse selection. *Journal of Public Economics*, 128:50–58.

Boyer, M. M. and Peter, R. (2018). Insurance fraud in a Rothschild-Stiglitz world. *Journal of Risk and Insurance*, forthcoming.

Crocker, K. J. and Snow, A. (1985). The efficiency of competitive equilibria in insurance markets with asymmetric information. *Journal of Public Economics*, 26(2):207–219.

Crocker, K. J. and Snow, A. (2008). Background risk and the performance of insurance markets under adverse selection. *Geneva Risk and Insurance Review*, 33(2):137–160.

Dahlby, B. (1981). Adverse selection and pareto improvements through compulsory insurance. *Public Choice*, 37(3):547–558.

Einav, L., Finkelstein, A., and Cullen, M. R. (2010a). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics*, 125(3):877–921.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, 103(1):178–219.

Einav, L., Finkelstein, A., and Schrimpf, P. (2010b). Optimal mandates and the welfare cost of asymmetric information: Evidence from the U.K. annuity market. *Econometrica*, 78(3):1031–1092.

Farinha Luz, V. (2017). Characterization and uniqueness of equilibrium in competitive insurance. *Theoretical Economics*, 12(3):1349–1391.

Finkelstein, A. (2002). The effect of tax subsidies to employer-provided supplementary health insurance: evidence from canada. *Journal of Public Economics*, 84:305–339.

Finkelstein, A., Poterba, J., and Rothschild, C. (2009). Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities. *Journal of Financial Economics*, 91(1):38–58.

Geruso, M. and Layton, T. J. (2017). Selection in health insurance markets and its policy remedies. *Journal of Economic Perspectives*, 31(4):23–50.

Hackmann, M. B., Kolstad, J. T., and Kowalski, A. E. (2015). Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, 105(3):1030–1066.

Handel, B., Hendel, I., and Whinston, M. D. (2015). Equilibria in health exchanges: Adverse selection versus reclassification risk. *Econometrica*, 83(4):1261–1313.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–2682.

Handel, B. R., Kolstad, J. T., and Spinnewijn, J. (2019). Information frictions and adverse selection: Policy interventions in health insurance markets. *Review of Economics and Statistics*, 101(2):326–340.

Hellwig, M. (2007). The undesirability of randomized income taxation under decreasing risk aversion. *Journal of Public Economics*, 91(3-4):791–816.

Helly, E. (1921). Über lineare Funktionaloperationen. *Sitzungsberichte der Naturwissenschaftlichen Klasse der Kaiserlichen Akademie der Wissenschaften, Wien*, 121:265–295.

Hendren, N. (2014). Unravelling vs unravelling: A memo on competitive equilibriums and trade in insurance markets. *Geneva Risk and Insurance Review*, 39(2):176–183.

Hoy, M. (1982). Categorizing risks in the insurance industry. *Quarterly Journal of Economics*, 97(2):321–336.

Inderst, R. (2005). Matching markets with adverse selection. *Journal of Economic Theory*, 121(2):145–166.

Landers, R. M., Rebitzer, J. B., and Taylor, L. J. (1996). Rat race redux: Adverse selection in the determination of work hours in law firms. *American Economic Review*, 86(3):329–348.

Levy, Y. J. and Veiga, A. (2019). Competitive insurance markets with unbounded cost. *Working Paper*.

Mimra, W. and Wambach, A. (2014). New developments in the theory of adverse selection in competitive insurance. *Geneva Risk and Insurance Review*, 39:136–152.

Mimra, W. and Wambach, A. (2019a). Contract withdrawals and equilibrium in competitive markets with adverse selection. *Economic Theory*, 67(4):875–907.

Mimra, W. and Wambach, A. (2019b). Endogenous insolvency in the Rothschild-Stiglitz model. *Journal of Risk and Insurance*, 86(1):165–181.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208.

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

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18:301–317.

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

Newhouse, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, 34(3):1236–1263.

Picard, P. (2014). Participating insurance contracts and the Rothschild-Stiglitz equilibrium puzzle. *Geneva Risk and Insurance Review*, 39(2):153–175.

Picard, P. (2019). Equilibrium in insurance markets with adverse selection when insurers pay policy dividends. *Journal of Risk and Insurance*, 86(4):887–914.

Puelz, R. and Snow, A. (1994). Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. *Journal of Political Economy*, 102(2):236–257.

Riley, J. G. (1979). Informational equilibrium. *Econometrica*, 47(2):331–359.

Rothschild, C. and Scheuer, F. (2013). Redistributive taxation in the roy model. *Quarterly Journal of Economics*, 128(2):623–668.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90(4):629–649.

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

Werning, I. (2007). Pareto efficient income taxation. *Working Paper*.

Weyl, E. G. and Veiga, A. (2017). Pricing institutions and the welfare cost of adverse selection. *American Economic Journal: Microeconomics*, 9(2):139–148.

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

A Appendix

A.1 Proof of Lemma 2

Proof. By assumption, $V_c < \xi < 0$. Incentive compatibility therefore implies that $\bar{V}^*(c)$ is decreasing in c and indeed, that $\bar{V}^*(c) - \bar{V}^*(c') > \xi(c' - c)$ for all $c' > c$.

Let $\bar{p}_c = \max\{p_c \geq 0 : V(q_c^*, p_c; c) \geq V(0, 0; c)\}$ be type c's maximum willingness to pay for her optimal actuarially fair contract $q_c^* \equiv \arg\max_q[V(q, qc; c)]$. Define $\bar{\pi} = \sup\{\bar{p}_c | c \in \mathcal{C}\}$ as the upper bound for maximum profits a firm can make on any type. Fix any $c^* < c$ and any $\delta > 0$. Use Assumption 1 to find an $\varepsilon > 0$ such that

$$\hat{V}^n(c, \bar{f}2\varepsilon\bar{\pi}) - \hat{V}^n(c, 0) < \delta. \tag{34}$$

Choose $\varepsilon < \delta$.

Now fix any $c \in (\underline{c}, c^*)$, any $c^+ \in \mathcal{C} \cap (c, c+\varepsilon)$ and $c^- \in \mathcal{C} \cap (c - \varepsilon, c)$, and an N such that $c^+, c^- \in C^N$. We will show that

$$\bar{V}^n(c^-) - \bar{V}^n(c^+) < (2k+1)\delta \quad \forall n \geq N. \tag{35}$$

The lemma will follow immediately.

Towards showing (35), first note that

$$\bar{V}^n(c^-) - \bar{V}^n(c^+) = \left[\bar{V}^n(c^-) - V(\vec{A}^n(c^+; c^-); c^+)\right] + \left[V(\vec{A}^n(c^+; c^-); c^+) - \bar{V}^n(c^+)\right] \tag{36}$$

$$\leq \left[\bar{V}^n(c^-) - V(\vec{A}^n(c^-; c^-); c^+)\right] + \left[\hat{V}^n(c^+, \bar{f}2\varepsilon\bar{\pi}) - \bar{V}^n(c^+)\right] \tag{37}$$

$$\leq 2\varepsilon k + \delta \leq (2k+1)\delta. \tag{38}$$

The third line follows from $\bar{V}^n(c^-) = V^n(\vec{A}^n(c^-; c^-); c^-)$, the uniform bound k on V_c, (34), $\hat{V}(c, 0) = \bar{V}(c)$, and the choice of ε.

The second line follows from two observations. First: $V(\vec{A}^n(c^+; c^-); c^+) \geq V(\vec{A}^n(c^-; c^-); c^+)$ by incentive compatibility for the c^+ types in the c^- type's MWS problem. Second, in the solution to the c^- type's MWS problem, there will be some negative profits $-T \leq 0$ on types in the range $[c^+, \bar{c}]$. Given T and the allocation for types in $[c^-, c^+)$, the allocation of the types in $[c^+, \bar{c}]$ must maximize the utility of the c^+ type subject to incentive compatibility for all types in $[c^-, \bar{c}]$, the minimum utility constraints for types $[c^+, \bar{c}]$, and the resource constraint relaxed by T. $V(\vec{A}^n(c^+; c^-); c^+)$ is the result of this maximization. Because of the (extra) incentive constraints associated with the lower types $[c^-, c^+)$, this maximization problem is *tighter* than the problem defining $\hat{V}^n(c^+, T)$. Hence, $V(\vec{A}^n(c^+; c^-); c^+) \leq \hat{V}^n(c^+, T)$. T is equal to the total profits on types in $[c^-, c^+)$, the mass of which is less than $\bar{f}2\varepsilon$ and the maximum profits from each of which is bounded by $\bar{\pi}$. Since $\hat{V}^n(c^+, T)$ is weakly increasing in T, it follows that $\hat{V}^n(c^+, T) \leq \hat{V}^n(c^+, \bar{f}2\varepsilon\bar{\pi})$. $\square$

A.2 Proof of Lemma 3

Proof. We need only to show that constrained efficiency of the step-2-like allocations (i.e., those featuring an interval of break even types below c^1) is equivalent to $g(c) \geq 0$. Following Werning (2007), take *any* incentive compatible, break-even allocation $(q^*(c), u^*(c))$ (where $u^*(c) \equiv v(q^*(c); c) - p^*(c)$). To test if this is constrained Pareto efficient, consider the problem of maximizing profits, subject to incentive compatibility and the "utility constraints" $0 = u(c) - u^*(c)$:

$$\max_{\{(q(c),u(c))\}_{c\in[\underline{c},\bar{c}]}} \int_{\underline{c}}^{\bar{c}} (v(q(c); c) - u(c) - q(c)c)\, f(c)dc \tag{39}$$

$$\textit{subject to } 0 = v_c(q(c); c) - u'(c) \qquad \forall c \in [\underline{c}, \bar{c}]; \quad [\text{multiplier } \eta(c)] \tag{40}$$

$$\textit{and } 0 = u(c) - u^*(c) \qquad \forall c \in [\underline{c}, \bar{c}]; \quad [\text{multiplier } \psi(c)]. \tag{41}$$

The Lagrangian, after integrating the incentive constraint by parts, is:

$$\begin{aligned}\mathcal{L} = &\int_{\underline{c}}^{\bar{c}} (v(q(c); c) - u(c) - cq(c))\, f(c)dc + \int_{\underline{c}}^{\bar{c}} (u(c) - u^*(c))\, \psi(c)dc \\ &- \int_{\underline{c}}^{\bar{c}} \left(\eta'(c)u(c) - \eta(c)v_c(q(c); c)\right) dc + [\eta(\underline{c})u(\bar{c}) - \eta(\underline{c})u(\hat{c})]\,.\end{aligned} \tag{42}$$

As in Werning (2007), the original allocation is constrained Pareto optimal precisely when the Lagrange multipliers $\psi(c)$ on the utility constraints are all nonnegative. The first order conditions for this problem are, with respect to $u(c)$,

$$\eta'(c) = \psi(c) - f(c) \tag{43}$$

and, with respect to $q(c)$,

$$(v_q(q(c); c) - c)f(c) - \eta(c)v_{cq}(q(c), c) = 0.$$

Solving the latter for η and differentiating yields

$$\eta'(c) = \frac{d}{dc}\frac{(v_q(q^*(c), c) - c)f(c)}{v_{cq}(q^*(c), c)} = \psi(c) - f(c).$$

Comparing with (43) shows that $g(c) = \psi(c)$. Hence, the constrained efficiency is equivalent to $g(c) \geq 0$. $\square$

– ONLINE APPENDIX –

B Equicontinuity in the canonical setting

Rothschild and Stiglitz (1976)'s insurance market model is a special case of the model considered in the main text. In this model, $c \in (0, 1)$ is interpreted as the probability of experiencing a loss of size $L > 0$ out of a fixed wealth W for a von Neumann-Morgenstern expected utility maximizer with a strictly concave utility function $u(\cdot)$. Without loss of generality, we normalize L to 1, so that the coverage level q can be interpreted as the gross indemnity (conditional on a loss), and the expected cost of selling such a contract to type c is qc. Upon buying a contract (q, p), type c's utility is $u_L = u(W - 1 + q - p)$ in the loss state, and it is $u_{NL} = u(W - p)$ in the no-loss state.

In this section, we prove that Assumption 1 from the main text is satisfied in this setting: the utility benefit of a small transfer in MWS equilibrium is uniformly bounded across MWS sub-problems and sufficiently fine discretization of the type space. The proof follows two steps. First, we prove an auxiliary lemma, stating that in a c-type's sub-problem either type c is bound away from full insurance[24] uniformly across fine discretizations or there are cross-subsidies from low to high risk types. Second, we leverage this result in computing the welfare effect of a small transfer in MWS equilibrium: if type c is uniformly bound away from full insurance, one can use this transfer to benefit only type c in an incentive compatible way; otherwise, the transfer can be used to increase utility of each type. In both cases, we can uniformly bound the transfer's impact on the c type's utility, which by the envelope theorem and concavity of the MWS program (in this canonical case) implies equicontinuity of the functions $\{\hat{V}^n(c, T)\}_{c \in [\underline{c}, c^*], n > N'}$.

Theorem B.1. *Assumption 1 is satisfied in the Rothschild and Stiglitz (1976) model.*

[24] i.e., c faces a deductible.

B.1 Auxiliary Lemma

Before proving the theorem, we set up some notation and establish an auxiliary lemma. Recall (from the main text) that $\vec{A}^n(c;\hat{c}) = (q^n(c;\hat{c}), p^n(c;\hat{c}))$ is the allocation that the c type receives in the MWS equilibrium sub-problem for the $\hat{c}$ type in the n^{th} discretization. Define $D^n(c;\hat{c}) = 1 - q^n(c;\hat{c})$, which is the insurance contract's deductible.

Lemma B.1. *For any $c^* < \bar{c}$, there exists a $\bar{D} > 0$ and an N such that for all $c < c^*$ and all $n \geq N$* either

1. $D^n(c;c) > \bar{D}$ or
2. $\sum_{c' \in [\hat{c},\bar{c}] \cap \mathcal{C}^n} f^n(c') \left(p^n(c';c) - c' q^n(c';c)\right) < 0 \quad \forall \hat{c} > c.$

In other words, for large enough n, either the type $c < c^*$ faces a "large" deductable in her own MWS sub-problem *or* the allocation for the c type's MWS sub-problem involves cross-subsidies for all subgroups $[\hat{c},\bar{c}]$ for every $\hat{c} > c$.

For any particular n and c, the fact that either property 1 or 2 holds is straightforward: if $D^n(c;c) = 0$, then all types get the same pooled fair allocation and property 2 clearly holds. The content of the proof is to construct a single $\bar{D}$ for which either 1 or 2 holds for all c and all sufficiently large n.

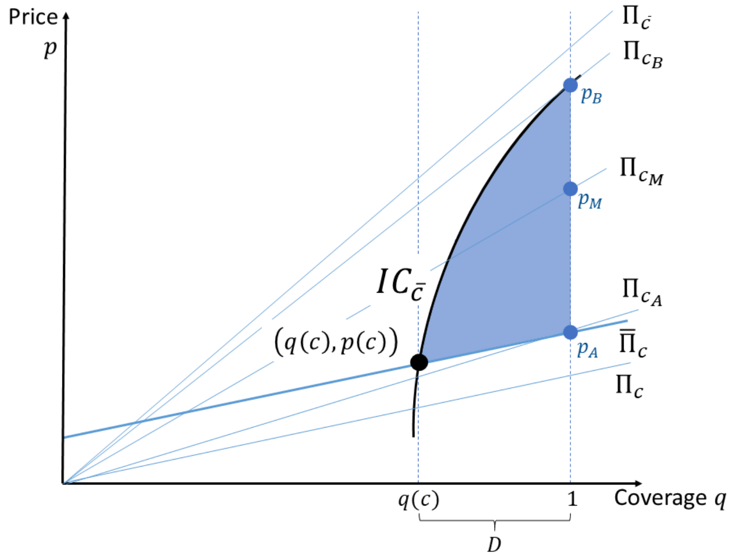
Proof. Fix any $c^* < \bar{c}$, $c \leq c^*$, $c \in \mathcal{C}$, and any n. Consider any allocation $\{(q(c'), p(c'))\}_{c' \in [c,\bar{c}] \cap \mathcal{C}^n}$ with $q(c') \leq 1$ for all c' (as will be true in any MWS sub-allocation since $q^* = 1$ is the individually optimal (first-best) insurance coverage in the RS model) and let $D = 1 - q(c)$. By incentive compatibility (and single crossing), $q(c') \geq q(c)$ for all $c' \geq c$, and $(q(c'), p(c'))$ must lie below the $\bar{c}$ type's indifference curve through $(q(c), p(c))$, labeled $IC_{\bar{c}}$ in Figure B.1.

Similarly, $(q(c'), p(c'))$ must lie above the c type's indifference curve through $(q(c), p(c))$ which, to the right of $(q(c), p(c))$ and for $q(c') < 1$, lies above the iso-profit line $\bar{\Pi}_c(q') = p(c) + c(q' - q(c))$.[25] All types' allocations must therefore

[25] The reason is that, in the RS model, a type c's marginal willingness-to-pay (and thus her indifference curve's slope) is steeper than her marginal cost c if $q(c) < 1$:

$$\frac{dp}{dq} = c \frac{u'_L}{c u'_L + (1-c) u'_{NL}} > c \;\Leftrightarrow\; u'_L > u'_{NL} \;\Leftrightarrow\; q < 1.$$

Figure B.1: Constructing D for Lemma B.1.



lie in the shaded area in Figure B.1. Denote by p_A and p_B the right-hand "corner prices" of this area. Formally: p_B is the price for full insurance that makes type $\bar{c}$ indifferent to $(q(c), p(c))$, satisfying

$$u(W - p_B) = (1 - \bar{c})u(W - p(c)) + \bar{c}u(W - p(c) - D),$$

and p_A is the price for full insurance on the c-type iso-cost line through $(q(c), p(c))$, i.e., $p_A = p(c) + cD$.

Given any type $c' \in [c, \bar{c}]$, the least profitable contract in the area is $(1, p_A)$ and the most profitable is $(1, p_B)$. Since the MWS sub-problem has exactly zero profits overall, this implies that p_B is above and p_A is below the joint pooling price

$$p_B \geq c_M^n(c) \geq p_A$$

where

$$c_M^n(c) \equiv \mathbb{E}_{F^n}[c'|c' \in [c, \bar{c}] \cap C^n]$$

is the expected cost in c type's sub-problem. Thus, $c_M^n(c)$ is the cost of a pooling contract for types $[\underline{c}, \bar{c}] \cap C^n$. It follows directly from the definition of p_A that

$$cD \leq c_M^n(c) - p(c). \tag{B.1}$$

Similarly, from the definition of p_B:

$$u(W - c_M^n(c)) \geq (1 - \bar{c})u(W - p(c)) + \bar{c}u(W - p(c) - D) \geq u(W - p(c) - D)$$

and hence

$$c_M^n(c) \leq p(c) + D. \tag{B.2}$$

The preceding formalizes the simple observation that if the deductible D is small then $p(c)$ must be close to the fair pooling price $c_M^n(c)$. *At* full insurance, it is obvious that all subgroups $[\hat{c}, \bar{c}]$ receive cross-subsidies from the lower risk types $[\underline{c}, \hat{c})$, for any $\hat{c} \in (\underline{c}, \bar{c})$. We will now show that the same is true for sufficiently large N and for sufficiently small D.

To that end, define:

$$c_M(c) \equiv \lim_{n\to\infty} c_M^n(c),$$

$$c_H^n(c) \equiv \mathbb{E}[c'|c' \in [c_M(c), \bar{c}], F^n], \quad \text{and} \quad c_H(c) \equiv \lim_{n\to\infty} c_H^n(c)$$

$$c_L^n(c) \equiv \mathbb{E}[c'|c' \in [c, c_H(c)], F^n], \quad \text{and} \quad c_L(c) \equiv \lim_{n\to\infty} c_L^n(c),$$

and

$$Z \equiv \min\left\{\min\{c_H(c') - c_M(c'), c_M(c') - c_L(c')\} | c' \in [\underline{c}, c^*]\right\} > 0,$$

where $Z > 0$ follows from $c^* < \bar{c}$.

The fact that the CDF F^n converges uniformly to the continuous distribution F implies that $c_i^n(c)$ converges uniformly for each $i \in \{L, M, H\}$. Hence, one can choose N such that, for all c, and $i \in \{L, M, H\}$, $n > N$ implies $|c_i^n(c) - c_i(c)| < Z/3$. For such n, then, $c_H^n(c) - c_M^n(c) > Z/3$ and $c_M^n(c) - c_L^n(c) > Z/3$.

For $n > N$ and any c, the total profits accruing to the set of types below any given $\hat{c} \in C^n$ in the MWS-sub-problem for type c are:

$$
\begin{aligned}
\Pi^*(\hat{c};c) = & \sum_{c' \in [c, \min\{\hat{c}, c_A^n\}) \cap C^n} f^n(c') \Pi(\vec{A}^n(c',c);c') \\
& + \sum_{c' \in [c_A^n, \min\{\hat{c}, c_B^n\}] \cap C^n} f^n(c') \Pi(\vec{A}^n(c',c);c') \\
& + \sum_{c' \in (c_B^n, \hat{c}] \cap C^n} f^n(c') \Pi(\vec{A}^n(c',c);c')
\end{aligned} \tag{B.3}
$$

where $c_A^n \equiv p_A$ and $c_B^n \equiv p_B$, and p_A and p_B are as in Figure B.1 (and where we use the convention that the sum over an "interval" of the form $[x,y]$ with $y < x$ is zero). By definition, total profits in the c sub-problem are zero, $\Pi^*(\bar{c};c) = 0$. Thus, showing that $\Pi^*(\hat{c};c) > 0$ for all $\hat{c} > c$ implies that profits of $(\hat{c}, \bar{c}]$ are negative which is equivalent to statement 2 in the lemma and will thus complete the proof. To that end, note first that all types $c' \in [c, c_A^n)$ have $\Pi(\vec{A}^n(c',c);c') > 0$ since all contracts in the shaded area with corners at $(1, p_A)$, $(1, p_B)$, $(q(c), p(c))$ are above the zero-profit lines for types $c' < c_A^n$. So $\Pi^*(\hat{c};c) > 0$ for all $\hat{c} \leq c_A^n$. Similarly, all types $c' \in (c_B^n, \bar{c}]$ have $\Pi(\vec{A}^n(c',c);c') < 0$, since contracts in the shaded area are below c' types' zero-profit line. Hence, profits for the sub-group $(c_B^n, \bar{c}]$ are negative and that for $[c, c_B^n]$ positive. Thus, $\Pi^*(\hat{c};c) > 0$ for all $\hat{c} \geq c_B^n$. It remains to establish $\Pi^*(\hat{c};c) > 0$ for $\hat{c} \in [c_A^n, c_B^n]$.

Observe that for each type,

$$
\Pi(\vec{A}^n(c',c);c') \geq p_A - c' = p(c) + cD - c' \geq c_M^n - c' - (1-c)D, \tag{B.4}
$$

where we use the bounds on p_A and $c_M^n(c)$ derived above (and the fact that $(1, p_A)$ is the least profitable contract in the shaded area of Figure B.1 for all types).

For $\hat{c} \in [c_A^n, c_B^n]$, taking $D < Z/3$, and taking an $f^n(c')$-weighted sum of expression (B.4) we have (using Equation (B.1), the definition of c_B^n and p_B, and the bound $u(W - p_B) \geq u(W - p(c) - D)$ to show $c_B^n \leq p(c) + D \leq c_M^n + (1-c)D <$

$c_M^n + Z/3 \le c_H$):

$$\begin{aligned}
\Pi^*(\hat{c};c) &\geq (F^n(\hat{c}) - F^n(c)) \left(c_M^n(c) - \mathbb{E}_{F^n}[c'|c' \in [c,\hat{c}]] - (1-c)D\right) \\
&> (F^n(\hat{c}) - F^n(c)) \left[(c_M - Z/3) - \mathbb{E}_{F^n}[c'|c' \in [c,c_B^n]] - Z/3\right] \\
&\geq (F^n(\hat{c}) - F^n(c)) \left[c_M - \mathbb{E}_{F^n}[c'|c' \in [c,c_H]] - 2Z/3\right] \\
&= (F^n(\hat{c}) - F^n(c)) \left[c_M - c_L^n - 2Z/3\right] \\
&\geq (F^n(\hat{c}) - F^n(c)) \left[c_M - c_L - Z\right] \geq 0
\end{aligned} \tag{B.5}$$

which completes the proof. □

B.2 Proof of Theorem B.1

In the following we prove that for any $c^* < c$ there exists an N' such that for all $K > 0$ there exists $\delta > 0$ such $|\hat{V}^n(c,T) - \hat{V}^n(c,0)| < \varepsilon$ for all $T < \delta$ and all $c \in [\underline{c}, c^*]$ and $n \geq N'$.

Recall that for $\hat{c} \in C^n$ and any $T \geq 0$, we define $\hat{V}(\hat{c},T)$ as

$$\hat{V}^n(\hat{c},T) \equiv \max_{\{\vec{A}^n(c;\hat{c})\}_{c\in[\hat{c},\bar{c}]\cap C^n}} V(\vec{A}^n(c;\hat{c});c) \tag{B.6}$$

subject to

$$V(\vec{A}^n(c;\hat{c});c) \geq V(\vec{A}^n(c';\hat{c});c) \quad \forall c,c' \geq \hat{c} \quad \text{with} \quad c,c' \in C^n \text{ and} \tag{B.7}$$

$$V(\vec{A}^n(c;\hat{c});c) \geq \bar{V}^n(c) \quad \forall c \geq \hat{c} \quad \text{with} \quad c \in C^n \text{ and} \tag{B.8}$$

$$\sum_{c\in[\hat{c},\bar{c}]\cap C^n} \Pi(\vec{A}^n(c;\hat{c});c) f^n(c) \geq -T. \tag{B.9}$$

By definition, $\bar{V}^n(c) = \hat{V}^n(c,0)$. For any $c_1, c_2 \in C^n$ with $c_1 \leq c_2$ define

$$T^n(c_2;c_1) = \sum_{c\in C^n\cap[c_1,c_2)} f^n(c)\Pi(\vec{A}^n(c;c_1);c) \tag{B.10}$$

as the profit of types $[c_1,c_2)$ in the original solution $\{\vec{A}^n(c;c_1)\}_{c\in C^n\cap[c_1,\bar{c}]}$ to the c_1 sub-problem. Since the sum of profits over all types $[c_1,\bar{c}] \cap C^n$ is zero in the c_1 type's MWS sub-problem for the nth discretization, $T^n(c_2;c_1)$ is the cross-subsidy to the group $[c_2,\bar{c}]$ in the solution to that sub-problem.

It is equivalent to formulate allocations in utility space (u_L, u_{NL}) instead of in coverage-price space (q, p) due to the binary wealth structure in the RS model. As a consequence, the program defining $\hat{V}^n(p,T)$ re-formulated in utility space has a linear objective function, linear incentive compatibility and minimum utility constraints, and a concave budget constraint. It follows that $\hat{V}^n(p,T)$ is concave in T and hence that

$$\hat{V}^n(c,T) - \hat{V}^n(c,0) \leq T\frac{\partial \hat{V}^n(c,0)}{\partial T}. \tag{B.11}$$

The following Lemma allows us to bound $\frac{\partial \hat{V}^n(c,0)}{\partial T}$ by some K uniformly in c and n (for $n > N$ for some sufficiently large N):

Lemma B.2. *For any $c^* < \bar{c}$ there exists an N and a K such that $\frac{\partial \hat{V}^n(c,0)}{\partial T} \leq K$ for all c in$[\underline{c}, c^*] \cap C^n$ and $n > N$.*

Proof. Choose N and D as in Lemma B.1 and consider any $c \leq c^*$ and any $n > N$. If case 2. of that Lemma holds for this c and n, then none of the minimum utility constraints bind in the MWS sub-problem for type c in the nth discretization. By the envelope theorem, we can compute the welfare effects of a small increase in T via a uniform marginal increase $\Delta > 0$ in utility across all types and both states, so:

$$\begin{aligned}\frac{\partial \hat{V}^n(c,0)}{\partial T} &= \frac{\Delta}{\sum_{c' \in C^n, c' \geq c} f^n(c')\Delta\left[\frac{1-c'}{u'(u^{-1}(u^n_{NL}(c',c)))} + \frac{c'}{u'(u^{-1}(u^n_L(c',c)))}\right]} & \text{(B.12)}\\ &\leq \frac{u'(W-1)}{1-F^n(c)} \leq \frac{u'(W-1)}{1-F^n(c^*)} \leq \frac{u'(W-1)}{1-F(c^*)} \equiv K_1, & \text{(B.13)}\end{aligned}$$

where the denominator of (B.12) is the total resource cost of marginally increasing everyone's utility by Δ in both states, and $u^n_L(c', c)$ and $u^n_{NL}(c', c)$ are the c' type's utility in the loss and no loss state in the c type's sub-problem, respectively.

If case 1. of Lemma B.1 holds, on the other hand, we can compute the welfare consequences of a small increase in T by using that transfer to slide the c type down and to the left along the $c^{n+} \equiv c + \frac{1}{2^n}$ type's (the next lowest type's) indifference curve. A straightforward computation of the welfare consequences of this marginal increase yields:

$$\frac{\partial \hat{V}^n(c,0)}{\partial T} = \frac{(1-c)/(1-c^{n+}) - c/c^{n+}}{f^n(c)\left(\frac{c}{u'(u^{-1}(u_L^n(c,c)))}(-1/c^{n+}) + \frac{1-c}{u'(u^{-1}(u_{NL}^n(c,c)))}/(1-c^{n+})\right)} \tag{B.14}$$

$$\leq \frac{1}{\underline{f}(1-c)c\left(\frac{1}{u'(u^{-1}(u_{NL}^n(c)))} - \frac{1}{u'(u^{-1}(u_L^n(c)))}\right)} \tag{B.15}$$

$$\leq \frac{1}{\underline{f}\min_{c'\in[\underline{c},\bar{c}]}(1-c')c'}\frac{1}{\min_{h\in[W-1,W]}\left[\frac{1}{u'(h)} - \frac{1}{u'(h-D)}\right]} \tag{B.16}$$

$$\equiv K_2 \tag{B.17}$$

Taking $K = \max\{K_1, K_2\}$ completes the proof. □

By Lemma B.2, for sufficiently large N, $n > N$ we find that:

$$\hat{V}^n(c,T) - \hat{V}^n(c,0) \leq TK, \tag{B.18}$$

which proves Theorem A.1.

C Equilibrium Construction

C.1 Setup

- A continuous distribution of types c with density $f(c)$ on support $[\underline{c},\bar{c}]$.
- Quasilinear preferences $V(q,p;c) = v(q;c) - p$ satisfying

$$\frac{\partial v}{\partial q} = v_q > 0, \frac{\partial^2 v}{\partial q^2} = v_{qq} < 0, \text{ and } \frac{\partial^2 v}{\partial q \partial c} = v_{qc} \geq 1.$$

- An MWS equilibrium is an allocation $\{q(c),p(c)\}_{c\in[\underline{c},\bar{c}]}$ solving

$$0 = \max_{\{(q(c),p(c))\}_{c\in[\underline{c},\bar{c}]}} \int_{\underline{c}}^{\bar{c}} (p(c) - q(c)c) f(c)dc \tag{C.1}$$

$$\text{subject to } v(q(c);c) - p(c) \geq v(q(c');c) - p(c') \quad \forall c,c' \in [\underline{c},\bar{c}], \tag{C.2}$$

$$\text{and } v(q(c);c) - p(c) \geq \bar{V}(c) \quad \forall c \in [\underline{c},\bar{c}], \tag{C.3}$$

where $\bar{V}$ is a function with the property that for all $\hat{c} \in [\underline{c}, \bar{c}]$ the value of the program

$$\max_{\{(q(c),p(c))\}_{c\in[\hat{c},\bar{c}]}} \int_{\hat{c}}^{\bar{c}} (p(c) - q(c)c)\, f(c)dc \tag{C.4}$$

$$\text{subject to } \nu(q(c);c) - p(c) \geq \nu(q(c');c) - p(c') \quad \forall c, c' \in [\hat{c}, \bar{c}] \tag{C.5}$$

$$\text{and } \nu(q(c);c) - p(c) \geq \bar{V}(c) \quad \forall c \in [\hat{c}, \bar{c}] \tag{C.6}$$

is 0.

In this setup, the following holds:

1. The assumptions on preferences imply that the solution $q^*(c)$ to the problem $\max_q[\nu(q;c) - qc]$ is unique and weakly increasing in c. They also imply that indifference curves in (q, p) space are concave and satisfy the single crossing property.

2. Instead of working with allocations $(q(c), p(c))$, it is equivalent to work with allocations $u(c) = \nu(q(c);c) - p(c)$ and $q(c)$ or $p(c)$.

3. It will turn out to be more convenient to work with $(q(c), u(c))$ allocations. In the next section, we take a "first order approach", assuming that insurance coverage is strictly increasing with cost type, $q(c') > q(c)$ for all $c' > c$. Then (due to single crossing), incentive compatibility is equivalent to the local incentive constraint

$$u'(c) = \nu_c(q(c);c). \tag{C.7}$$

If $q(c)$ is not strictly increasing, bunching needs to be considered, as we do in Section C.3.

C.2 Construction

C.2.1 Step 1

Fix $\hat{c} \in [\underline{c}, \bar{c}]$. Consider the solution to the relaxed, primal (sub-)problem for types $[\hat{c}, \bar{c}]$:

$$\max_{\{(q(c),u(c))\}_{c\in[\hat{c},\bar{c}]}} u(\hat{c}) \tag{C.8}$$

subject to

$$u'(c) = \nu_c(q(c);c) \qquad \forall c \in [\hat{c}, \bar{c}]; \quad [\text{multiplier } \lambda\eta(c)] \tag{C.9}$$

and

$$\int_{\hat{c}}^{\bar{c}} (\nu(q(c);c) - u(c) - q(c)c)\, f(c)dc \geq K \qquad [\text{multiplier } \lambda]. \tag{C.10}$$

We call this Program $A_K(\hat{c})$

The Lagrangian for $A_K(\hat{c})$, after integrating by parts the incentive constraint, is:

$$\begin{aligned}\mathcal{L} = u(\hat{c}) + \lambda \int_{\hat{c}}^{\bar{c}} \left(\nu(q(c);c) - u(c) - cq(c) - \hat{K}\right) f(c)dc \\ - \lambda \int_{\hat{c}}^{\bar{c}} \left(\eta'(c)u(c) - \eta(c)\nu_c(q(c);c)\right) dc + \lambda \left[\eta(\bar{c})u(\bar{c}) - \eta(\hat{c})u(\hat{c})\right],\end{aligned} \tag{C.11}$$

with $\hat{K} = K\left(\int_{\hat{c}}^{\bar{c}} f(c)dc\right)^{-1}$. The first order conditions for this Lagrangian with respect to $u(c)$ and $q(c)$ can be used to fully characterize the solution. From the former:

$$\eta'(c) = -f(c), \tag{C.12}$$

and, from the latter, for interior c:

$$\left(\nu_q(q(c);c) - c\right) f(c) = \eta(c) v_{qc}(q(c);c). \tag{C.13}$$

Since the solution will have no distortion at the top ($\nu_q(q(\bar{c});\bar{c}) - \bar{c} = 0$ at $q(\bar{c}) =$

$q^*(\bar{c})$), the solution must, by continuity, have $\eta(\bar{c}) = 0$, and hence, from (C.12),

$$\eta(c) = 1 - F(c). \tag{C.14}$$

We can then write (C.13) as:

$$\frac{v_q(q(c);c) - c}{v_{qc}(q(c);c)} = \frac{1 - F(c)}{f(c)}. \tag{C.15}$$

Let $q^0(c)$ be the solution to (C.15). Note that this solution is *independent* of $\hat{c}$ and K.

Denote by $q(c;\hat{c})$ the solution of Program $A_0(\hat{c})$ for type $c \geq \hat{c}$ in type $\hat{c}$'s sub-problem. It follows that the allocation solving $A_0(\hat{c})$ for any given $\hat{c} \in [\underline{c}, \bar{c})$ has:

$$q(c;\hat{c}) = q^0(c) \quad \text{and} \quad u(c;\hat{c}) = u^0(c) + T^0(\hat{c}), \tag{C.16}$$

where, by (C.9),

$$u^0(\bar{c}) \equiv v(q^0(\bar{c});\bar{c}) - \bar{c}q^0(\bar{c}), \tag{C.17}$$

$$u^0(c) \equiv u^0(\bar{c}) - \int_c^{\bar{c}} v_c(q^0(c');c')dc', \tag{C.18}$$

and

$$T^0(\hat{c}) \equiv \frac{1}{1 - F(\hat{c})} \int_{\hat{c}}^{\bar{c}} \left[v(q^0(\tilde{c});\tilde{c}) - u^0(\tilde{c}) - \tilde{c}q^0(\tilde{c}) \right] f(\tilde{c})d\tilde{c}. \tag{C.19}$$

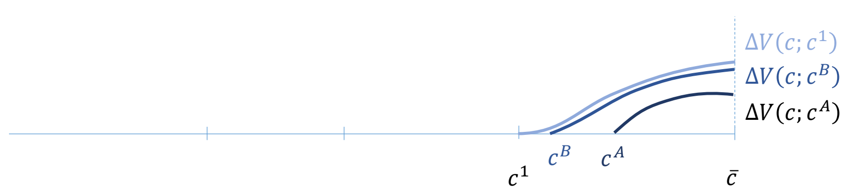
The intuition for the preceding is as follows. First, we know that the solution must have $q(c;\hat{c}) \equiv q^0(c)$ for any K and $\hat{c}$. If we knew $u(\bar{c};\hat{c})$, we could then integrate the (local) incentive constraint to find $u(c)$ for any c. We find $u(\bar{c};\hat{c})$ in two steps. First, we normalize $u(\bar{c};\hat{c})$ to optimal and fair insurance, as in (C.17). Second, we find the associated $u(c;\hat{c})$ allocation for all cs (by integrating the local incentive constraints (C.18)) and compute the resulting per-person *surplus* $T^0(\hat{c})$ (per (C.19)) associated with the resulting allocation. Third, we balance the resource constraint by reducing prices (thereby increasing utility) uniformly by $T^0(\hat{c})$, holding $q(c;\hat{c})$ fixed. Since this uniform transfer maintains incentive compatibility and balances the budget, the result is the solution.

The solution to $A_0(\hat{c})$ will be the solution to the MWS sub-problem for type $\hat{c}$ if (1) $q^0(c)$ is weakly increasing and (2) the minimum utility constraints (C.6) are satisfied (since the MWS sub-problem program differs only from Program $A_0(\hat{c})$ in that the latter drops those constraints). If $T^0(\hat{c})$ is decreasing in the neighborhood around $\bar{c}$, then we can, per the following lemma, easily verify that the minimum utility constraint $\bar{V}(\hat{c}) = u^0(\hat{c}) + T^0(\hat{c})$ "works" for $\hat{c}$s in this neighborhood. Intuitively: as $\hat{c}$ is lowered within this range, the per-person transfer $T^0(\hat{c})$ and hence $u(c;\hat{c})$ increase—so the minimum utility constraints are indeed slack for all $c > \hat{c}$ types. Figure C.1 illustrates this increasing "slack" in the minimum utility constraint when reducing $\hat{c}$. This leads us to the following lemma:

Lemma C.1. *Let* $c^1 \equiv \sup\left\{c \in [\underline{c}, \bar{c}] | \frac{dT^0(c)}{dc} > 0\right\}$. *Then for all* $\hat{c} \in [c^1, \bar{c}]$, $\bar{V}(\hat{c}) = u^0(\hat{c}) + T^0(\hat{c})$ *are MWS reservation utilities on* $[c^1, \bar{c}]$.

Proof. The result is trivial if $c^1 = \bar{c}$. If $c^1 < \bar{c}$, observe that, because $q^0(c)$ is independent of $\hat{c}$ and T^0 is non-increasing on $[c^1, \bar{c}]$, we have, for $\hat{c} \in [c^1, \bar{c}]$ and $c' > \hat{c}$, $u(c';\hat{c}) \geq u(c';c') \equiv u^0(c') + T^0(c') = \bar{V}(c')$. Hence, with $\bar{V}(\hat{c})$ so defined, the solution to Program $A_0(\hat{c})$ satisfies the minimum utility constraints (and yields 0 profits). It is thus a solution to the MWS (sub)-problem for type $\hat{c}$. $\square$

Figure C.1: Step 1: Illustration.
The figure illustrates the slack in minimum utility constraints during lowering $\hat{c}$ from c^A to c^B and c^1 in step 1 of the construction.

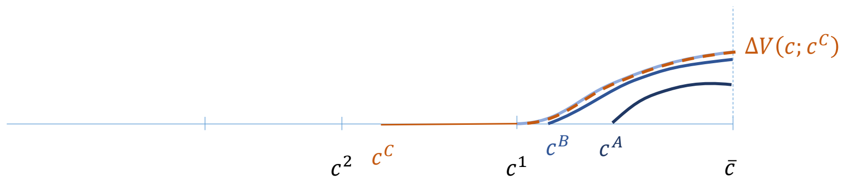


C.2.2 Step 2

By construction, $T^0(\hat{c})$ is increasing in $\hat{c}$ to the left of c^1, so the solution to Program $A_0(c^1 - \varepsilon)$ *violates* the minimum utility constraints for type $c^1 - \varepsilon$ for sufficiently small ε. The minimum utility constraints must therefore bind in the solution to the MWS sub-problem for types $c^1 - \varepsilon$ (again, for sufficiently small ε), as Figures C.2

illustrates. That is, there will be an interval of break-even types to the left of c^1 (and c^1 breaks even in the solution of Program $A_0(c^1)$).

Figure C.2: Step 2 – Intermediate stage: Illustration.
The figure illustrates the slack in minimum utility constraints during lowering $\hat{c}$ in step 2 of the construction.



For these break-even types, it is $u(c) = v(q(c);c) - cq(c)$, and hence

$$u'(c) = v_q(q(c);c)q'(c) + v_c(q(c);c) - q(c) - cq'(c), \tag{C.20}$$

or, using incentive compatibility:

$$q'(c) = \frac{q}{v_q(q(c);c) - c}. \tag{C.21}$$

This differential equation can be solved (uniquely) for $q^1(c)$ given the initial condition $q^1(c^1) = q^0(c^1)$. The local incentive constraint can then be integrated to find $u^1(c)$ using the initial condition $u^1(c^1) = \bar{V}(c^1)$.[26]

To find the the lowest break-even type, define

$$g(c) \equiv \frac{d}{dc}\left[\frac{(v_q(q^1(c);c) - c)f(c)}{v_{cq}(q^1(c);c)}\right] + f(c), \tag{C.22}$$

and let $c^2 \equiv \sup\left\{c \in [\underline{c}, c^1) | g(c) < 0\right\}$. Then, $\bar{V}$ for types $\hat{c} \in [c^2, c^1)$ may be found by evaluating these types' utility at $(q^1(\hat{c}), u^1(\hat{c}))$, as the next lemma shows and Figure C.3 illustrates.

Lemma C.2. *For types $\hat{c} \in [c^2, c^1)$, the solution to the $\hat{c}$-subproblem is*

$$q(c;\hat{c}) = \begin{cases} q^0(c), & \text{if } c \in [c^1, \bar{c}], \\ q^1(c), & \text{if } c \in [c^2, c^1) \end{cases} \quad \text{and} \quad p(c;\hat{c}) = \begin{cases} u^0(c) + T^0(c^1), & \text{if } c \in [c^1, \bar{c}], \\ u^1(c), & \text{if } c \in [c^2, c^1) \end{cases}$$

[26]The solution satisfies $u'(c) = v_c(q(c);c)$ or, equivalently, $p'(c) = v_q(q(c);c)q'(c)$.

and the MWS reservation utility is $\bar{V}(\hat{c}) = u^1(\hat{c})$ for $\hat{c} \in [c^2, c^1)$.

Proof. The allocation $\{(\hat{q}(c), \hat{u}(c))\}_{c\in[\hat{c},\bar{c}]}$ with $(\hat{q}(c), \hat{u}(c)) = (q^1(c), u^1(c))$ for $c \in [\hat{c}, c^1]$ and $(\hat{q}(c), \hat{u}(c)) = (q^0(c), u^0(c) + T^0(c^1))$ for $c \in [c^1, \bar{c}]$ is incentive compatible, satisfies the minimum utility constraints with the $\bar{V}$ defined in the claim, and yields exactly zero profits. Moreover, as we will show shortly, it is (constrained) Pareto optimal as long as $g(c) \geq 0$. Since $g(\hat{c}) \geq 0$ for all $\hat{c} \in [c_2, c_1]$, it follows that for each such $\hat{c}$ this allocation solves the MWS problem given this $\bar{V}$.

To see that it is constrained Pareto optimal when $g(c) \geq 0$, we follow Werning (2007). Take any incentive compatible, break-even allocation $(q^*(c), u^*(c))$. To test if this is constrained Pareto efficient, consider the Program X of maximizing profits, subject to incentive compatibility and the "utility constraints" $0 = u(c) - u^*(c)$:

$$\max_{\{(q(c),u(c))\}_{c\in[\underline{c},\bar{c}]}} \int_{\underline{c}}^{\bar{c}} \left(v(q(c);c) - u(c) - q(c)c\right) f(c)dc \tag{C.23}$$

$$\textit{subject to } 0 = v_c(q(c);c) - u'(c) \qquad \forall c \in [\underline{c}, \bar{c}]; \quad [\text{multiplier } \eta(c)] \tag{C.24}$$

$$\textit{and } 0 = u(c) - u^*(c) \qquad \forall c \in [\underline{c}, \bar{c}]; \quad [\text{multiplier } \psi(c)]. \tag{C.25}$$

The Lagrangian, after integrating the incentive constraint by parts, is:

$$\begin{aligned}\mathcal{L} = &\int_{\underline{c}}^{\bar{c}} \left(v(q(c);c) - u(c) - cq(c)\right) f(c)dc + \int_{\underline{c}}^{\bar{c}} \left(u(c) - u^*(c)\right) \psi(c)dc \\ &- \int_{\underline{c}}^{\bar{c}} \left(\eta'(c)u(c) - \eta(c)v_c(q(c);c)\right) dc + \left[\eta(\underline{c})u(\bar{c}) - \eta(\underline{c})u(\hat{c})\right].\end{aligned} \tag{C.26}$$

As in Werning (2007), the original allocation is constrained Pareto optimal precisely when the Lagrange multipliers $\psi(c)$ on the utility constraints are all non-negative. The first order conditions for this problem are, with respect to $u(c)$,

$$\eta'(c) = \psi(c) - f(c)$$

and, with respect to $q(c)$,

$$(v_q(q(c);c) - c)f(c) - \eta(c)v_{cq}(q(c),c) = 0.$$

Solving the latter for η, differentiating, substituting into the former, and evaluating at the allocation proposed in this lemma yields

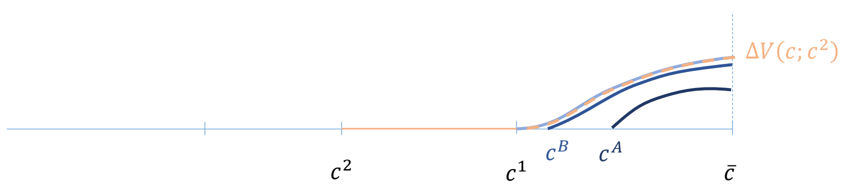
$$\eta'(c) = \frac{d}{dc}\frac{(v_q(q^*(c),c) - c)f(c)}{v_{cq}(q^*(c),c)} = \psi(c) - f(c)$$

and, thus,

$$g(c) = \psi.$$

Hence, the proposed allocation is constrained Pareto optimal precisely when $g(c) \geq 0$. $\square$

Figure C.3: End of Step 2: Illustration.
The figure illustrates the slack in minimum utility constraints at the end of step 2 of the construction.



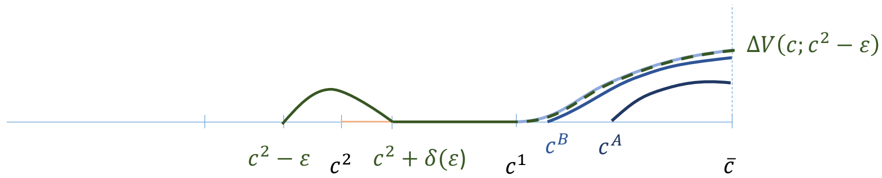
C.2.3 Step 3

As we lower c to $c^2 - \varepsilon$, we (by construction) fail the constrained Pareto optimality test if assigning the break-even allocation from Step 2 to types $[c^2 - \varepsilon, c^2)$. Since the allocation for types $c \in [c^2, \bar{c}]$ is constrained Pareto optimal, this failure can only happen insofar as there is scope for Pareto-improving transfers from types $c \in [c^2 - \varepsilon, c^2)$ to the types above. As these Pareto-improving transfers are implemented, minimum utility constraints become slack for some interval $[c^2 - \varepsilon, c^2 + \delta]$ (with the same MWS-sub-problem allocation obtaining above $c^2 + \delta$). Figure C.4 provides an illustration.

In order to find $\bar{V}(c^2 - \varepsilon)$, we need to characterize the relationship between δ and ε and to find a condition that determines the start of the next lower-cost-type "break-even interval" c^3. The rationale is as follows: for any ε, use the FOCs (C.15) and the "initial conditions" at $c^2 + \delta$ to solve for the allocation for types $[c^2 - \varepsilon, c^2 + \delta]$ (given that minimum utility constraints are slack). The initial condition

Figure C.4: Step 3 – Intermediate Stage: Illustration.

Minimum utility constraints become slack as Pareto-improving transfers from types $[c^2-\varepsilon, c^2]$ to $[c^2, c^2+\delta(\varepsilon)]$ are implemented.



$q^2(c^2+\delta) = q^1(c^2+\delta)$ implies by (C.13) that

$$\eta(c^2+\delta) = \frac{\left(v_q(q^1(c^2+\delta); c^2+\delta) - (c^2+\delta)\right) f(c^2+\delta)}{v_{qc}(q^1(c^2+\delta); c^2+\delta)}. \tag{C.27}$$

Combining this with the FOC from (C.12) yields

$$\eta(c) = \eta(c^2+\delta) + \int_c^{c^2+\delta} f(c')dc'. \tag{C.28}$$

From there, we solve

$$\frac{v_q(q(c); c) - c}{v_{qc}(q(c), c)} = \frac{\eta(c)}{f(c)} \tag{C.29}$$

to characterize the optimal coverage $q^2(c)$ when minimum utility constraints do not bind. Finally, u^2 is given by the initial condition $u^2(c^2+\delta) = u^1(c^2+\delta)$ and integrating the local incentive compatibility constraint $u'(c) = v_c(q(c); c)$.

We find the upper end of the interval $c_\delta = c^2 + \delta(\varepsilon)$ by integrating profits "up" from $c^2+\varepsilon$ until the break-even point, which defines c_δ by

$$\int_{c^2-\varepsilon}^{c_\delta} v(q^2(c'); c') - u^2(c') - q^2(c')c'dc' = 0. \tag{C.30}$$

Defined as such, the minimum utility at $c_\varepsilon = c^2 - \varepsilon$ is

$$\bar{V}(c_\varepsilon) = u^2(c_\varepsilon). \tag{C.31}$$

To determine the end-point (i.e., lowest ε) of Pareto-improving transfers, define by $\varepsilon^* > 0$ the smallest point such that either (A) $\delta(\varepsilon^*) = c^1 - c^2$, *or* (B) type $c^2 - \varepsilon^*$

exactly breaks even.

In case (A), the whole break-even segment disappears. If there exists a larger break-even segment $(c'_\varepsilon, c'_\delta)$, test for Pareto-improving transfers from $(c^2 - \varepsilon^* - \varepsilon', c'_\varepsilon)$ to $(c'_\varepsilon, c'_\delta)$ for small ε', following step 3. If there does not exist a larger break-even segment, test for Pareto-improving subsidies up to the highest cost type by calculating total profits akin to step 1, and add lower cost types as long as these weakly increase total profit. Once there exist no Pareto-improving transfers, the lowest considered cost type breaks even and a new break-even segment begins, as Figure C.5 illustrates.

In case (B), a new break-even segment begins at c_{ε^*}, as in Figure C.6. Then, in both cases, refer to step 2 starting at c^1 redefined as $c^1 = c_{\varepsilon^*}$.

Figure C.5: Step 3 (Case (A)) and reiteration of step 2: Illustration.
A new break-even segment begins after implementing Pareto-improving transfers across $[c^3, c^2 + \delta^*]$ that partially "eat up" the previous break-even segment.

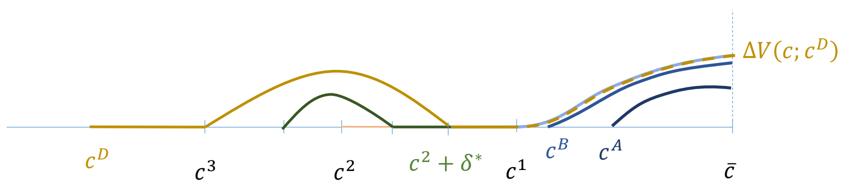
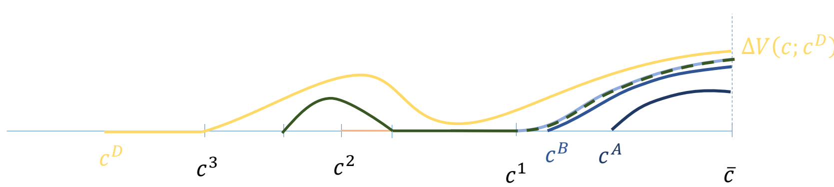


Figure C.6: Step 3 (Case (B)) and reiteration of Step 2: Illustration.
A new break-even segment begins after implementing Pareto-improving transfers that fully "eat up" the previous break-even segment.



Finally, note that the above procedure does not only determine the reservation utilities $\bar{V}(c)$ but also the equilibrium allocation, which is the allocation that determines $\bar{V}(\underline{c})$.

C.3 Bunching

The solution described in the previous section does not necessarily satisfy the monotonicity constraint $q(c) \geq q(c')$ for all $c \geq c'$. In particular, while $q'(c) \geq 0$ by construction in an break-even interval (see step 2), the construction in steps 1 and 3 may violate monotonicity. In this case, we need to add the term $\int \xi(c) q'(c) dc$ to the Lagrangian, where $\xi(c)$ is the multiplier on the monotonicity constraint. By integrating this term by parts, we yield

$$\xi(\bar{c}')q(\bar{c}') - \xi(\hat{c})q(\hat{c}) - \int \xi'(c) q(c) dc \tag{C.32}$$

for each interval $[\hat{c}, \bar{c}']$ of non-binding minimum utility constraints. At the upper end-point $\bar{c}'$ (which is either $\bar{c}$ or the lowest cost type in the next break-even group), monotonicity holds by construction, and thus $\xi(\bar{c}') = 0$. Adding (C.32) does not affect the FOC w.r.t. u, thus, η stays the same (i.e., as in Section C.2). However, it changes the FOC for interior q by ξ',

$$\left(\nu_q(q(c);c) - c\right) f(c) - \xi'(c) = \eta(c) v_{qc}(q(c), c) \tag{C.33}$$

if the minimum utility constraint does not bind for c. If, instead, it binds, it is $\xi'(c) = 0$.

Denote by $\tilde{q} : [\hat{c}, \bar{c}] \to [0, \infty)$ the solution constructed as in Section C.2. Assume that $\tilde{q}'(c) < 0$ in some interval $\mathcal{I} \subset [\hat{c}, \bar{c}]$. Following the rationale of Mussa and Rosen (1978), we will choose $q^r > 0$ and $c_A, c_B \in [\hat{c}, \bar{c}], c_A < c_B, \mathcal{I} \subset [c_A, c_B]$, and allocate the same coverage q^r to all types $c' \in [c_A, c_B]$. Suppose, first, that monotonicity does not bind at c_A and c_B. Since η is unchanged, a necessary condition is that $\tilde{q}(c') = q^r(c')$ at both $c' \in \{c_A, c_B\}$. Since $\xi(c') = 0$ for $c' \in \{c_A, c_B\}$, it also holds that

$$\int_{c_A}^{c_B} \xi'(c) dc = 0.$$

And we know ξ' given c_B: from (C.33) and the fact that $q(c) = \tilde{q}(c_B) \equiv q^r$ for $c \in [c_A, c_B]$, it is

$$\xi'(c) = \left(\nu_q(q^r, c) - c\right) f(c) - \eta(c) v_{qc}(q^r, c) \quad \forall c \in [c_A, c_B]. \tag{C.34}$$

These conditions are sufficient to determine c_A, c_B, and q^r.

Suppose, second, that the monotonicity binds at c_A. This is the case only if $\tilde{q}'(\hat{c}) < 0$, which requires $\hat{c} \equiv c_A$. Then, $\xi(c_A) \neq 0$ while $\xi(c_B) = 0$. Hence,

$$\xi(c_A) = \xi(c_B) - \int_{c_A}^{c_B} \xi'(c')dc' = -\int_{c_A}^{c_B} \xi'(c')dc'. \tag{C.35}$$

$\xi(c_A) = \xi(\hat{c})$ may be found by using the FOC with respect to $q(\hat{c})$, which is

$$\left(\nu_q(q^r, \hat{c}) - \hat{c}\right) f(\hat{c}) - \xi(\hat{c}) - \xi'(\hat{c}) = \eta(\hat{c}) v_{qc}(q^r, \hat{c}) \tag{C.36}$$

and thus

$$\xi(\hat{c}) = \left(\nu_q(q(\hat{c}), \hat{c}) - \hat{c}\right) f(\hat{c}) - \eta(\hat{c}) v_{qc}(q(\hat{c}), \hat{c}) - \xi'(\hat{c}) = -\int_{c_A}^{c_B} \xi'(c')dc', \tag{C.37}$$

which implies

$$\left(\nu_q(q(\hat{c}), \hat{c}) - \hat{c}\right) f(\hat{c}) - \eta(\hat{c}) v_{qc}(q(\hat{c}), \hat{c}) = \lim_{x \to 0} - \int_{c_A + x}^{c_B} \xi'(c')dc' = \int_{c_A}^{c_B} \xi'(c')dc'. \tag{C.38}$$

Together with (C.34), this determines c_B and q^r.

C.4 Numerical Implementation

C.4.1 Preliminaries

We start with a discretization of the type space $C^N = \{c_1, ..., c_N\}$, enumerated such that $c_i < c_{i+1}$, where the number of types N describes the degree of discretization. Assume that $c_{i+1} - c_i \equiv \Delta > 0$ for all i. The discretized distribution of types is described by masses $\hat{f}_i$, $i = 1, ..., N$.[27] Type c_i yields utility $V(q, p; c_i) = \nu(q; c_i) - p$ from contract coverage q at price p. We determine the allocation $(\vec{A}_i)_{i=1,...,N} = (q_i, p_i)_{i=1,...,N}$ by constructing reservation utilities $\bar{V}_i$. In this process, we will determine whether minimum utility constraints hold in equilibrium, which will enable us to simultaneously determine the equilibrium allocation.

[27] Note that $\hat{f}_i / \Delta$ is thus an approximation of $f(c_i)$.

C.4.2 Constructing $\bar{V}$ and solving for the equilibrium allocation

In the following, we provide a pseudo-code that implements the construction of $\bar{V}$ and equilibrium allocations based on Section C.2.

Step 1. Non-binding minimum utility constraints.
Initialize with the highest cost type's optimal contract at a fair price, $q_N^0 = q_N^*$ and $p_N^0 = q_N^* c_N$. For $i = 1, ..., N-1$, determine q_i^0 by numerically solving[28]

$$\frac{v_q(q_i; c_i) - c_i}{v_{qc}(q_i; c_i)} = \frac{1 - \sum_{j=1}^{i} \hat{f}_j}{\hat{f}_i \Delta^{-1}} \tag{C.39}$$

for q_i, and then use the solution to compute

$$p_i^0 = p_{i+1}^0 + v_q(q_{i+1}^0; c_{i+1})(q_i^0 - q_{i+1}^0) \tag{C.40}$$

and

$$T_i^0 = \sum_{j=i}^{N} \left(p_j^0 - c_j q_j^0\right) \frac{\hat{f}_j}{\sum_{k=i}^{N} \hat{f}_k}. \tag{C.41}$$

To find the cut-off value c^1, define[29]

$$i^1 = \max\left\{i = 2, ..., N : T_i^0 - T_{i-1}^0 > 0\right\}. \tag{C.42}$$

If $\min\{q_{j+1}^0 - q_j^0 : j = i, ..., N-1\} \geq 0$, the solution satisfies monotonicity and we can directly use it. Otherwise, use ironing techniques a la Mussa and Rosen (1978) for each set $\{i, ..., N\}$ as described in Section C.4.3. It is

$$\bar{V}_i = v(q_i^0; c_i) - p_i^0 + T_i^0 \quad \forall i = i^1, ..., N. \tag{C.43}$$

Record as candidates for the final equilibrium allocation

$$q_i^{fin} = q_i^0 \text{ and } p_i^{fin} = p_i^0 - T_{i^1}^0 \quad \forall i = i^1, ..., N. \tag{C.44}$$

If $i^1 = 1$, we have found reservation utilities $\bar{V}_{i=1,...,N}$ and the equilibrium alloca-

[28] To account for potentially multiple local optima, it is useful to specify the optimal coverage for the largest cost type, q_N^*, as an upper bound when solving (C.39).
[29] If $T_i^0 - T_{i-1}^0 \leq 0$ for all $i = 2, ..., N$, then let $i^1 = 1$.

tion $(q_i^{fin}, p_i^{fin})_{i=1,...,N}$; else, go to Step 2.

Step 2. Break-even types.
Define $q^1_{i^1} = q^0_{i^1}$ and $p^1_{i^1} = p^0_{i^1}$.[30] The coverage for break-even types is found by recursively solving[31]

$$q^1_j = q^1_{j+1} + (c_j - c_{j+1}) \frac{q^1_{j+1}}{\nu_q(q^1_{j+1}; c_{j+1}) - c_{j+1}} \quad \forall j = 1, ..., i^1 - 1, \tag{C.45}$$

and prices are[32]

$$p^1_j = p^1_{j+1} + (c_j - c_{j+1}) \frac{\nu_q(q^1_{j+1}; c_{j+1}) q^1_{j+1}}{\nu_q(q^1_{j+1}; c_{j+1}) - c_{j+1}} \quad \forall j = 1, ..., i^1 - 1. \tag{C.46}$$

The cut-off value may be found by first defining

$$\hat{g}_j = \hat{f}_j \Delta^{-1} \frac{\nu_q(q^1_j; c_j) - c_j}{v_{qc}(q^1_j; c_j)}, \quad \forall j = 1, ..., i^1 - 1, \tag{C.47}$$

then numerically differentiating it,

$$\hat{g}'_j = \frac{\hat{g}_j - \hat{g}_{j-1}}{c_j - c_{j-1}}, \quad \forall j = 2, ..., i^1 - 1, \tag{C.48}$$

and letting $\hat{g}'_1 = \hat{g}'_2$. Finally, defining the cut-off value by

$$i^2 = \max \left\{ i = 1, ..., i^1 - 1 : \hat{g}'_i + \hat{f}_i \Delta^{-1} \leq 0 \right\}, \tag{C.49}$$

it is

$$\bar{V}_i = \nu(q^1_i; c_i) - p^1_i \quad \forall i = i^2, ..., i^1 - 1. \tag{C.50}$$

[30] As the type space becomes more dense, it is $q^1_{i^1} c_{i^1} \approx p^0_{i^1}$ by construction.

[31] Alternatively, one may use a discrete variant by solving $V(q_i, q_i c_i; c_{i+1}) = V(q^1_{i+1}, q^1_{i+1} c_{i+1}; c_{i+1})$ for q_i. In some instances we tested, this provided a numerically more stable implementation.

[32] Alternative, one may directly specify break-even prices, $p^1_j = q^1_j c_j$.

Record as candidates for the final equilibrium allocation

$$q_i^{fin} = q_i^1 \text{ and } p_i^{fin} = p_i^1 \quad \forall i = i^2, ..., i^1 - 1. \tag{C.51}$$

If $i^2 = 1$, we have found reservation utilities $\bar{V}_{i=1,...,N}$ and the equilibrium allocation $(q_i^{fin}, p_i^{fin})_{i=1,...,N}$; else, go to Step 3.

Step 3. Pareto-improving transfers.
Let $\varepsilon = 0$, $i_\varepsilon = i^2$, and $K_\varepsilon > 0$. As long as $K_\varepsilon > 0$ repeat:

(1) Let $\varepsilon \leftarrow \varepsilon + 1$ and $i_\varepsilon \leftarrow i^2 - \varepsilon$.

(2) Let $K > 0$ and $\delta = 0$. As long as $\delta < i^1 - i^2$ and $K > 0$, repeat:

   (i) Let $\delta \leftarrow \delta + 1$ and define $i_\delta = i^2 + \delta$, $q_{i_\delta}^2 = q_{i_\delta}^1$, $p_{i_\delta}^2 = q_{i_\delta}^1 c_{i_\delta}$, and

$$\eta_{i_\delta} = \frac{(\nu_q(q_{i_\delta}^2; c_{i_\delta}) - c_{i_\delta})\hat{f}_{i_\delta}\Delta^{-1}}{v_{qc}(q_{i_\delta}^2; c_{i_\delta})}.$$

   (ii) For all $j = i_\varepsilon, ..., i_\delta - 1$: solve

$$\frac{\nu_q(q_j^2; c_j) - c_j}{v_{qc}(q_j^2; c_j)} = \frac{\eta_{i_\delta} + \sum_{h=(j+1)}^{i_\delta} \hat{f}_h}{\hat{f}_j \Delta^{-1}} \tag{C.52}$$

   to determine q_j^2 (if there is bunching, use ironing as in Section C.4.3 for each j), use the incentive constraint to recursively determine the price,

$$p_j^2 = p_{j+1}^2 + \nu_q(q_{j+1}^2; c_{j+1})(q_j^2 - q_{j+1}^2), \tag{C.53}$$

   and define as the profit of types $\{i_\varepsilon, ..., i_\delta\}$

$$K = \sum_{h=i_\varepsilon}^{i_\delta} (p_h^2 - q_h^2 c_h)\hat{f}_h. \tag{C.54}$$

(3) If $\delta = i^1 - i^2$, the break-even group disappears. If there exists another ("higher") break-even group, check for Pareto-improving transfers to the lowest cost types

in this higher break-even group using the same procedure above.[33] If $\delta = i^1 - i^2$ and there does not exist another break-even group, compute the total profit in $\{i_\varepsilon, ..., N\}$ akin to Step 1.

Otherwise,

$$\bar{V}_{i_\varepsilon} = v(q^2_{i_\varepsilon}; c_{i_\varepsilon}) - p^2_{i_\varepsilon} \tag{C.55}$$

and

$$K_\varepsilon = p^2_{i_\varepsilon} - q^2_{i_\varepsilon} c_{i_\varepsilon}. \tag{C.56}$$

If $K_\varepsilon > 0$, go to (1). If $K_\varepsilon \leq 0$ and $i_\varepsilon > 1$, another break-even interval begins: go to Step 2. If $i_\varepsilon = 1$, stop.

Finally, store the equilibrium allocation (q^{fin}, p^{fin}) which is given by $(q^2_j, p^2_j)_{i_\varepsilon,...,i_\delta}$.

C.4.3 Bunching and ironing

We start with a function $q : \{1, ..., n\} \to [0, \infty)$ that does not satisfy monotonicity, i.e., is strictly decreasing for some set $I \subseteq \{1, ..., n\}$. Thus, it displays "wiggles", i.e., local minima and maxima. We implement the ironing technique described in Section C.3. The procedure is as follows:

Step 1. Determine local maxima and minima.

Define by

$$h = \{i \in \{1, ..., n-1\} : q_{i+1} - q_i < 0\} \tag{C.57}$$

the set of all (left-)points at which q is decreasing. Define by

$$b = \{j \in h : (j-1) \notin h\} \tag{C.58}$$

the set of the smallest (in a neighborhood) of these points. These are local maxima. We enumerate this set as $b_1, ..., b_H$ with $H = |b|$ such that $b_i < b_{i+1}$ and $\{b_1, ..., b_H\} = b$. Similarly, define by

$$l = \{i \in \{1, ..., n-1\} : q_{i+1} - q_i > 0\} \tag{C.59}$$

[33] It is possible that such transfers exist but were not detected earlier due to the availability of even lower cost types $j < i_\varepsilon$.

the set of all (left-)points at which q is increasing and by

$$L' = \{j \in h : (j-1) \notin h\} \tag{C.60}$$

the smallest (in a neighborhood) of these points, which are local minima. Let $L = L' \cup \{1, n\}$, enumerated as $L_1 < ... < L_{H_L}$ with $H_L = |L|$.

Step 2. Ironing.
As long as $|b| > 0$:

(i) Let $b_h = \arg\min b$. Define $b_l = \max\{i \in L : i \leq b_h\}$ and $b_u = \min\{i \in L : i > b_h\}$. Let $\hat{b} = \arg\min_{i \in \{b_h, ..., n\}} q_i$.

(ii) For $j = 0, ..., \hat{b} - b_h$ determine

(a) $q^r = \begin{cases} q_{b_h - j}, & \text{if } b_h - j \geq 1 \text{ \& } b_h - j \geq b_l \\ q_{b_h + j}, & \text{else,} \end{cases}$

$c_A = \min\{i \in \{b_l, ..., 1\} : q_i \geq q^r\}$, and $c_B = \min\{i \in \{b_u, ..., n\} : q_i \geq q^r\}$.

(b) For $k = c_A, ..., c_B$ define

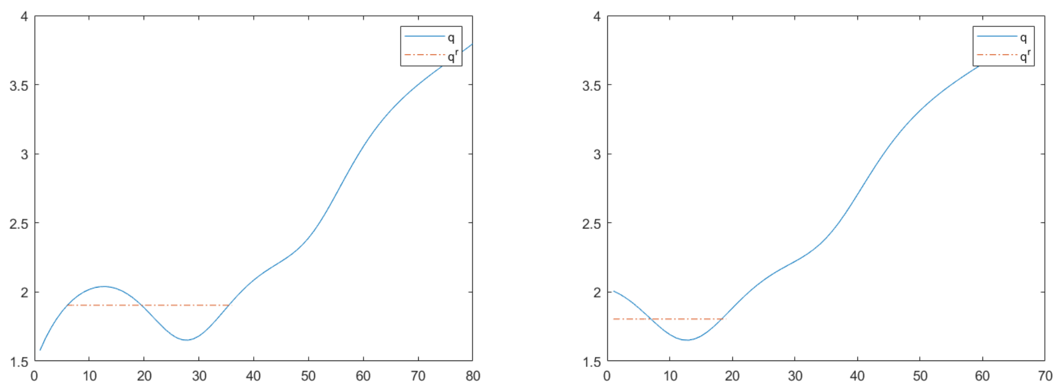
$$\xi'_k = (\nu_q(q^r; c_k) - c_k)\hat{f}_k \Delta^{-1} - \eta_k \nu_{qc}(q^r; c_k), \tag{C.61}$$

where η is defined as in Section C.4, and define

$$K_h = \sum_{k=c_A}^{c_B} \xi'_k \Delta.$$

(iii) Determine the optimal ironing threshold by $\hat{h} = \arg\min\{|K_h|\}$ and use the corresponding q^r, c_A, and c_B. Remove all $b_l < c_B$ from b.

Figure C.7: Illustration of ironing.



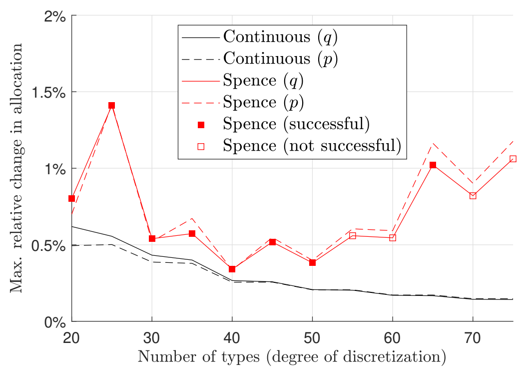
(a) Typical case.

(b) Knife-edge case.

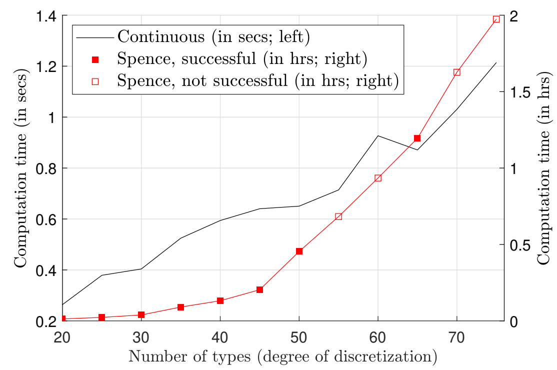
C.5 Numerical stability and efficiency

Figure C.8: Numerical stability and efficiency of continuous equilibrium construction compared to Spence's recursive programs.

Figures depict (a) the maximum relative change in the equilibrium vector of prices p (coverage q) when increasing the equidistant discretization of the type space by 5 additional types and (b) the computation time to solve for the equilibrium. We distinguish between solving Spence's recursive optimization programs (using Matlab's *fmincon* routine with an interior-point algorithm, a maximum of 5,000 iterations, and 10,000 function evaluations) and solving the continuous construction we propose in Section 4. We indicate by filled (empty) squares that the final Spence optimization program for the lowest cost-type converges (does not converge), whereby convergence is defined by reaching an allocation that satisfies first-order optimality and constraints with tolerance 1e-6. The example uses a uniform distribution of types with preferences as in Section 5.



(a) Convergence of allocations.



(b) Computation time.