Abstract

A finite number of vertically differentiated firms simultaneously compete for and screen a continuum of agents with private information about their ability or willingness to pay for quality or quantity. Firms compete through menus of wage-effort or transfer-quality pairs. In equilibrium, higher firms serve higher segments of types. In each segment, the allocation is distorted downward on types below a threshold, but upwards above. The equilibrium approaches the competitive limit as the number of firms grows large. The welfare effects of asymmetric information may be reversed from the monopoly setting. If firms are sufficiently differentiated, then any strategy profile that satisfies the necessary conditions is an equilibrium, and an equilibrium exists.

Keywords. Adverse Selection, Screening, Oligopoly, Incentive Compatibility, Positive Assortative Matching, Vertical Differentiation.

We thank Meghan Busse, Bruno Jullien, Natalia Kovrijnykh, Michael Powell, Luis Rayo, and seminar participants at Northwestern, 2018 SED–Mexico City, and Universidad Alberto Hurtado–Chile for their helpful comments.
# Contents

1 Introduction 1

2 Related Literature 6

3 The Model and Equilibrium 9
   3.1 Interpretation as a Product Market 10

4 Necessity 11
   4.1 Positive Profits on Each Agent Served (PP) 11
   4.2 No Poaching (NP) 11
   4.3 Positive Sorting (PS) 11
      4.3.1 No Extraneous Offers 12
   4.4 Internal Optimality (IO) 13
   4.5 Optimal Boundaries (OB) 15

5 Other Implications of Necessity 16
   5.1 Underprovision at the Bottom and Overprovision at the Top 16
   5.2 Discounts and (Non-)Implementation by Linear Contracts 17
   5.3 The Competitive Limit 18
   5.4 Who Does Asymmetric Information Help or Hurt? 19

6 Existence and Sufficiency 19
   6.1 Stacking and Strict Regularity 20
   6.2 The Main Results 20
   6.3 The Reformulation 21
   6.4 Relating the Original and the Relaxed Problem 21
   6.5 Unique Best Responses 23
   6.6 Sufficiency 25
   6.7 Existence 25

7 Conclusion 26

8 Appendix A: Proofs For Sections 4–5 27
   8.1 Proof of PP 27
   8.2 Proof of NP 28
   8.3 Proof of PS 29
   8.4 Proof of IO 31
   8.5 Proof of OB 35
8.6 Proofs for Section 5 ................................................................. 36

9 Appendix B: Proofs for Section 6 ................................................. 37

9.1 Proofs for Section 6.3 ............................................................... 37
9.2 Proofs for Section 6.4 ............................................................... 38
9.3 Proofs for Section 6.5 ............................................................... 40
  9.3.1 Local Properties of \( r \) ......................................................... 40
  9.3.2 Essentially Unique Optimality ........................................... 44
9.4 Proofs for Section 6.6 ............................................................... 50
9.5 Proofs For Section 6.7 ............................................................. 52
1 Introduction

Screening is an important feature of labor and product markets, and is a pervasive topic in economic theory and applied work. Indeed, the principal-agent model with adverse selection developed in Mussa and Rosen (1978) and Maskin and Riley (1984)—where a firm screens a consumer who has private information about his valuation by providing different quantities or qualities of a good at different prices—is a workhorse in economics of information. In this model, the firm is a monopolist. At the other extreme is the competitive screening model of Rothschild and Stiglitz (1976) and variations thereof—where identical insurance companies screen consumers who differ in and have private information of their riskiness.

Many markets with screening do not fall at these extremes. Instead, a small number of heterogeneous firms both screen their own customers and compete for them in the first place. Examples abound in the health care, labor, and product markets. For example, Saint Laurent must both screen among the customers of its luxury handbags, and compete with Hermes above them and Coach below. Similarly, firms that compete for talent often face the problem of screening their workers into appropriate roles, and competition with vertically differentiated rivals.

Although there are some notable attempts in the literature (see the literature review below), the lack of a standard workhorse for this case has hindered progress in the literature. This is true both for theoretical work, where oligopolistic screening is not well understood, and for recent empirical work, which estimates models of insurance and product markets with an oligopolistic structure but where informational frictions and screening take a restrictive or reduced form.

This paper is an attempt to fill this gap. We develop a natural extension of the Mussa and Rosen (1978) and Maskin and Riley (1984) principal-agent paradigm to an oligopolistic setup with heterogeneous firms. We provide necessary conditions that equilibria must satisfy and then show that, under some economically interpretable assumptions on primitives, these conditions are sufficient, a result that also allows us to prove equilibrium existence. We shed light on the properties of the competitive limit of the model as the number of firms grows large, and study the welfare effects of asymmetric information.

For definiteness, we cast the analysis as a labor market where a finite number of heterogeneous firms that are vertically differentiated in their technology compete for a continuum of heterogeneous workers who differ in their privately known ability or talent. The model can be immediately reinterpreted as one where these firms compete for consumers who differ in their privately known valuation for quality, or as a model of quantity screening under exclusive dealing.

In the model, there is a finite set of firms that differ in the technology by which they transform a worker’s effort into revenue. These technologies are ordered by a single-crossing property, so firms with higher index have a higher marginal revenue for effort. Hence, our model is one of
vertical differentiation across firms. The technology of each firm is additive over all the workers hired (no peer effects). On the other side of the market there is a continuum of workers with quasilinear preferences who differ in their marginal disutility of effort. Ability is a worker’s private information, and can take on a continuum of values. Competition is modelled as the following two-stage game. In stage one, firms simultaneously post menus of contracts, where a menu consists of wage-effort pairs, or, equivalently, utility-effort pairs, one for each type. In stage two, workers each choose the firm and the contract from its menu that suits him best, resolving ties across firms equiprobably. By imposing the workers’ sequentially rational behavior as constraints, we analyze the problem as a simultaneous game among the firms—which are restricted to choosing incentive compatible menus—and then study the properties of its Nash equilibria. A challenge we face is that the resulting game has an infinite dimensional strategy space and discontinuous payoffs.

We first derive a set of properties that any equilibrium exhibits. Several of these necessary conditions are closely related to ones that Jullien (2000) derives in the case of single principal who faces a type-dependent participation constraint, where in our setting, this constraint depends on the most attractive contract the worker of any given ability faces from an alternative employer. We provide alternative proofs for these conditions because some details of our setting are different (in specific at ties) and because we think there is value in elementary proofs that avoid optimal control tools. We will discuss this connection in detail below.

Since our model has private values—the type of a worker enters the firm’s profit only through the contract chosen by the agent—in equilibrium firms make strictly positive profit on each type of worker they hire. Any equilibrium also satisfies no poaching: if a firm does not win a type, then imitating the menu offered by the incumbent to that type yields negative profit to the imitating firm. An implication is that the worker is matched to the firm that most efficiently uses the effort level he exerts. This does not imply that the allocation is efficient: it may be that more total surplus would be generated were the worker to be matched to a different firm and effort level.

Since our model embeds a nontrivial matching problem between firms and workers, an important task is to study equilibrium sorting patterns. We show that any equilibrium entails positive sorting: firms with a higher index hire sets of workers with higher types. Adjacent firms can tie on an interval of types at the boundaries, but if so, they offer a pooling contract to those types and competition drives profits on those types to zero. Equilibrium sorting highlights the dual role that menus play in our model: they are used to screen internally the types of the workers hired, and they serve to attract the right pool of types for the firm. Positive sorting is straightforward in the complete information version of our model due to the supermodularity assumptions and the absence of peer effects within firms. It is somewhat more subtle under incomplete information given the cross-type constraints imposed by incentive compatibility.

---

1. A model with both vertical and horizontal differentiation would also be of interest in many settings, but this is beyond the scope of this paper.
Positive sorting proves fundamental in the analysis of the properties of equilibrium menus and distortions. Since firms hire intervals of types, we can focus on each firm solving for the optimal menu and interval endpoints, subject to the endogenous type-dependent participation constraints induced by the menus offered by the other firms.

On the interval of types for which any given firm is the uniquely best choice in equilibrium, the menu offered by the firm satisfies internal optimality—the effort level of each type is pinned down by a condition that generalizes the standard trade off between efficiency and information rents to reflect that the firm serves only a segment of the market. In addition, each firm must satisfy optimal boundary conditions that determine the endpoints of the relevant interval of types. These conditions reflect the trade off that changing the effort given to a boundary type alters the efficiency for that type, and also attracts or loses some marginal types.

The internal optimality and optimal boundary conditions yield a clear-cut pattern of equilibrium distortions. The highest firm distorts effort downwards for all types. This reflects the standard intuition that lowering the effort of any given type employed lowers the information rents of types above that. In turn, the lowest firm distorts effort upwards for all types. This follows because for the lowest firm, the outside option (of working for someone else) binds for the highest type of worker hired. All other types earn rents. Raising the effort level of any given type thus lowers the information rents of all types below that type. For middle firms, the key is that, holding fixed the utility of the lowest and highest worker the firm employs—the two types on which the participation constraint will turn out to bind—the firm can lower the information rents of workers in the middle of the interval employed by simultaneously lowering the effort of low types employed and raising the effort of high types employed. Hence, the firm distorts effort downwards on workers below a threshold, and upwards for those above.

In an interesting precedent, Dupuis (1848) observes that a rail company provides roofless carriages in third class to “frighten the rich.” As confirmed by the extensive literature on screening, the low quality third class car helps to sell second class seats at a higher price. But Dupuis goes further. The first class passenger receives a “superfluous” level of quality. In the standard model, this is not so. But, in our model, if the rail firm is a middle firm that competes for its richest customers against, for example, a private carriage, then the high type served does indeed receive an inefficiently high level of service. From their point of view, the extra quality is nearly worth the extra cost, and so the extra cost can be largely reflected in the price. But, the superfluous extra quality—with the extra price implied—reduces the temptation for the second class passenger to ride first, and this also helps to reduce their information rents.

Another implication of the analysis is that, when firms are sufficiently differentiated, there are effort gaps at the boundaries between adjacent firms. This is testable: in our setting, we should observe better firms having strictly more productive workers. Similarly, we should observe that products of certain qualities are simply not offered in some markets.
Next, we study the behavior of our oligopoly model as the number of firms grows large and the vertical differentiation between successive firms becomes small. We show that each firm’s profit is bounded above by a constant that goes to zero, as in the competitive limit. In turn, the utility that each type receives in equilibrium converges to the surplus in the efficient match and effort level, again as in the competitive limit.

We finish our examination of the implications of necessity by comparing the outcome of the oligopoly market with and without asymmetric information. In a monopoly, it is unambiguous that full information hurts the worker by destroying information rents, and helps the firm, which gains back the information rents, and can now induce efficient effort. Here, we have a surprising reversal, driven by the new force that under full information a firm can now poach the workers of another firm without worrying about potentially attracting their own existing workers. This competitive force dominates at least for types where the second most efficient firm for the type is close to the first, and full information helps the worker and hurts the firm.

We then turn to the analysis of sufficient conditions for an equilibrium and to the question of equilibrium existence. We show that if firms are differentiated enough—a condition we call stacking—then first, any strategy profile that satisfies positive assortative matching plus the internal optimality and optimal boundaries conditions is essentially an equilibrium; and, second, an equilibrium exists. We view these results as a central contribution of the paper, since first, they are the most challenging technical problems that one must tackle in this setting, second, they are fundamental for economic applications of our model, and third, they are novel in the literature. Indeed, this is a game whose payoff functions are discontinuous and are not quasiconcave in the strategy profile. Thus, the fact that our necessary conditions are also sufficient is both surprising and non-trivial. The lack of quasiconcavity makes convexity of each player’s best responses non-obvious, which complicates the use of off-the-shelf existence results.

The first step of our attack is to restrict attention to menus that yield positive profits on all types, and whose effort levels are within certain bounds that our necessary conditions suggest are reasonable. We show that under stacking, if other firms choose menus with these two properties, then each firm will best respond with menus that also satisfy them. More importantly, stacking then implies that the resulting strategy profile has positive sorting and that the optimal boundary conditions also imply that no poaching holds as well, which deals with inframarginal types.

The second step is to effectively reduce the problem of finding a best response to two dimensions. In particular, we show that under the stated properties, each firm can restrict attention to making a suitable choice of the upper and lower boundaries of the interval of types hired, with the rest of the menu pinned down by internal optimality. Using stacking, we show that payoffs are continuous in this parameterization. A straightforward implication of this result is that the set of best responses is nonempty for each firm.

Even viewed as a two-dimensional optimization problem, we still face significant technical
challenges. Our third key step is an exercise in topography. Fix the behavior of a firm’s opponents, and consider a landscape given by the payoffs to the firm, where the choice of the bottom endpoint is a choice from west to east, while the choice of the top endpoint is a choice from south to north. This landscape has valleys, and indeed, local minima. But, when the other firms play strategies from the previously defined set, we show that the firm has available positive profit strategies. So, consider the “islands” where payoffs are positive. These islands have terrain that is kinked, because the participation constraint—given by the utility offered to each type by the toughest competitor of the firm—has kinks at each type where the relevant opponent changes. Nor need payoffs be quasi-concave even on a given island. We show first that on such islands, any place where the first order conditions are satisfied is also a local maximum, so that any local minima are underwater. Next, fix any position from south to north (a choice of the top endpoint) such that there is some land at that latitude. We show that this defines an interval of latitudes. Fix some such latitude, and consider moving west to east. We show that there is a single interval where one is above water, and that payoffs are strictly quasi-concave as one moves east in this interval. But then, there is a single island, and there is a unique path running down the length of the island that defines the highest point as one moves from west to east. Finally, we show that payoffs are strictly quasi-concave as one moves northward along this path. It follows that the island has a unique peak, and that any point that satisfies the first-order conditions—the optimal boundary conditions—along with the positive profit condition, is in fact that maximum.

Our sufficiency result follows from these steps. Under stacking, if a strategy profile satisfies positive sorting, internal optimality, and the optimal boundary conditions, then it pins down uniquely the interval of types hired by each firm and the optimal effort function, from which one constructs the utility offered to each type hired. And outside the interval of types hired by each firm, one can easily modify the menu offered by each firm to comply with the bounds on effort and positive profits assumed, completing the construction of an equilibrium strategy profile.

It remains to show that an equilibrium exists. To this end, we further restrict attention to a class of strategies that also satisfies a natural bound on the slope of the effort function offered by each firm, and a natural lower bound on the utility that is offered to the highest type of worker. We show that if other firms use strategies that satisfy these conditions, then each firm has a best response with these properties. This class of strategies is sufficiently well-behaved to permit the application of a standard fixed point theorem, delivering existence. In particular, because we have shown that for any given behavior of its opponents, the firm has a unique optimal interval served, and because internal optimality ties down what is happening on that interval, any two best responses will differ only in inessential ways, and the set of such best responses will be convex.

The next section presents a literature review. Section 3 describes the model. Sections 4 derives the necessary conditions and their implications. Section 5 studies distortions, quantity discounts, the competitive limit, and the welfare effects of asymmetric information. Section 6
focuses on sufficiency and existence, presenting the main results and describing the main steps of the proofs. Section 7 concludes. Appendices A and B contain all proofs omitted from the text.

2 Related Literature

The paper is clearly related to the huge literature on principal-agent models with adverse selection, as in Mussa and Rosen (1978), and Maskin and Riley (1984), and the host of papers that build on them (see Laffont and Martimort (2002) for a survey). It is more related to the small literature on oligopoly and price discrimination under adverse selection, nicely surveyed by Stole (2007). The most relevant related papers are Champsaur and Rochet (1989), Biglaiser and Mezzetti (1993), and Stole (1995). Finally, since competition makes each agent’s reservation utility endogenous and type-dependent, the paper is also related to the general analysis of the principal-agent problem with adverse selection and type-dependent reservation utility in Jullien (2000).

Champsaur and Rochet (1989) analyze a two-stage game where two identical firms first choose intervals of qualities they can produce, and then in the second stage offer price schedules to a continuum of consumers. The two-stage nature of their model gives scope for firms to cede parts of the market before price competition takes place, giving very different underlying economics.

Closer to our paper is Stole (1995), who analyzes an oligopoly setting with screening in which a continuum of customers differ along both a vertical and a horizontal dimension, but only one of them is allowed to be private information, so as to keep the analysis one dimensional. When the horizontal differentiation parameter is private information, then the model is an extension of Spulber (1989), and so again not that close to ours. The more interesting case for our purposes is when the vertical dimension is private information while the horizontal one is common knowledge, and where the firms can tailor their offerings to the horizontal type of the agent. In this case, and under symmetric marginal costs, when there are two firms the market partitions into two intervals with each firm serving all vertical types of those customers closest to it. Competition leads the distant firm to offer the product at marginal cost (which is assumed constant across quality levels), and the optimal menu makes customers up to a threshold type indifferent between buying from either firm, while above that type the menu is as in the monopoly screening case. With multiple firms located in a circle, the paper shows how entry reduces distortions in quality.

The critical aspect of Stole’s analysis is that the cost of providing quality is the same across firms. Because of this, if a customer is closer to Firm 1 than to Firm 2, it is efficient for Firm 1 to serve the customer regardless of the customer’s vertical type. But then, the close-by firm faces a standard monopoly screening problem with the outside option of the customer determined endogenously by the best offer the more distant firm can make without losing money.

Spulber (1989), working in a Salop (1979) circular model of horizontal differentiation, considers screening on quantities. The relationship between his analysis and ours is not that close. In particular, his surplus schedule has the same structure as in monopoly, with the intercept determined by the level of competition.
In contrast, our model has no horizontal type. But, in our product market interpretation, the

cost to providing incremental quality differs across firms. It is presumably relatively expensive for
Ferrari to produce a basic economy vehicle, and prohibitively expensive for Hyundai to produce
a purebred performance automobile. Hence, it is not clear to which firm the customer should be
assigned. Indeed, how customers are matched to firms is at the heart of our analysis.

Biglaiser and Mezzetti (1993) analyzes a model with adverse selection and “false moral hazard”
between two heterogeneous principals who differ in their marginal and average cost of production,
and one agent with a continuum of possible types. In equilibrium, the principal with lower
marginal cost serves an interval of types higher than a threshold, the other principal dominates
the low-type segment of the market, and the two principals offer the same pooling contract for
intermediate types, driving profits to zero on that segment. Putting aside the false moral hazard,
this is a special case of our general setup, for which we provide a complete equilibrium analysis.
A crucial assumption in Biglaiser and Mezzetti (1993) is that ties are broken in favor of the firm
that gains the most from that type. This tames payoff discontinuities at ties in a crucial way, but
is less economically natural than our equiprobable tie-breaking assumption.

Another important reference is Jullien (2000). He provides a sophisticated analysis of optimal
menus in a principal-agent model with exogenously given type-dependent reservation utility, and
shows that both upward and downward distortions can emerge. Holding fixed the behavior of the
other firms, the problem facing each of our firms is similar to the one in Jullien (2000), with the
key difference being that he assumes that if the firm offers the worker surplus equal to his outside
option, then the agent accepts, while in the oligopoly setting, ties are broken equiprobably. This
makes some difference at a technical level. Existence of a best-response is no longer guaranteed,
and it becomes harder to analyze the problem using standard control techniques. Because of
this, we have to work harder than Jullien (2000) to prove, for example, that any optimal contract
implements actions that are continuous over the range of types employed.

Because of the similarity of the underlying problems those of our necessary conditions that
derive solely from the implications of best-responses on a firm-by-firm basis are the same as those
in Jullien (2000). In particular, our positive profit, internal optimality, and optimal boundary
conditions each has a close relative in Jullien. In turn, those of our necessary conditions which
are derived from the interplay of the incentives of one firm and another are novel. And since
our model with competition endogenizes the agents’ type-dependent reservation utility, we can
provide more clear-cut predictions of the pattern of distortions that must arise in any equilibrium.

In his Theorem 4, Jullien (2000) shows that under one of two conditions (which our model
satisfies) his necessary conditions are also sufficient with full participation, while in Section 4, he
describes how to extend his model to also handle cases where full participation is not optimal. It
is tempting to conclude that, suitably modified, Jullien’s analysis also implies that our necessary

\footnote{This generalizes the main insight in Maggi and Rodriguez-Clare (1995).}
conditions are sufficient for optimality in our setting where each firm hires only some of the workers. However, as that paper recognizes (p.17, second paragraph), one must be extremely careful in applying the sufficiency part of Theorem 4 and the ideas of Section 4 at the same time.

To see the issue, the idea of Section 4 in Jullien (2000) is to add an artificial technology that mimics the action and surplus that the agent gets at his outside option (in our setting, the agent’s favorite offering from the other firms) but does so at zero profit for the firm. The firm’s profits are then the maximum of those associated with the original production function and the artificial technology. With this modified production function, the principal is willing to hire all workers, and so, Jullien argues, we can solve the full participation problem per his original analysis, and then simply drop workers in the (zero profit) regions where the modification was relevant.

The problem is that the heart of the proof of sufficiency in Theorem 4 (in either variation) is that the benefit to the firm from effort is concave. But, because the modified profit function introduced in Section 4 of Jullien (2000) involves a maximum, it will typically have an upward kink where it transitions from one function to the other. Hence, the modified production function is not concave, and Jullien’s analysis does not apply. One major contribution of our paper is to provide a proof of sufficiency that allows for the fact that with competition, each firm will have less than full participation. This construction is also at the heart of our existence proof.

Since our model embeds a nontrivial matching problem between firms and sets of workers, the paper relates to the literature on many-to-one matching problems with transfers, as in the seminal papers of Crawford and Knoer (1981) and Kelso Jr and Crawford (1982), who provide conditions under which a competitive equilibrium exists without private information. A recent paper that sheds light on sorting in matching models with “large” firms (and complete information) is Eeckhout and Kircher (2018). Our model can be thought of as a matching setting under one-sided incomplete information, where firms that differ in their technology compete in a noncooperative fashion for sets of workers with private information about their disutility of effort.

Finally, there is a large literature on competitive markets with adverse selection in the tradition of Rothschild and Stiglitz (1976), including some recent contributions that feature imperfect competition driven by search frictions, as in Guerrieri, Shiuler, and Wright (2010) and Lester, Shourideh, Venkateswaran, and Zetlin-Jones (2018). Our setup differs in two fundamental ways from that literature: first, we have a small number of heterogeneous firms or principals, and thus there is oligopolistic competition and a nontrivial sorting problem; and second, unlike the insurance problem ours is a model with private values. A tighter connection between our approach and this literature must await the extension of our analysis to common values.

---

4Our results thus also prove sufficiency for a class of type-dependent reservation utility models not covered by Jullien’s analysis. In essence, one needs the slope of the type-dependent reservation utility to satisfy a “steep or shallow” condition similar to our stacking condition. This is natural in our setting if firms are sufficiently vertically differentiated, and may be natural in others.

5Another example in the literature with positive sorting under incomplete information can be found in Liu, Mailath, Postlewaite, and Samuelson (2014).
We assume function adds tractibility and we believe does not subtract significantly from the economics of the situation.

\[ \alpha \]

compatible menus. Thus, a menu can equally well be described as (wherever it is clear which firm we are talking about, we suppress the
\[ n \]

objects on either side have strictly the same sign. We follow the hierarchy Lemma, Proposition, Theorem. Finally, of \[ x \]

needed, and similarly with positive and negative, and concave and convex. Also, for any function \[ f \] and argument \[ x \] of \[ f \], we write \( (f)_x \) for the total derivative of \( f \) with respect to \( x \). We use the symbol \( =_s \) to indicate that the objects on either side have strictly the same sign. We follow the hierarchy Lemma, Proposition, Theorem. Finally, wherever it is clear which firm we are talking about, we suppress the \( n \) superscript.

\[ \frac{\alpha}{\theta} \]

adverse selection, we assume that effort is observable, thus ruling out moral hazard issues.

There is a unit measure of agents (workers) and \( N \) principals (firms). Agents differ in a parameter \( \theta \in [\underline{\theta}, \overline{\theta}] \) with cumulative distribution function (cdf) \( H \) with strictly positive and \( C^1 \) density \( h[\theta] \).

We assume \( H \) and \( 1 - H \) are strictly log-concave.\(^7\)

Each worker exerts effort \( a \geq 0 \) at cost \( c(a, \theta) = (1 - \theta)a \). If type \( \theta \) exerts effort \( a \) and obtains wage \( w \), then his utility is \( w - c(a, \theta) \). For simplicity, we assume that the worker has no outside option beyond choosing among the offers of the various firms. To zero in on competition under adverse selection, we assume that effort is observable, thus ruling out moral hazard issues.

If firm \( n \) pays wage \( w \) to a worker who exerts effort \( a \), then the firm’s payoff is \( B^n(a) - w \), where \( B^n \) is \( C^2 \), increasing, and strictly concave in \( a \). We assume that \( B^n_\theta > B^n_\theta^{n-1} \) for all \( n \), that is, \( B^n \) is supermodular in \( (a, n) \). Firms do not have capacity constraints and their technology is additively separable in the workers they hire. Purely to avoid corner solutions, it is convenient to assume that \( \lim_{a \to 0} B^n_\theta(a) = \infty \), and \( \lim_{a \to \infty} B^n_\theta(a) = 0 \).

Firms simultaneously offer menus of contracts, where Firm \( n \)’s menu is a pair of functions \( (\alpha^n, w^n) \), where \( \alpha^n(\theta) \) is the action required of a worker who chooses Firm \( n \) and announces type \( \theta \), and \( w^n(\theta) \) is his wage. Contracts are exclusive: each worker can work for only one firm.

Let \( v^n \) be the surplus function for a worker who takes the contract of firm \( n \), given by \( v^n(\theta) = w^n(\theta) - c(\alpha^n(\theta), \theta) \). By the revelation principle, it is without loss that firms offer incentive compatible menus. Thus, a menu can equally well be described as \( (\alpha^n, v^n) \), where incentive compatibility is equivalent to requiring that \( \alpha^n \) increasing and that \( v^n_\theta(\theta) = -c_\theta(\alpha^n(\theta), \theta) \) for almost all \( \theta \). We will do so henceforth. Let \( \pi^n(\theta, a, v_0) = B^n(a) - c(a, \theta) - v_0 \) be the principal’s profit if the agent of type \( \theta \) takes action \( a \), and is given utility \( v_0 \). For any \( \alpha \) and \( v \), we write \( \pi^n(\theta, \alpha, v) \) as shorthand for \( \pi^n(\theta, \alpha(\theta), v(\theta)) \).

After observing the posted menus, workers sort themselves to the most advantageous firm. Formally, for each \( n \), let \( S^n \) be the set of pairs \( (\alpha^n, v^n) \) where \( \alpha^n \) is increasing and \( v^n \) is continuous, with \( v^n_\theta(\theta) = -c_\theta(\alpha^n(\theta), \theta) = \alpha^n(\theta) \) for almost all \( \theta \). Let \( s^n = (\alpha^n, v^n) \) be a typical element of \( S^n \). The joint strategy space is \( S = S^1 \times \cdots \times S^N \) with typical element \( s \). Let \( s^{-n} \in \times_{n' \neq n} S^{n'} \) be a typical strategy profile for players other than \( n \). For any \( s \in S \), and for any \( n \) and \( \theta \), define the scalar-valued function

\[ v^{-n}(\theta) = \max_{n' \neq n} v^{n'}(\theta) \]

\(^6\)We use increasing and decreasing in the weak sense of nondecreasing and nonincreasing, adding ‘strictly’ when needed, and similarly with positive and negative, and concave and convex. Also, for any function \( f \) and argument \( x \) of \( f \), we write \( (f)_x \) for the total derivative of \( f \) with respect to \( x \). We use the symbol \( =_s \) to indicate that the objects on either side have strictly the same sign. We follow the hierarchy Lemma, Proposition, Theorem. Finally, wherever it is clear which firm we are talking about, we suppress the \( n \) superscript.

\(^7\)As is standard, our model is equivalent to one with a single worker drawn from \( H \).

\(^8\)It natural in this interpretation to assume \( [\frac{\theta}{\theta}, \frac{\theta}{\theta}] \subseteq [0, 1] \), but this plays no formal role. The simplified cost function adds tractibility and we believe does not subtract significantly from the economics of the situation.
as the most surplus offered by any of $n$’s competitors. Let $a_n^{-1}$ be the scalar-valued associated action function. That is, $a_n^{-1}$ is an increasing function that is almost everywhere equal to $v_{\theta}^{-n}$. From the point of view of $n$, $(a_n^{-1}, v_n^{-1})$ summarizes all the relevant information about the strategies of his opponents. Define

$$
\varphi^n(\theta, s) = \begin{cases} 
0 & v^n(\theta) < v_n^{-1}(\theta) \\
1 & v^n(\theta) = v_n^{-1}(\theta) \\
\#\{n' \in \{1, \ldots, N\} | v_{n'}(\theta) = v_n^{-1}(\theta)\} & v^n(\theta) > v_n^{-1}(\theta)
\end{cases}
$$

That is, ties are broken equiprobably. Define

$$
\Pi^n(s_n) = \int \pi^n(\theta, \alpha^n_n, v^n_n) \varphi^n\left(\theta, s_n\right) h(\theta) \, d\theta.
$$

Let $BR^n(s_n) = \arg \max_{s_n} \Pi^n(s_n, s_n^{-1})$. Strategy profile $s$ is a Nash equilibrium (in pure strategies) of the game $(S^n, \Pi^n)_{n=1}^N$, if for each $n$, $s_n \in BR^n(s)$. Define $\alpha^n_n(\theta) = \arg \max_a (B^n(a) - c(a, \theta))$, and let $v^n_n(\theta) = B^n(\alpha^n_n(\theta)) - c(\alpha^n_n, \theta)$ be the most surplus $n$ can offer type $\theta$ without losing money. Define $v_n(\theta) = \max_{n \in \{1, \ldots, N\}} v^n_n(\theta)$ as the most surplus that any firm can offer type $\theta$, and let $v_n^{-1}(\theta) = \max_{n' \neq n} v^n_n(\theta)$ be the most surplus that any firm other than $n$ can offer $\theta$. We will assume that for all $n$ there exists $\theta^n \in \Theta$ such that

$$
v^n_n(\theta^n) > v_n^{-1}(\theta^n) \quad (1)
$$

That is, $n$ is the unique firm that can create maximal surplus for type $\theta^n$. Because $B^n - c$ is strictly supermodular in $(a, \theta)$, the intervals over which each firm is the unique maximizer of $v^n_n(\theta)$ will be ordered by the names of the firms, and so in particular, we can take $\theta_1^n = \theta$, and $\theta_N^n = \bar{\theta}$. The existence of $\theta^n$ for all $n$ will imply that each firm in equilibrium employs a positive measure of workers and earn strictly positive profits.

### 3.1 Interpretation as a Product Market

To reinterpret our model as one of product quality, assume that each customer has a unit demand, and that the value of consuming a unit of a product of quality $a$ is $-c(a, \theta) = (\theta - 1)a$, where we take $\theta > 1$, and where since $-c_{a\theta} = 1$, higher $\theta$ types put higher marginal value on quality. The production cost of quality $a$ to firm $n$ is $-B^n(a)$, so that higher indexed firms have lower incremental cost of providing quality. A reinterpretation in terms of quantity provision under exclusive dealing is also straightforward.
4 Necessity

In this section, we present a set of necessary conditions for a Nash equilibrium in pure strategies. We begin by stating the theorem. Then, we define each of the terms, show why it must hold, and flesh out its economic implications. All formal proofs are in Appendix A.

Theorem 1 (Necessity) Every pure strategy Nash equilibrium with no extraneous offers satisfies positive profits, no poaching, positive sorting, internal optimality, and optimal boundaries.

4.1 Positive Profits on Each Agent Served (PP)

The positive profits condition (PP) is satisfied if for each \( n \), the probability that \( n \) hires a worker on whom he strictly loses money is 0. We prove the stronger statement that for any \( s = (s^n, s^{-n}) \) (equilibrium or not), \( s^n \) can be transformed to a strategy that is equivalent anywhere it earns positive profits, but eliminates any situation where it loses money. To see the intuition, let \( P \) be the set of types on which \( n \) makes money. Eliminate all action-wage offerings for workers not in \( P \), and let \( \hat{v} \) be the resultant surplus function. Workers in \( P \) have fewer deviations available, workers not in \( P \) who go to another firm save the firm money, and workers not in \( P \) who accept the same contract as some worker in \( P \) are now profitable, using private values. One key implication of PP is that each firm earns strictly positive profits in equilibrium: since firms do not lose money, a firm that offers \( v^n - \varepsilon \) will win near \( \theta^n \). See Corollary 3. Another key implication is that there is no cross-subsidization. Losing money on some types does not enhance the profits earned on others.

4.2 No Poaching (NP)

The no poaching condition (NP) holds if for all \( \theta \) such that \( \varphi(\theta, s) < 1 \), \( \pi^n(\theta, a^{-n}, v^{-n}) \leq 0 \). That is, if \( \theta \) is not always hired, then imitating \( \theta \)'s equilibrium contract is unprofitable. Intuitively, Firm \( n \) can offer a worker not in \( [\theta_l, \theta_h] \) a deal that replicates that of the incumbent plus some \( \varepsilon \) in surplus without affecting the behavior of any type in \( (\theta_l, \theta_h) \), on whom it is currently offering uniquely the best deal. As such, NP is about stealing the inframarginal workers of another firm. Under NP, each firm offers a deal at least as good as could the second most capable firm at the action chosen. This bound is strongest when firms have similar capabilities (when \( \pi^n \) and \( \pi^{n+1} \) are similar). The emphasis is important: a firm might profitably outcompete \( n \) on type \( \theta \) with another action, but only at the cost of attracting some of its existing workers in a detrimental way.

4.3 Positive Sorting (PS)

Our sorting condition states that the sets of workers employed by consecutive firms are consecutive intervals. If these intervals overlap, then the tied firms earn zero profits on the relevant workers.
For \( n \in \{1, \ldots, N - 1\} \), let \( \hat{\alpha}^n \) be the unique solution to \( B^n(a) = B^{n+1}(a) \). Say that \( s \) has positive sorting (PS) if there is \( \{\theta^n_h, \theta^n_l\}_{n=1}^N \) such that
\[
\theta = \theta^1_l < \theta^1_h < \theta^2_l < \cdots < \theta^N_l < \theta^N_h = \bar{\theta},
\]  
with \( \alpha^n(\theta^n_l) \leq \alpha^{n+1}(\theta^{n+1}_l) \) for all \( n \in \{1, \ldots, N - 1\} \), with
\[
\varphi^n(\theta, s) = \begin{cases} 
0 & \text{if } \theta < \theta^n_{h-1} \text{ or } \theta > \theta^n_{l+1} \\
\frac{1}{2} & \text{if } \theta \in [\theta^n_{h-1}, \theta^n_l] \text{ or } \theta \in [\theta^n_{l+1}, \theta^n_{h-1}] \\
1 & \text{if } \theta \in (\theta^n_l, \theta^n_h) 
\end{cases}
\]  
and where on \( (\theta^n_h, \theta^n_{l+1}) \), \( \alpha^n(\theta) = \hat{\alpha}^n = \alpha^{n+1}(\theta) \), and \( v^n(\theta) = B^n(\hat{\alpha}^n) - c(\hat{\alpha}^n, \theta) = B^{n+1}(\hat{\alpha}^n) - c(\hat{\alpha}^n, \theta) = v^{n+1}(\theta) \). That is, (i) higher indexed firms hire consecutively higher intervals of workers, with actions monotone at boundaries, (ii) firms hire all workers in \( (\theta^n_l, \theta^n_h) \) and half of those in regions of overlap, and (iii) if two firms strictly overlap, then they do so at action \( \hat{\alpha}^n \) and profit zero. An implication is that \( v^n(\theta^n_l) = v^{-n}(\theta^n_l) \) and \( v^n(\theta^n_h) = v^{-n}(\theta^n_h) \), where “=” is relaxed to “≥” at \( \theta^1_l = \theta \) and \( \theta^N_h = \bar{\theta} \).

**Proposition 1** Any pure strategy equilibrium with no extraneous offers has PS.

To see the intuition for PS, fix \( \theta' > \theta \). By incentive compatibility, \( \theta' \) is taking an action at least as high as \( \theta \) in equilibrium. But, \( B^n(a) \) is strictly supermodular in \( n \) and \( a \). Hence, if \( n \) sometimes hires \( \theta' \) and \( n' > n \) sometimes hires \( \theta \), then, by PP and NP, either \( n \) will want to always hire \( \theta' \), or \( n' \) will want to always hire \( \theta' \). The only exception is if both firms are indifferent about hiring both \( \theta \) and \( \theta' \), and this can only happen if actions are constant and \( B^n \) is equal to \( B^n' \) on the tied interval, and profits are dissipated.

Say that \( s \) has strictly positive sorting (SPS) if \( \theta^n_n = \theta^{n+1}_n \) for all \( n \in \{1, \ldots, N - 1\} \), so that there are no regions of ties. Under SPS, there will often be gaps in the effort level induced (or set of products offered) as one moves from one firm to the next. Indeed, a gap seems the generic outcome, as \( v^n \) must cross \( v^{n+1} \) from below at \( \theta^n_{l+1} \), and only for carefully chosen parameters will this crossing be tangential. Figure 1 shows a typical example with SPS and three firms.

### 4.3.1 No Extraneous Offers

Without some refinement, one can have equilibria in which firm \( n \) wins “always” on \( (\theta^n_h, \theta^n_{l+1}) \), but at some zero measure set of points, \( \varphi^n(\theta) = 1/2 \), since \( v^n(\theta) \) is equal, for example, to \( v^{n+1}(\theta) \). While these offers by \( n + 1 \) may lose money when accepted, they do not hurt, since they have measure zero. Similarly, there may be discontinuities in \( \alpha^{n+1} \) outside of the region where \( n + 1 \) wins that create complicated incentives for \( n \).
Figure 1: **An example with three firms and strict positive sorting** Firm 2 hires workers between $\theta_l^2$ and $\theta_h^2$, with Firm 1 hiring lower workers, and Firm 3 higher workers.

For technical and aesthetic reasons, we wish to rule out these difficulties. In Lemma 6 we show that any best response for $n$ must be continuous on $(\theta_{n-1}^l, \theta_n^l)$, the region over which $n$ ever wins. We will consider equilibria in which each $\alpha_n$ is continuous everywhere, and in which $\alpha_n^{n+1} - \alpha_n^{n} > 0$ outside of $[\theta_{n-1}^h, \theta_n^l]$, a condition we will term *no extraneous offers* (NEO). Condition NEO will hold, for example, if each $n$ offers the action to types above $\theta_n^{n+1}$ that it offers to $\theta_n^l$ and the same action to types below $\theta_n^{n-1}$ that it offers to $\theta_n^h$.

### 4.4 Internal Optimality (IO)

Fix $n$, and define $\gamma_n$ by

$$\pi_n(a, \gamma_n(a, \kappa), v) = \frac{\kappa - H(b)}{h(b)}.$$  \hspace{1cm} (4)

Note that $\gamma_n$ is strictly decreasing in $\kappa$. Strategy profile $s$ satisfies *internally optimality* (IO) if for each $n$, there is $\kappa^n \in (H(\theta_l^n), H(\theta_h^n))$ such that $\alpha^n = \gamma^n(\cdot, \kappa^n)$ on $(\theta_l^n, \theta_h^n)$, where $\kappa^1 = 0$, and $\kappa^N = 1$.

---

9It is an open question whether there are interesting settings in which this refinement rules out existence.

10Lemma 7 in Section 8.4 shows that log-concavity of $H$ and $1 - H$ imply that $(\kappa - H(\cdot))/h(\cdot)$ is decreasing for all $\kappa \in [0, 1]$, and so, since $\pi_{a\theta} = 1 > 0$, $\gamma^n(\cdot, \kappa)$ is indeed strictly increasing under IO.
To see why IO holds, fix \(\theta_l\) and \(\theta_h\), and let \(\mathcal{P}(\theta_l, \theta_h)\) be the optimization problem given by

\[
\max_{(\alpha, v)} \int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v) h(\theta) d\theta \\
\text{s.t. } v(\theta_l) \geq v^{-n}(\theta_l) \tag{5} \\
v(\theta_h) \geq v^{-n}(\theta_h), \text{ and} \tag{6} \\
v_\theta(\theta) = \alpha(\theta) \text{ for almost all } \theta \in [\theta_l, \theta_h], \tag{7}
\]

This is a relaxation of the problem actually faced by the firm, since we drop monotonicity of \(\alpha\), ignore the value of \(v\) except at \(\theta_l\) and \(\theta_h\), and check relaxed versions of the constraint at \(\theta_l\) and \(\theta_h\). We will deal with the optimal choice of \(\theta_l\) and \(\theta_h\) in the next section.

Define

\[z(\theta_l, \theta_h, \kappa) = v^{-n}(\theta_h) - v^{-n}(\theta_l) - \int_{\theta_l}^{\theta_h} \gamma(\theta, \kappa) d\theta.\]

Note that \(z(\theta_l, \theta_h, \cdot)\) is increasing. Define \(\tilde{\kappa}(\theta_l, \theta_h)\) as \(H(\theta_l)\) if \(z(\theta_l, \theta_h, H(\theta_l)) > 0\), as \(H(\theta_h)\) if \(z(\theta_l, \theta_h, H(\theta_h)) < 0\), and as the solution to \(z(\theta_l, \theta_h, \kappa) = 0\) otherwise.

**Lemma 1 (Relaxed Problem)** Problem \(\mathcal{P}(\theta_l, \theta_h)\) has a solution \(\tilde{s}(\theta_l, \theta_h) = (\tilde{\alpha}, \tilde{v})\). On \([\theta_l, \theta_h]\), \(\tilde{s}\) is unique and has \(\tilde{\alpha} = \gamma(\cdot, \tilde{\kappa}(\theta_l, \theta_h))\), where if \(\tilde{\kappa}(\theta_l, \theta_h) > H(\theta_l)\) then \(\tilde{v}(\theta_l) = v^{-n}(\theta_l)\), and if \(\tilde{\kappa}(\theta_l, \theta_h) < H(\theta_h)\) then \(\tilde{v}(\theta_h) = v^{-n}(\theta_h)\).

To see the intuition for the lemma, start from the case where \(6\) is slack. Raise effort a little at any given \(\theta\) in \((\theta_l, \theta_h)\). This gains \(\pi_a(\theta, \alpha, v)\) on \(h(\theta)\) workers, but raises the surplus of the \(H(\theta_h) - H(\theta)\) workers above \(\theta\) at rate \(-c_{a\theta} = 1\). For this perturbation not to be profitable, we must have \(\pi_a(\theta, \alpha, v) h(\theta) - (H(\theta_h) - H(\theta)) = 0\), or equivalently, \(\tilde{\alpha} = \gamma(\cdot, H(\theta_h))\), where we note that the standard monopoly screening problem has \(\theta_h = \bar{\theta}\), and hence \(\kappa = 1\). Similarly, if \(5\) is slack, then \(\kappa = H(\theta_l)\).

It cannot be that both \(5\) and \(6\) are slack, since then a reduction of \(v\) by a constant is profitable. Assume that both \(5\) and \(6\) bind at the optimum. For any given \(\theta'\) and \(\theta'' > \theta'\), with \(\theta', \theta'' \in (\theta_l, \theta_h)\), raise effort near \(\theta'\) while lowering effort near \(\theta''\) by an equal amount, raising the surplus of workers between \(\theta'\) and \(\theta''\), but leaving the utility at \(\theta_l\) and \(\theta_h\) unchanged. The benefit to the firm is thus

\[
\pi_a(\theta', \alpha, v) h(\theta') - (H(\theta'') - H(\theta')) - \pi_a(\theta'', \alpha, v) h(\theta''). \tag{8}
\]

At an optimal menu, the value of this perturbation must be zero, and so \(\kappa = H(\theta) + \pi_a(\theta, \alpha, v) h(\theta)\) is constant on \((\theta_l, \theta_h)\), giving \(4\). Note that \(\kappa\) is tied down uniquely by \(5\) and \(6\), and that if \(\kappa > H(\theta_h)\), then it is better to lower \(\kappa\) to \(H(\theta_l)\), making \(6\) slack, and if \(\kappa < H(\theta_l)\) then it is better to raise \(\kappa\) to \(H(\theta_h)\), making \(5\) slack.
To prove that the firm’s optimal solution must also be of this form, assume that \((\alpha, v)\) does not agree with \(s\) on \((\theta_l, \theta_h)\). We begin by perturbing \((\alpha, v)\) linearly in the direction of \(s\), but modify the perturbation to keep payoffs greater than \(v - n\) on \((\theta_l, \theta_h)\), so that the firm continues to hire those workers. Conducting this modification while maintaining monotonicity of \(\alpha\) may mean hiring some workers outside of \((\theta_l, \theta_h)\). Using \(PP\), we next purge any unprofitable workers. We show that despite these two modifications, the initial impact of this perturbation is at least as profitable as simply moving in the direction of \(s\). But, since \(s\) is the unique solution to \(P(\theta_l, \theta_h)\), and since the objective function in \(P(\theta_l, \theta_h)\) is concave, moving in this direction raises profits, a contradiction. We will show in the next section that for \(n \notin \{1, N\}\), \(\kappa\) is interior in equilibrium.

An economic implication of \(IO\) is that there is complete sorting of workers—or, alternatively, a complete product line—within the interval of types served by each firm. This of course depends on the absence of a fixed cost per menu item.

### 4.5 Optimal Boundaries (\(OB\))

Strategy profile \(s\) satisfies the optimal boundary condition (\(OB\)) if

\[
\pi^n(\theta^n_l, \alpha^n, v^n) - \pi^n_a(\theta^n_l, \alpha^n, v^n)(\alpha^n(\theta^n_l) - a^{-n}(\theta^n_l)) = 0, \quad (9)
\]

\[
\pi^n(\theta^n_h, \alpha^n, v^n) + \pi^n_a(\theta^n_h, \alpha^n, v^n)(a^{-n}(\theta^n_h) - \alpha^n(\theta^n_h)) = 0, \quad (10)
\]

where (9) is discarded for Firm 1, and (10) for Firm \(N\).

Each equation asks for a balance between the direct profit on the boundary type of worker and a term which is the product of (i) the marginal profit of requiring a higher action of the boundary worker and (ii) the difference in action between firm \(n\) and the adjacent firm at the boundary type. Recall that by \(PS\) these differences are positive.

In contrast to \(NP\), which is about stealing potentially distant workers, \(OB\) reflects that small changes in the set of workers employed are not profitable.

To see the intuition for \(OB\), fix \(n\) and increase the effort of types near \(\theta_h\) a little. This has direct benefit \(\pi_a(\theta_h, \alpha, v)h(\theta_h)\). But, as \(v(\theta_h)\) is raised, \(\theta_h\) is increased at rate \(1/(a^{-n}(\theta_h) - \alpha(\theta_h))\) and so the profit on the new workers hired is \(\pi(\theta_h, \alpha, v)h(\theta_h)/(a^{-n}(\theta_h) - \alpha(\theta_h))\). Cancelling \(h(\theta_h)\) and rearranging yields \(10\). The derivation of \(9\) is similar, noting that to lower \(\theta_l\) and gain extra workers, one reduces effort on types near \(\theta_l\), holding fixed the surplus of higher types.

**Lemma 2** Properties \(IO\) and \(OB\) together imply \(PP\).

To see this, note that by \(10\), since by \(IO\) \(\pi_a(\theta_h, \alpha, v) \leq 0\), and since by \(PS\), \(\alpha^{n+1}(\theta_h) \geq \alpha^n(\theta_h)\), we must have \(\pi(\theta_h, \alpha, v) \geq 0\), and similarly, by \(9\) and \(IO\), \(\pi(\theta_l, \alpha, v) \geq 0\). But then, \(\pi(\cdot, \alpha, v)\), which is strictly single-peaked, must be positive everywhere between \(\theta_l\) and \(\theta_h\). Con-
dition $PP$ is noted because it is of independent interest, and because proving $PP$ is an essential building block toward proving $IO$ and $OB$.

Finally, let us see that $\kappa$ is interior for $n \notin \{1, N\}$. Assume $\kappa = H(\theta_h)$. Note that

$$(\pi(\theta, \alpha, v))_{\theta} = \pi_a(\theta, \alpha, v) \alpha(\theta),$$

because $-c_\theta(\alpha(\theta), \theta) = \alpha(\theta) = v_\theta(\theta)$. By (4) and (11), $\pi(\theta, \alpha, v)$ is thus increasing in $\theta$, and so, since profits are positive by Lemma 3, $\pi(\theta_h, \alpha, v)$ is strictly positive, and since $\kappa = H(\theta_h)$, we have $\pi_a(\theta_h, \alpha, v) = 0$, and so (10) is violated. Essentially increasing effort on types near $\theta_h$ both moves them closer to the efficient action and gains some extra workers on whom profits are strictly positive. Similarly, $\kappa > H(\theta_1)$.

The definition of $OB$ discards (9) for $n = 1$ and (10) for $n = N$, rather than replacing them with inequalities. The reason is that given the above discussion, $IO$ implies that, holding fixed $\theta_h$, Firm 1 is better off with $\kappa = 1$ and $\theta = \theta$ than with any higher $\theta_1$. Hence, checking optimality of $\theta_h$ for Firm 1 is enough, and similarly for $N$.

5 Other Implications of Necessity

5.1 Underprovision at the Bottom and Overprovision at the Top

Let $\theta_0 \in [\theta_l, \theta_h]$ be such that $H(\theta_0) = \kappa$. A surprising feature of the solution is that $\pi_a = B_a - c_a < 0$ for $\theta \in (\theta_0, \theta_h]$, an interval which is non-empty except for $n = N$. That is, each firm except the highest chooses to ask too much effort for workers towards the top of his range of participation (over-provide quality for his most quality-sensitive customers). In turn, the firm asks too little from workers towards the bottom of his range of participation, that is, $\pi_a = B_a - c_a > 0$ for workers with $\theta \in [\theta_l, \theta_0)$, an interval that is non-empty except for $n = 1$. These two distortions from efficiency allow the principal to lower the surplus of the agents in the middle. Since $\kappa^1 = 0$, Firm 1 distorts effort upward for all types except $\bar{\theta}$, while $\kappa^N = 1$, and so Firm $N$ acts as in a standard screening model, distorting effort downward for all types except $\bar{\theta}$. See Figure 2 for an illustration with three firms.

By (11), profits are single peaked with maximum at $\theta_0$. This has some intuition. Customers in the middle of the participation range find neither of the alternative firms very attractive.
5.2 Discounts and (Non-)Implementation by Linear Contracts

Let the tariff $T^n$ paid by the firm to the worker associated with action $a$ be implicitly defined by $T^n(\alpha^n(\theta)) = v^n(\theta) + c(\alpha^n(\theta), \theta)$. Then,

$$T^n_a(\alpha^n(\theta))\alpha^n_\theta(\theta) = v^n_\theta(\theta) + c_a(\alpha^n(\theta), \theta)\alpha^n_\theta(\theta) - \alpha^n(\theta)$$

and hence, $T^n_a(\alpha^n(\theta)) = c_a(\alpha^n(\theta), \theta)$. But then, using the properties of $c$, $T^n_{aa}(\alpha^n(\theta)) = -1$, and so $T$ is strictly concave. It follows first that there are ‘quantity discounts’: the wage per unit of effort decreases in the amount of effort, and hence higher types obtain a lower wage per unit of effort. Further, since $T^n$ is strictly concave, it cannot be implemented using a menu of its tangents, that is, using linear contracts (see Laffont and Martimort (2002), Section 9.5).

Similarly, in the product interpretation of the model the amount paid by the consumer to the firm is $\tilde{T}^n(\alpha^n(\theta)) = -v^n(\theta) - c(\alpha^n(\theta), \theta)$, which, arguing as before, is strictly convex, and hence once again cannot be implemented by a menu of linear contracts. It can also be shown that $\tilde{T}/a$ increases in $a$, and therefore there are quantity premia.

---

This is one place where the simple cost function we use plays a role.
5.3 The Competitive Limit

We now explore the behavior of our economy as \( N \) grows. Let \( d_1 = \max_{a,n} (B^n(a) - \max_{n' \neq n} B^n'(a)) \).

When \( d_1 \) is small, then for any firm \( n \) and action \( a \), there is another firm for whom \( B^n'(a) \) is nearly as large as \( B^n(a) \). Also, for each firm \( n \), define \( (a^n_l, a^n_h) \) as the interval of actions over which \( B^n(a) > \max_{n' \neq n} B^n'(a) \), and define \( d_2 = \max_{n} (a^n_h - a^n_l) \) as the longest such interval. Each of \( d_1 \) and \( d_2 \) is a measure of how far firms are apart.

**Example 1** Let \( B(a,t) = a - (a - t)^2 \) and for each \( n \in \{1, \ldots, N\} \) let \( B^n(a) = B(a,n/N) \).

Then, \( d_1 = 1/N^2 \) and \( d_2 = 1/N \).

In this example, as \( N \) grows large, \( d_1 \) and \( d_2 \) both converge to 0. One would in general expect that \( d_1 \) and \( d_2 \) will be small in economies with many firms.\(^{12}\)

**Lemma 3** For all \( n \), and for all \( \theta \) that \( n \) serves,

\[
0 \leq \pi^n(\theta, \alpha^n, v^n) \leq B^n(\alpha^n(\theta)) - \max_{n'} B^n'(\alpha^n(\theta)) \leq d_1,
\]

and so, in particular, firm \( n \) is the most efficient firm at \( \alpha^n(\theta) \).

The proof follows directly from PP and NP (see Appendix A, Section 8.6). By Lemma 3, conditional on the effort level asked of \( \theta \), the match between the firm and the agent is efficient. Note, however, that it may well be that \( \arg \max_{n'} B^n'(a^n_*(\theta)) - c(a^n_*(\theta), \theta) \) is not equal to \( n \). There are thus three sources that pull the surplus of the agent down from the competitive equilibrium level. First, effort will typically be distorted from \( a^*_n(\theta) \). Second, the worker may be mismatched. Third, the firm to whom the worker is matched earns rents.

Let \( \delta = \max_{a,n,\theta} (c_{aa}(a, \theta) - B^n_{aa}(a)) \) bound the absolute value of \( B^n_{aa} - c_{aa} \).

**Theorem 2 (Limit Efficiency)** Let \( s \) be an equilibrium. Fix \( \theta \), and let \( n \) be the firm that serves \( \theta \). Then \( \pi^n(\theta, \alpha^n, v^n) \leq d_1 \), and \( v_*(\theta) - v^n(\theta) \leq d_1 + \frac{1}{2} d_2^2 \delta \).

Thus, as \( d_1 \to 0 \) and \( d_2 \to 0 \), the payoff to both firms and workers converges to the competitive limit. In the example above, \( \pi^n(\theta, \alpha^n, v^n) \leq 1/N^2 \), and \( v_*(\theta) - v^n(\theta) \leq (1 + \frac{1}{2} \delta)/N^2 \), and so convergence is fast.

\(^{12}\)For an example where this is false, start from a two-firm example, and then create \( N - 1 \) copies of Firm 1 while retaining a single copy of Firm 2. For an example with slow convergence spread \( \lfloor \sqrt{N} \rfloor \) firms out evenly as in Example 1, and make the remainder copies of Firm 1.
5.4 Who Does Asymmetric Information Help or Hurt?

Consider the version of our model without workers’ private information. Under monopoly the effect is clear: the firm is better off. It can undo any inefficiency, raising total surplus, and then extract all the surplus, as information rents disappear. The workers are clearly worse off.

In the oligopoly case, there is a third effect. With asymmetric information, worker $\theta$ might be hired by firm $n$, even though some other firm $n'$ could, by an appropriate choice of wage and effort, both attract and earn profits on $\theta$. What holds $n'$ back is that such an offer might also attract some of $n'$’s existing workers at lower profits. Without asymmetric information, there is no such cross-type constraint, making it easier to compete for a worker currently hired by a competitor. Hence, while $\theta$ no longer receives information rents, his outside option may have increased.

To examine this issue, note first that equilibrium with perfect information in the oligopoly case suffers from the classic discontinuity-at-ties problem. To sidestep this, let us assume that when the worker faces two offers giving him the same surplus, he chooses the firm that earns more surplus in hiring him. Then, competition at each type is Bertrand between differentiated firms, and so in equilibrium each worker will be hired by the firm that can use him best, effort will be efficient, and surplus will equal the surplus that the second most efficient firm for that type can provide. Since the allocation is efficient, matching will once again be positively assortative.

**Theorem 3 (Welfare)** Let $\hat{\theta}$ be on the boundary between the regions of types efficiently hired by two consecutive firms. Then, for all $\theta$ in some interval containing $\hat{\theta}$, $\theta$ is strictly worse off, and the firm hiring $\theta$ strictly better off, under asymmetric information than under full information.

The proof is simple. Since $\hat{\theta}$ can be efficiently hired by two firms, the Bertrand logic implies that in the full-information equilibrium, $\hat{\theta}$ earns the efficient surplus, and the firms earn zero. In the asymmetric information case, we have already proven that firms earn strictly positive profits, and so, since total surplus is at most the efficient surplus, the surplus of the worker must be strictly lower than in the full-information equilibrium. The argument is completed by noting that profits and surplus are continuous in type.

That is, contrary to the case of a monopoly firm, it is workers who are harmed by asymmetric information, and firms who are helped, at least over ranges of types near points where two firms can offer the efficient surplus. We do not have clear results or intuition for how the two forces—one in favor of more competition and the other against—balance outside of these ranges.

6 Existence and Sufficiency

To complete the analysis, we now turn to sufficiency and existence. We will provide a set of conditions for a strategy profile $s$ to be an equilibrium, and for an equilibrium to exist.
6.1 Stacking and Strict Regularity

The possibility of ties at the boundaries between players substantially complicates things. So, we begin by imposing some simplifying structure on the problem.

**Definition 1** Stacking is satisfied if for all \( n < N \), \( \gamma^{n+1}(\cdot, 1) > \gamma^n(\cdot, 0) \).

Under stacking, when each \( \kappa \) is restricted to lie in \([0, 1]\), the action schedule for player \( n + 1 \) always lies strictly above that of player \( n \). Stacking holds if firms are sufficiently differentiated, and simplifies our analysis since in any strategy profile that will be relevant to us, the surplus functions of adjacent players will cross strictly, precluding ties.\(^{13}\) We will henceforth impose stacking.

6.2 The Main Results

We will state the main results first, and then, in the next several subsections, discuss how to prove them. The relevant proofs are in Appendix B.

Fix \( n \) and \( s^{-n} \). Say that \( s^n = (\alpha^n, v^n) \) and \( \hat{s}^n = (\hat{\alpha}^n, \hat{v}^n) \) are essentially equivalent if \( \varphi(\cdot, (s^n, s^{-n})) = \varphi(\cdot, (\hat{s}^n, s^{-n})) \), and if anywhere that \( \varphi(\cdot, (s^n, s^{-n})) > 0 \), we have \( \alpha^n = \hat{\alpha}^n \) and \( v^n = \hat{v}^n \). That is, \( s^n \) and \( \hat{s}^n \) agree anywhere that is relevant given \( s^{-n} \). Two strategy profiles are essentially equivalent if they are essentially equivalent for each \( n \).

**Theorem 4 (Sufficiency)** Assume stacking. Then any strategy profile satisfying \( PS \), \( IO \), and \( OB \) is essentially equivalent to a Nash equilibrium.

This is non-trivial, because \( \Pi^n(\cdot, s) \) is not quasi-concave: if we fix \( s^{-n}, s^n, \) and \( \hat{s}^n \), then, a convex combination of \( s^n \) and \( \hat{s}^n \) will win a set of types different from either \( s^n \) or \( \hat{s}^n \), and so it unclear how its profits will relate to those of either \( s^n \) or \( \hat{s}^n \). But then satisfying the first-order conditions need not imply optimality.

**Theorem 5 (Existence)** Assume stacking. Then, a Nash equilibrium exists.

Existence is not trivial since \( \Pi(\cdot, s) \) is not continuous on \( S \). For example, let \( N = 2 \) and \( v^2 = v^1 + \varepsilon \). Then \( \varphi^2(\cdot, s) = 0 \) for all \( \varepsilon < 0 \), while \( \varphi^2(\cdot, s) = 1 \) for all \( \varepsilon > 0 \). Further, since \( \Pi \) is not quasi-concave, the set of best-responses may be non-convex.

\(^{13}\)If firms are not very differentiated, then equilibria must involve ties. To see this, consider \( N = 2 \), and assume that \( \gamma^2(\cdot, 1) < \gamma^1(\cdot, 0) \). Then, there must be a tie (with associated action equal to \( \hat{a}_1 \)), since if \( \theta^2_1 = \theta^1_0 \), then \( \alpha^1(\theta^1_0) = \gamma^1(\theta^1_0, 0) > \gamma^2(\theta^2_1, 1) = \alpha^2(\theta^2_1) \), contradicting \( PS \).
6.3 The Reformulation

Let us reformulate the problem of finding a best response. Fix \( n \) and \( s^{-n} \). Strategy \( s^n \) is dominant on \((\tau_l, \tau_h)\) if \((\tau_l, \tau_h)\) is a maximal interval such that \( v^n > v^{-n} \). Say that \( s^n \) is single dominant on \((\tau_l, \tau_h)\) if in addition \( v^n < v^{-n} \) for \( \theta \notin [\tau_l, \tau_h] \). That is, Firm \( n \) wins with probability one on \((\tau_l, \tau_h)\), and probability zero outside of \([\tau_l, \tau_h]\).

The first key step is to show that if the other firms are doing something “reasonable,” then the firm can optimize over single-dominant strategies that are of the \( \gamma \) form, with \( \kappa \in [0, 1] \). To formalize “reasonable” note first that while the convex combination of two \( \gamma \) strategies each with \( \kappa \in [0, 1] \) need not be a \( \gamma \) strategy, it will always satisfy the following condition.

**C1** \( \alpha^n \) is continuous, with \( \alpha^n(\theta) \in [\gamma^n(\theta, 1), \gamma^n(\theta, 0)] \) for all \( \theta \).

Given Proposition \( \boxed{3} \) in the Appendix, it is also innocuous to assume that firms never offer a surplus above \( v^n_* \), the most surplus they can offer without losing money.

**C2** \( v^n \leq v^n_* \).

By C2 and \( \boxed{1} \) it follows that \( n \), in any best response to \( s^{-n} \), earns positive profits.

**Lemma 4** Assume stacking, let \( s \) satisfy C1, and assume that \( n \) sometimes wins. Then, \( s^n \) is single dominant on some non-empty interval, and if \( s^n \) satisfies OB, it satisfies NP as well.

The first result follows since by C1 and stacking, \( n \)'s actions are always strictly higher than any action of lower indexed competitors and strictly lower than those of higher indexed competitors, and so \( v^n \) can only cross \( v^{-n} \) twice, and does so strictly.

The second result follows since \( a^{-n} \) is above the efficient level for \( n \) to the right of \( \theta_h \), and hence by \( \boxed{11} \), the profits to poaching are decreasing. And, we show that near \( \theta_h \), \( \boxed{10} \) implies that \( n \) prefers to gain an extra worker by moving \( \theta_h \) than by poaching. Similarly for \( \theta < \theta_l \).

**Corollary 1** Under stacking, any equilibrium has SPS.

This follows immediately since we have already shown that in any equilibrium, all players sometimes win, and that their strategies are of the \( \gamma \) form with \( \kappa \in [0, 1] \), and hence satisfy C1.

6.4 Relating the Original and the Relaxed Problem

Our goal is to move the analysis of \( n \)'s problem from the original infinite-dimensional problem to the two-dimensional–and hence relatively tractable–problem of maximizing \( r \). To do so, we need first to relate \( r \) to \( \Pi \). This is accomplished in the next three claims. Recall that \( \tilde{s}(\theta_l, \theta_h) \) is the solution to the relaxed problem \( \mathcal{P}(\theta_l, \theta_h) \), and let \( r(\theta_l, \theta_h) \) be the resulting value.
Lemma 5 Assume stacking. Fix n, and let $s^n$ satisfy C1 and C2. Then r has a maximum, and at any maximum $(\theta_l, \theta_h)$ of r,

(i) $\tilde{s}(\theta_l, \theta_h) \in S$,
(ii) if $\theta_l > \theta$, then $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$ and if $\theta_h < \theta$, then $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$, and
(iii) $\tilde{s}(\theta_l, \theta_h)$ is single dominant on $(\theta_l, \theta_h)$ with $\Pi(\tilde{s}(\theta_l, \theta_h), s^{-n}) = r(\theta_l, \theta_h)$.

A maximum exists since r is continuous on the compact set $\{\theta_l, \theta_h | \theta \leq \theta_l \leq \theta_h \leq \theta\}$. Part (i) follows since $\tilde{s}(\theta_l, \theta_h) \in [0, 1]$, so that $\gamma(\cdot, \tilde{s})$ is increasing. To show (ii), note first that C1 and stacking imply that $\tilde{v}$ crosses $v^{-n}$ at most once and from below where the opponent is $n' < n$, and at most once and from above where the opponent is $n' > n$. The key to the proof is to show that if for example $\tilde{v}(\theta_h) > v^{-n}(\theta_h)$, then either $\theta_h = \tilde{\theta}$, in which case single-dominance again holds, or $r_{\theta_h}(\theta_l, \theta_h) > 0$, which violates that $\theta_h$ is optimal. Part (iii) follows immediately from (ii).

Proposition 2 Assume stacking. Fix n and $s^{-n}$ satisfying C1 and C2. Then, for each $\tilde{s}$ there is $(\theta_l, \theta_h)$ with $\Pi(\tilde{s}, s^{-n}) = r(\theta_l, \theta_h)$.

Before we discuss the proof, we note that Lemma 5 and Proposition 2 between them justify the desired reparameterization:

Corollary 2 Assume stacking. Fix n and $s^{-n}$ satisfying C1 and C2. Then, $\tilde{s}$ is a maximum of $\Pi(\cdot, s^{-n})$ if and only if $\tilde{s} = \tilde{s}(\theta_l, \theta_h)$, where $(\theta_l, \theta_h)$ maximizes r.

The proof of Proposition 2 uses two lemmas. Lemma 9 (Section 9.2) shows that there is an interval $[\underline{m}, \overline{m}]$ of types such that n makes money imitating his opponent if and only if $\theta \in [\underline{m}, \overline{m}]$. This follows from (11) since by C1 and stacking, $a^{-n}$ is first strictly below n’s efficient action level and then strictly above, and so the profits to imitation are single-peaked.

Assume that $v$ is dominant on some interval $(\tau_l, \tau_h)$. Then (Lemma 10) $(\tau_l, \tau_h)$ and $[\underline{m}, \overline{m}]$ overlap. For intuition, assume that $\tau_l \geq \overline{m}$, so that $v(\tau_l) = v^{-n}(\tau_l)$. Then, by the definition of dominance, for some $\tau$ just to the right of $\tau_l$, we must have $\alpha > a^{-n}$ and $v > v^{-n}$. At such a point, since $a^{-n}$ is already inefficiently high for n, n must be losing money, contradicting Proposition 3.

Armed with these two lemmas, to prove the proposition, let $\overline{m}^* \geq \overline{m}$ capture any region of dominance of $v$ that contains $\overline{m}$, and let $\underline{m}^* \leq \underline{m}$ similarly capture any region of dominance of $v$ that contains $\underline{m}$. Relative to $\tilde{s}$, the firm strictly benefits by removing any worker outside of $[\underline{m}^*, \overline{m}^*]$, and adding any worker in $(\underline{m}, \overline{m})$ that it does not already hire with probability one. But, $\tilde{s}(\underline{m}^*, \overline{m}^*)$ does exactly this, and does so optimally in the relaxed problem, and hence its associated payoff $r(\underline{m}^*, \overline{m}^*)$ is at least as high as $\Pi(s^n, s^{-n})$. 

22
6.5 Unique Best Responses

In this section, we discuss the building-block results we will use to prove sufficiency and existence. We will begin by showing that \( r \) is sufficiently well-behaved that it has a unique maximum for given \( s^{-n} \) satisfying \( C1 \) and \( C2 \), and that any critical point of \( r \) is that maximum.

We face three challenges. First, \( v^{-n} \) has a kink point at each \( \theta \) where the relevant opponent changes, and hence \( r \) has kinks wherever either the top or bottom opponent changes. Second, \( r \) can have troughs and so single-peakedness fails, complicating a proof of uniqueness. Finally, because our choice set is two dimensional, it is not obvious that single-peakedness alone is enough.

Note that by \( C1 \), each kink point of \( v^{-n} \) is a point at which one transitions from one opponent to the next, and hence there are at most \( N - 1 \) such points. Let \( \vec{R} = [\iota_l, \iota_h] \times [\iota'_l, \iota'_h] \) be a maximal rectangle with the property that the opponent on \( (\iota_l, \iota_h) \) is constant, the opponent on \( (\iota'_l, \iota'_h) \) is constant, and \( \iota_h \leq \iota'_l \). Using \( C1 \), \( v^{-n} \) is continuously differentiable on \( \vec{R} \), with kinks in \( r \) constrained to the boundaries between rectangles.

We begin by showing any optimum of \( r \) is in the rectangle \( R = [\theta, \theta^x] \times [\theta^x, \bar{\theta}] \) illustrated in Fig. 3. Let the locus \( L_N \) be defined by \( z(\theta_l, \theta_h, H(\theta_l)) = 0 \), and the locus \( L_S \) be defined by
\[ z(\theta_l, \theta_h, H(\theta_h)) = 0. \] These are the north and south boundaries of the set

\[ \Theta = \{ (\theta_l, \theta_h) \in R | z(\theta_l, \theta_h, \tilde{\kappa}(\theta_l, \theta_h)) = 0 \}. \]

Assume first that \( L_S \) hits the western boundary of \( R \), let \( \theta_T \leq \tilde{\theta} \) be the latitude at which \( L_N \) hits the boundary of \( R \), and let \( A \) be the (possibly empty) segment of the western boundary of \( R \) above \( \theta_T \). Using Lemmas \[ 1 \] and \[ 5 \] we show that any maximum of \( r \) occurs either in \( \Theta \), with both the utility constraints \[ 5 \] and \[ 6 \] binding, or in \( A \), with the utility constraint at \( \theta \) slack.

Next we show (Lemma \[ 14 \]) that, on any given \( \tilde{R} \cap \Theta \), if \( r_{\theta_l} = 0 \) then \( r \) is locally strictly concave in \( \theta_l \). Similarly, if \( r_{\theta_h} = 0 \), \( r \) is locally strictly concave in \( \theta_h \), and anywhere that \( r_{\theta_l} = r_{\theta_h} = 0 \), \( r \) is locally strictly concave in \((\theta_l, \theta_h)\). Some intuition comes from \[ 25 \], where we show that after some cancelations (for which it is very convenient that \( c_{a\theta} = -1 \)), \( r_{\theta_h, \theta_h} \) has the same sign as

\[ \pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) (\gamma(\theta_h, \tilde{\kappa}))_{\theta_h} (a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})) + \pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) a_{\theta}^{-n}(\theta_h), \]

both terms of which are negative at an optimum. The first term reflects that as \( \theta_h \) increases, effort is distorted further above the efficient level, while the second term reflects that as the action of the opponent gets steeper, the rate at which the firm must distort effort to move \( \theta_h \) increases.

The proof from here follows the topographical intuition from the introduction. For each \( \theta_h \), let \( \Theta(\theta_h) \) be the interval of \( \theta_l \) such that \( (\theta_l, \theta_h) \in \Theta \cup A \), so that for \( \theta'_h > \theta_T \), \( \Theta(\theta'_h) = \{ \theta \} \). Define

\[ \psi(\theta_h) = \max_{\theta_l \in \Theta(\theta_h)} r(\theta_l, \theta_h), \]

so that we begin by maximizing \( r \) moving east-west. Let \( D \) be the set of \( \theta_h \) such that \( \psi > 0 \). Fix \( \theta_h \in D \) with \( \theta_h < \theta_T \). In Lemma \[ 17 \] we show that \( r(\cdot, \theta_h) \) is strictly single-peaked where it is positive and has a unique maximum \( \lambda(\theta_h) \). One implication of this is that any local minima are under water. The proof rests on Lemma \[ 14 \] but accounts for the fact that our terrain is kinked at the boundaries where an opponent changes.

The function \( \lambda \) is the path described in the introduction. We show (Lemma \[ 20 \]) that \( \lambda \) is continuous, and hence so is \( \psi \). We also show (Lemma \[ 19 \]) that \( D \) is an interval. We show that \( \lambda \) never runs along \( L_N \), because at any point in \( L_N \), profits are strictly decreasing in \( \theta_l \). The path may run along \( L_S \), but we show that where \( \lambda \) is on \( L_S \), \( \psi \) is strictly increasing, where the intuition is that on \( L_S \), the firm is better off to strictly increase \( \theta_h \), and is also benefited by the fact that \( L_S \) is less binding as \( \theta_h \) increases. So, consider any \( \hat{\theta}_h \) such that \( \lambda(\hat{\theta}_h) \) is in the interior of \( \Theta(\hat{\theta}_h) \). We show (Lemma \[ 22 \]) that the left and right derivatives of \( \psi \) at \( \hat{\theta}_h \) and the left and right partial derivatives of \( r \) with respect to \( \theta_h \) at \((\lambda(\hat{\theta}_h), \hat{\theta}_h)\) agree. Given that \( \lambda(\hat{\theta}_h) \) maximizes \( r(\cdot, \theta_h) \), this follows from the Envelope Theorem. The proof again deals with kinks in \( v^{-n} \) at either \( \theta_h \) or \( \lambda(\theta_h) \).

Using Lemma \[ 19 \] and Lemma \[ 22 \], we show (Lemma \[ 23 \]) that \( \psi \) is strictly single-peaked—and
thus has a unique maximum—on the interval $D$, which is to say, as one walks northward along the path. This uses the concavity properties already established for $r$, with the usual complexities at kink points. Finally, we show (Lemma 24) that if $\theta^*_h$ is the unique maximizer of $\psi$, then $(\lambda(\theta^*_h), \theta^*_h)$ is the unique maximizer of $r$.

Assume that instead of hitting $R$’s western boundary, $L_S$ instead hits $R$’s northern boundary at $(\tilde{\theta}_T, \bar{\theta})$. Then, we can argue as before that any optimum of $r$ occurs either in $\Theta$, with both constraints binding, or on the segment of the northern boundary of $R$ with $\theta_l \leq \tilde{\theta}_T$ with the constraint at $\bar{\theta}$ slack. We can thus perform the same analysis as above, but exchange the roles of $\theta_l$ and $\theta_h$, so that one defines $\tilde{\lambda}(\theta_l)$ by first maximizing along north-south slices where $\theta_l$ is held constant, and then walks eastward along the path defined by $\tilde{\lambda}$.

### 6.6 Sufficiency

Let us now discuss sufficiency. Fix $\hat{s}$ satisfying $PS$, $IO$, and $OB$. We wish to show that there is a strategy profile $s$ that is essentially equivalent to $\hat{s}$ and is a Nash equilibrium. The key is that $PS$, $IO$ and $OB$ imply that, for each $n$, $\hat{s}^n$ corresponds to a critical point of $r$.

The first step is to modify each $\hat{s}^n$ outside of $[\theta_l, \theta_h]$ so as to satisfy $C_1$ and $C_2$ there as well, so that the results of the previous section apply. We do this while maintaining continuity of actions, and hence, $OB$ is unaffected.

Let $s$ be the strategy profile constructed in this way. Consider first $n \notin \{1, N\}$. Since $n$ earns positive profits, and since by $IO$, the associated $\kappa$ is in $(H(\theta_l), H(\theta_h))$, Lemma 17 applies and $r(\cdot, \theta_h)$ is single-peaked where it is positive. But then, since $\partial_\theta r(\theta_l, \theta_h) = 0$ by $OB$, we must have $\theta_l = \lambda(\theta_h)$, and so, $\partial_\theta r(\lambda(\theta_h), \theta_h) = \psi_\theta(\theta_h) = 0$ by Lemma 22. Thus by Lemma 23, $\theta_h = \theta^*_h$, and so $s^n$ is a best response to $s^{−n}$. The argument for $n = 1$ and $N$ is similar.

### 6.7 Existence

Let us turn to existence. Recall that in general, $\Pi$ can be discontinuous, and that there is no reason to believe that the set of best responses is convex. Our plan is to restrict the strategy space so that continuity and convexity of best-responses hold, and to show that the equilibrium of the restricted game is an equilibrium of the original game.

To begin, we need a convex and compact set of strategies. Let $\eta$ be a bound on both the slope and value of any $\gamma$ strategy with $\kappa$ in $[0, 1]$. Impose (C3) that action profiles have slope bounded by $\eta$. Choose $\beta$ small enough that if surplus at $\bar{\theta}$ is strictly less than $\beta$, then (10) is guaranteed to fail, and impose (C4) that surplus functions give surplus at least $\beta$ at $\bar{\theta}$. Finally (C5), impose that $v$ has the standard integral form. Let $S^n_R$ be the subset of $S^n$ such that C1–C5 hold, with $S_R$ and $S^n_R$ defined in the usual way.

We first show (Lemma 25) that if other firms choose from $S^n_R$, then Firm $n$ has a best response
in $S_R^n$. The idea is that every best response is a $\gamma$ strategy where it is single-dominant, and that $\eta$ and $\beta$ were chosen to not bind for such strategies, so that $C3$ and $C4$ are non-binding. Satisfying $C1$ and $C2$ involves inessential modification of the strategy outside of $[\theta_l, \theta_h]$.

Given Lemma 25, it is enough to show that $(S^n_R, \Pi^n)$ has an equilibrium. We first establish that $S^n_R$, and hence $S_R$, is a Banach space with a norm yielding continuous payoffs. The key to continuity is that $C1$, $C2$, and stacking imply that $v^n$ and $v^{n+1}$ strictly single cross, and hence boundaries move continuously as strategies vary. Compactness and convexity follow since the relevant action profiles are equicontinuous by $C3$, and since $C1$–$C4$ can be phrased as a collection of weak inequalities. Since payoffs are continuous on $S_R$, $BR^n_R$ has a closed and non-empty valued graph. Finally, to show that $BR^n_R(s^{-n})$ is convex, observe that for any $s^{-n}$, Section 6.5 implies that any two best responses are essentially equivalent. But then, their convex combination is essentially equivalent to either of them, and so is also a best response. We thus have all the conditions to apply the Kakutani-Fan-Glicksberg Theorem, and hence a Nash equilibrium exists.

7 Conclusion

We analyze an oligopoly market with heterogeneous vertically-differentiated firms and workers with privately known ability. The model is a natural extension to an oligopolistic setting of the ubiquitous principal-agent problem in Mussa and Rosen (1978) and Maskin and Riley (1984). Firms post menus to both screen workers and attract the right pool of applicants. Our analysis uncovers several insights regarding sorting, distortions, and gaps in productivity across firms. We examine the model’s competitive limit. Contrary to the monopoly model, asymmetric information can help firms and hurt workers. Finally, we show that under enough firm heterogeneity these conditions are sufficient for a strategy profile to be an equilibrium, and that an equilibrium exists.

There are many potential extensions of our analysis that are worth pursuing, some for completeness and some more drastic. First is to allow for more general disutility of effort. We conjecture that this will primarily present technical complications. Second is to extend the existence and sufficiency results to the case where firms are less vertically differentiated, so that stacking does not hold. Our existing proof relied hard on stacking to establish continuity. Third is to extend the model to allow both horizontal and vertical differentiation. Finally, a pressing but challenging extension is to allow for common values and risk-averse workers, so as to apply the framework to insurance markets and also to incorporate moral hazard in a nontrivial way.
8 Appendix A: Proofs For Sections 4–5

8.1 Proof of PP

We begin with a preliminary result. It shows that there is zero probability that a firm hires a worker on whom it strictly loses money, and that among each firm’s best responses is always a menu in which no offer, whether accepted with positive probability or not, strictly loses money.

**Proposition 3** Fix \( n, s^{-n}, \text{and} \ s^n = (\alpha, v) \). Let \( P \equiv \{ \theta | \pi(\theta, \alpha, v) \geq 0 \} \). Then, there is \((\hat{\alpha}, \hat{v})\) with \( \pi(\cdot, \hat{\alpha}, \hat{v}) \geq 0 \) that agrees on \( P \) with \((\alpha, v)\). If \((\alpha, v)\) is a best response, then \( \pi(\theta, \alpha, v) \geq 0 \) for almost all \( \theta \) where \( \varphi > 0 \).

**Proof** Fix \( n \), and let \( G(\theta, v) \) be the subdifferential to \( v \) at \( \theta \). Since \( v \) is convex, \( G \) is singleton-valued almost everywhere, and every selection from \( G \) is increasing. Thus, since \( G \) is upper hemicontinuous in \( \theta \) and compact-valued, it is wlog to assume that \( \alpha(\theta) \in \arg \max_{a \in G(\theta, v)} \pi(\theta, a, v) \) for all \( \theta \). But then, \( \pi(\cdot, \alpha(\cdot), v) \) is upper semicontinuous (Aliprantis and Border (2006), Lemma 17.30, p. 569), and so \( P \equiv \{ \theta | \pi(\theta, \alpha, v) \geq 0 \} \) is closed.

For each \( \theta' \in [\theta, \bar{\theta}] \), let \( v_L(\theta, \theta') = v(\theta') + (\theta - \theta')\alpha(\theta') \). Note that \( v_L(\theta, \theta) = v(\theta) \), and that

\[
(\pi(\theta, \alpha(\theta'), v_L(\theta, \theta')))_\theta = (B(\alpha(\theta')) - c(\alpha(\theta'), \theta) - v_L(\theta, \theta'))_\theta = \alpha(\theta') - \alpha(\theta') = 0,
\]

using \( c(\alpha, \theta) = -a \). Hence, along \( v_L(\cdot, \theta') \), the profits to the firm are constant.

If \( P \) is empty, set \((\alpha, v) = (\alpha_*, v_*)\), and we are done. If \( P \) is non-empty, define

\[
\hat{v}(\theta) = \max_{\theta' \in P} v_L(\theta, \theta').
\]

Then, \( \hat{v} \) is convex, with \( \hat{v} = v \) on \( P \) and \( \hat{v} \leq v \). Let \( \hat{\alpha} \) be a selection from \( G(\cdot, \hat{v}) \), where we can take \( \hat{\alpha} = \alpha \) on \( P \), and where at any \( \theta \notin P \), we can take \( \hat{\alpha}(\theta) = \alpha(\theta') \) for some \( \theta' \in \arg \max_{\theta' \in P} v_L(\theta, \theta') \). Then by using \((\hat{\alpha}, \hat{v})\), the firm earns the same on \( P \) as it used to, and earns positive profits on any other worker, since that worker is now imitating a worker in \( P \).

**Corollary 3** Each firm earns strictly positive profits in equilibrium.

**Proof** By assumption there is \( \theta^n_* \) such that \( v^n_*(\theta^n_*) > v^{-n}_*(\theta^n_*) \), and so, by continuity, \( v^n_*(\theta) > v^{-n}_*(\theta) \) for all \( \theta \) in some interval \( I \) around \( \theta^n_* \). Assume that on a positive measure set of \( I \), \( v^{-n}(\theta) \geq \)
Then, since \( v^*_n(\theta) > v^{*-n}_n(\theta) \) on \( I \), either some firm other than \( n \) is winning with positive probability and is losing money, or \( n \) is winning having offered surplus \( v^n(\theta) > v^{-n}(\theta) \geq v^*_n(\theta) \). (Note that if \( n \) offers \( v^*_n(\theta) \), then firms other than \( n \) win with positive probability since ties are broken equiprobably.) Either case violates Proposition 3. But then, for \( \varepsilon \) sufficiently small but positive, the strategy of offering all types surplus \( v^*_n(\theta) - \varepsilon \) and action \( a^*_n(\theta) \) earns at least \( \varepsilon \) on a positive measure set of types. Hence, \( n \) must earn strictly positive profits in equilibrium. \( \square \)

8.2 Proof of NP

We now establish property NP.

**Proposition 4** Let \((\alpha, v)\) be optimal, and for each \( \omega \geq 0 \), let \( \Theta_{\omega} = \{\theta | \pi(\theta, a^{-n}, v^{-n}) > \omega\} \). Then,
\[
\int_{\Theta_0} (1 - \varphi(\theta, s)) h(\theta) d\theta = 0,
\]
so that \( n \) wins with probability one anywhere she could strictly profitably imitate the incumbent.

**Proof** Assume the lemma is false. Then, there exists \( \omega > 0 \) such that
\[
\int_{\Theta_{\omega}} (1 - \varphi(\theta, s)) h(\theta) d\theta > 0.
\]
Let
\[
\hat{\Theta}_{\omega} \equiv \Theta_{\omega} \cap \{\theta | \varphi(\theta, s) < 1\},
\]
and for each \( \varepsilon > 0 \), let
\[
v_{\varepsilon}(\theta) = \max \left( v(\theta), \varepsilon + \sup_{\theta' \in \hat{\Theta}_{\omega}} (v^{-n}(\theta') + (\theta - \theta)a^{-n}(\theta')) \right) \leq \max (v(\theta), v^{-n}(\theta) + \varepsilon),
\]
where the inequality follows since \( v^{-n} \) is convex. Let \( \alpha_{\varepsilon} \) be an associated subgradient. In words, \( n \) adds to her menu all contracts that her opponents were offering on \( \hat{\Theta}_{\omega} \) but offers \( \varepsilon \) more surplus than the incumbent.

Since \( v_{\varepsilon} \) is the supremum of convex functions, it is convex, and so \( \alpha_{\varepsilon} \) is increasing. Consider any \( \theta \) at which \( \varphi(\theta, s) = 1 \). Then, \( v(\theta) > v^{-n}(\theta) \), and so for all \( \varepsilon \) sufficiently small, \( v_{\varepsilon}(\theta) = v(\theta) \), from which wlog, \( \alpha_{\varepsilon}(\theta) = \alpha(\theta) \), and hence \( \pi(\theta, \alpha_{\varepsilon}, v_{\varepsilon}) = \pi(\theta, \alpha, v) \). The loss compared to \( \pi(\theta, \alpha, v) \) where \( \varphi = 1 \) under \( v \) but \( v_{\varepsilon} \neq v \) is bounded, and the probability of such \( \theta \) goes to 0 as \( \varepsilon \to 0 \).

Note next that for \( \theta \in \hat{\Theta}_{\omega} \), \( v_{\varepsilon}(\theta) = v^{-n}(\theta) + \varepsilon \), since \( v(\theta) < v^{-n}(\theta) + \varepsilon \), and since \( v^{-n} \) is convex. But then, we can take \( \alpha_{\varepsilon}(\theta) = a^{-n}(\theta) \), and so the firm earns at least \( \omega - \varepsilon \) from hiring \( \theta \). Since \( \varphi \) is raised from at most \( 1/2 \) to 1, as \( \varepsilon \to 0 \), the total profit from this deviation thus approaches at least \( (\omega/2) \int_{\hat{\Theta}_{\omega}} (1 - \varphi(\theta)) h(\theta) d\theta > 0 \), contradicting optimality. \( \square \)
8.3 Proof of $PS$

Let us first prove that any Nash equilibrium (with or without $NEO$) satisfies a condition slightly weaker than $PS$. Say that $s$ has quasi-positive sorting ($QPS$) if it satisfies the conditions for $PS$ except that each condition on $\varphi$ is allowed to fail on a zero-measure subset.

**Proposition 5** Every Nash equilibrium has $QPS$.

**Proof** Let $n' > n$, let $\theta_{\text{inf}}^{n'}$ be the infimum of the support of $\varphi^{n'}$ and let $\theta_{\text{sup}}^{n}$ be the supremum of the support of $\varphi^n$. Assume that $\theta_{\text{inf}}^{n'} < \theta_{\text{sup}}^{n}$. Conditional on $\varphi^{n'}(\theta, s) > 0$, with probability one $\pi^{n'}(\theta, \alpha^{n'}, v^{n'}) \geq 0$ by Proposition 3 and $\pi^n(\theta, \alpha^{v^n}, v^{n'}) \leq 0$ by Lemma 4. Hence, for any $\varepsilon \in (0, (\theta_{\text{sup}}^{n} - \theta_{\text{inf}}^{n'})/2)$ there is $\theta_1 \in [\theta_{\text{inf}}^{n'}, \theta_{\text{inf}}^{n'} + \varepsilon]$ where $\varphi^{n'}(\theta_1) > 0$ and

$$\pi^{n'}(\theta_1, \alpha^{n'}, v^{n'}) \geq 0 \geq \pi^n(\theta_1, \alpha^{n'}, v^{n'}) \quad (12)$$

and similarly, there is $\theta_2 \in [\theta_{\text{inf}}^{n} - \varepsilon, \theta_{\text{inf}}^{n}]$ where $\varphi^n(\theta_2) > 0$ and

$$\pi^n(\theta_2, \alpha^{n}, v^{n'}) \geq 0 \geq \pi^{n'}(\theta_2, \alpha^{n}, v^{n'}) \quad (13)$$

By incentive compatibility, since $\theta_2 > \theta_1$ and since $\varphi^{n'}(\theta_1) > 0$ and $\varphi^n(\theta_2) > 0$, it must be that $\alpha^n(\theta_2) \geq \alpha^{n'}(\theta_1)$. Adding (12) and (13) and cancelling common terms,

$$B^{n'}(\alpha^{n'}(\theta_1)) + B^n(\alpha^n(\theta_2)) \geq B^n(\alpha^{n'}(\theta_1)) + B^{n'}(\alpha^n(\theta_2)).$$

Since $B^n(a)$ is strictly supermodular, $\alpha^{n'}(\theta_1) = \alpha^n(\theta_2) = \tilde{a}$, and so, by incentive compatibility, and since $\varepsilon$ was arbitrary, $\alpha^{n'}(\theta) = \alpha^n(\theta) = \tilde{a}$ for all $\theta \in (\theta_{\text{inf}}^{n'}, \theta_{\text{sup}}^{n})$. From (12), $B^{n'}(\tilde{a}) \geq B^n(\tilde{a})$, while from (13), $B^n(\tilde{a}) \leq B^n(\tilde{a})$, and so $B^n(\tilde{a}) = B^n(\tilde{a}) = b$. But then, from (12), $\pi^{n'}(\theta_1, \alpha^{n'}, v^{n'}) = 0$, and from (13), $\pi^n(\theta_2, \alpha^n, v^{n'}) = 0$. Finally, on $(\theta_{\text{inf}}^{n'}, \theta_{\text{sup}}^{n})$, $(\pi(\theta, \alpha, v))_\theta = \pi_\alpha(\alpha, \alpha, \alpha) = 0$, using $-c_\theta(\alpha(\theta), \theta) = \alpha(\theta) = v_\theta(\theta)$. Hence $\pi^n = \pi^{n'} = 0$ on $(\theta_{\text{inf}}^{n'}, \theta_{\text{sup}}^{n})$.

Assume that $n' \neq n + 1$, and let $n < n'' < n'$. Assume first that $B^{n''}(\tilde{a}) \leq b = B^n(\tilde{a})$. Then since $n'' > n$ and $B^n(a)$ is strictly supermodular, $B^{n''}(a) < B^n(a)$ for all $a < \tilde{a}$, and similarly, $B^{n''}(a) < B^n(a)$ for all $a > \tilde{a}$, contradicting that $B^{n''}$ is somewhere uniquely maximal. Thus $B^{n''}(\tilde{a}) > b$, and so $\pi^{n''}(\tilde{a}, v^{n''}) > 0$ on $(\theta_{\text{inf}}^{n'}, \theta_{\text{sup}}^{n})$, which contradicts Lemma 4 since by definition of $\theta_{\text{inf}}^{n'}$ and $\theta_{\text{inf}}^{n}$, $\varphi^{n''} < 1$ on $(\theta_{\text{inf}}^{n'}, \theta_{\text{inf}}^{n})$. Thus, $n'' = n + 1$, and $\tilde{a} = \hat{a}^n$. Letting $\theta_h^n = \theta_{\text{inf}}^{n'}$ and $\theta_l^{n+1} = \theta_{\text{sup}}^{n}$, we have the claimed structure at ties.

Finally, it must be that $\theta_l^n < \theta_h^n$, since by Corollary 3 $n$ earns strictly positive expected profit, but on each type above $\theta_h^n$ and below $\theta_l^n$ either loses for sure or ties but earns 0. □

**Corollary 4** Every Nash Equilibrium that satisfies NEO has $PS$. 
Proof Assume that for some \( n' > n \), and for some \( \hat{\theta} \in (\theta_l, \theta_h) \), \( v^{n'} = v^n \). Then, since by NEO, \( \alpha^{n'} \geq \alpha^n \), and hence \( v^{n'}(\theta) - v^n(\theta) \) is increasing, \( v^{n'} \geq v^n \) everywhere on \([\hat{\theta}, \theta_h]\), contradicting that \( n \) wins with probability one conditional on \( \theta \in (\theta_l, \theta_h) \).

Lemma 6 Fix \( n, s^{-n} \), and \( \hat{s} = (\hat{\alpha}, \hat{v}) \). If \( \hat{s} \) is a best-response, then \( \hat{\alpha} \) must be continuous on any open interval where \( v^n \geq v^{-n} \).

Proof Consider some maximal interval \([\tau_l, \tau_h] \) on which \( \hat{v} \geq v^{-n} \), and in which there is a jump point \( \theta_j \in (\tau_l, \tau_h) \) where \( a \equiv \lim_{\theta \uparrow \theta_j} \hat{\alpha}(\theta) < \bar{a} \equiv \lim_{\theta \downarrow \theta_j} \hat{\alpha}(\theta) \). Let us first show that \( (\hat{\alpha}, \hat{v}) \) is sub-optimal. Fix \( 0 < q < \min(\theta_j - \tau_l, \tau_h - \theta_j) \), and let

\[
\alpha_1(\theta) = \begin{cases} 
\hat{\alpha}(\theta) & \theta \notin [\theta_j - \epsilon, \theta_j + \epsilon] \\
\hat{\alpha}(\theta) + q & \theta \in [\theta_j - \epsilon, \theta_j] \\
\hat{\alpha}(\theta) - q & \theta \in (\theta_j, \theta_j + \epsilon]
\end{cases}
\]

and let

\[
v_1(\theta) = \begin{cases} 
\hat{v}(\theta) & \theta \notin [\theta_j - \epsilon, \theta_j + \epsilon] \\
\hat{v}(\theta) + q(\theta - \theta_j + \epsilon) & \theta \in [\theta_j - \epsilon, \theta_j] \\
\hat{v}(\theta) + q(\theta - \theta_j + \epsilon) & \theta \in (\theta_j, \theta_j + \epsilon]
\end{cases}
\]

where we note that \( v_1 \geq \hat{v} \), and that \( v_1 \) has slope \( \alpha_1 \) almost everywhere. Thus, since \( \pi(\cdot, \hat{\alpha}, \hat{v}) \geq 0 \),

\[
\Pi^n((\alpha_1, v_1), s^{-n}) - \Pi^n(s) = \int_{\theta_j - \epsilon}^{\theta_j + \epsilon} (\pi(\theta, \alpha_1, v_1)\varphi(\theta, ((\alpha_1, v_1), s^{-n})) - \pi(\theta, \hat{\alpha}, \hat{v})\varphi(\theta, ((\hat{\alpha}, \hat{v}), s^{-n})))h(\theta)d\theta
\]

\[
\geq \int_{\theta_j - \epsilon}^{\theta_j + \epsilon} (\pi(\theta, \alpha_1, v_1) - \pi(\theta, \hat{\alpha}, \hat{v}))h(\theta)d\theta
\]

\[
= j(\epsilon)
\]

where the inequality uses that \( \varphi(\theta, ((\alpha_1, v_1), s^{-n})) = 1 \) since on \( (\theta_j - \epsilon, \theta_j + \epsilon) \), \( v_1 > \hat{v} \geq v^{-n} \), and that \( \pi(\cdot, \hat{\alpha}, \hat{v}) \geq 0 \).

Now,

\[
\ j_\epsilon(\epsilon) = (\pi(\theta_j - \epsilon, \hat{\alpha}(\theta_j - \epsilon) + q, \hat{v}(\theta_j - \epsilon)) - \pi(\theta_j - \epsilon, \hat{\alpha}, \hat{v}))h(\theta_j - \epsilon) - q \int_{\theta_j - \epsilon}^{\theta_j} h(\theta)d\theta
\]

\[
= (\pi(\theta_j + \epsilon, \hat{\alpha}(\theta_j + \epsilon) - q, \hat{v}(\theta_j + \epsilon)) - \pi(\theta_j + \epsilon, \hat{\alpha}, \hat{v}))h(\theta_j + \epsilon) - q \int_{\theta_j}^{\theta_j + \epsilon} h(\theta)d\theta,
\]

and thus, taking limits as \( \epsilon \to 0 \),

\[
\ j_\epsilon(0) = ((\pi(\theta_j, a + q, \hat{v}) - \pi(\theta_j, a, \hat{v})) - (\pi(\theta_j, a, \hat{v}) - \pi(\theta_j, a - q, \hat{v})))h(\theta_j) > 0,
\]
since \( \pi \) is strictly concave in \( a \), and since \( a + q < \bar{a} - q \) by definition of \( q \). Thus, since \( j(0) = 0 \), for small \( \varepsilon \), we have that \( j(\varepsilon) > 0 \), and hence \( \Pi^n((\alpha_1, v_1), s^{-n}) - \Pi^n(s) > 0 \). Thus, \( \hat{s} \) is sub-optimal, proving that no such jump point can exist if \( \hat{s} \) is a best response. \( \square \)

8.4 Proof of IO

We begin with two preliminary lemmas.

**Lemma 7** Let \( \kappa \in [0, 1] \). Then,

\[
\left( \frac{\kappa - H(\theta)}{h(\theta)} \right)_{\theta} = -1 - \left( \frac{(\kappa - H(\theta)) h'(\theta)}{h^2(\theta)} \right) \leq 0. \tag{14}
\]

**Proof** Assume first that \( h'(\theta) \leq 0 \). Then,

\[
\left( \frac{(\kappa - H(\theta)) h'(\theta)}{h^2(\theta)} \right) \geq \frac{(1 - H(\theta)) h'(\theta)}{h^2(\theta)} \geq -1,
\]

since \( 1 - H \) is log-concave. And if \( h'(\theta) > 0 \), then the result follows since \( H \) is log-concave. \( \square \)

**Lemma 8** Fix \( n \), and for any feasible \( \alpha \) and \( v \), define

\[
M(\theta, \alpha, v) = B(\alpha(\theta)) - c(\alpha(\theta), \theta) - v(\theta_l) - \alpha(\theta) \frac{H(\theta_h) - H(\theta)}{h(\theta)}.
\]

Then,

\[
\int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v) h(\theta) d\theta = \int_{\theta_l}^{\theta_h} M(\theta, \alpha, v) h(\theta) d\theta. \tag{15}
\]

**Proof** Note first that for any \( \alpha \) and \( v \),

\[
\int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v) h(\theta) d\theta = \int_{\theta_l}^{\theta_h} \left( B(\alpha(\theta)) - c(\alpha(\theta), \theta) - v(\theta_l) - \int_{\theta_l}^{\theta} \alpha(\tau) d\tau \right) h(\theta) d\theta,
\]

and that, integrating by parts,

\[
\int_{\theta_l}^{\theta_h} \left( \int_{\theta_l}^{\theta} \alpha(\tau) d\tau \right) h(\theta) d\theta = \int_{\theta_l}^{\theta_h} \alpha(\theta)(H(\theta_h) - H(\theta)) d\theta.
\]

Substituting and rearranging yields (15). \( \square \)

We now prove Lemma [1], which derives the solution to the relaxed problem \( \mathcal{P}(\theta_l, \theta_h) \).

**Proof of Lemma [1]** Existence is standard and uniqueness follows since the set of feasible strategies is convex, and the objective function is strictly concave. Fix \( (\theta_l, \theta_h) \), fix the optimum \( (\hat{\alpha}, \hat{v}) \),
and define
\[
\xi(\theta) = \pi_a(\theta, \hat{\alpha}, \hat{\nu}) h(\theta) + H(\theta).
\]
Let us first show that \(\xi\) is constant on \((\theta_1, \theta_h)\). In particular, choose any two points \(\theta'' > \theta'\) in \((\theta_1, \theta_h)\), and fix \(0 < \varepsilon < \frac{1}{2} \min \{\theta'' - \theta', \theta' - \theta_1, \theta_h - \theta''\}\). Define \(\hat{\alpha}(\cdot, y)\) to be \(\hat{\alpha} - y/2\varepsilon\) on \([\theta' - \varepsilon, \theta' + \varepsilon]\), \(\hat{\alpha} + y/2\varepsilon\) on \([\theta'' - \varepsilon, \theta'' + \varepsilon]\), and \(\hat{\alpha}\) elsewhere, and define
\[
\hat{\nu}(\theta, y) = \hat{\nu}(\theta_1) + \int_{\theta_1}^\theta \hat{\alpha}(\tau, y) d\tau,
\]
noting that \(\hat{\nu}(\theta_h, y) = \hat{\nu}(\theta_h)\), and so for each \(y\), \(\hat{s}(y) = (\hat{\alpha}(\cdot, y), \hat{\nu}(\cdot, y))\) is feasible in \(P(\theta_1, \theta_h)\).

Let profits as a function of \(y\) be \(j(y) = \int_{\theta_1}^{\theta_h} \pi(\theta, \hat{s}(y)) h(\theta) d\theta\). Then, since \(\pi_v = -1\),
\[
j_y(y) = \int_{\theta_1}^{\theta_h} (-\pi_a(\theta, \hat{s}(y)) \frac{1}{2\varepsilon} - \hat{\nu}(\theta, y)) h(\theta) d\theta + \int_{\theta_1}^{\theta_h} \hat{\nu}(\theta, y) h(\theta) d\theta
\]
\[
+ \int_{\theta_1}^{\theta_h} (\pi_a(\hat{\theta}, \hat{s}(y)) \frac{1}{2\varepsilon} - \hat{\nu}(\hat{\theta}, y)) h(\hat{\theta}) d\hat{\theta},
\]
where between \(\theta' + \varepsilon\) and \(\theta'' - \varepsilon\) we use \(\hat{s}_y = 0\) and \(\hat{\nu}_y = -1\). By the Mean Value Theorem, there is \(\tau' \in [\theta' - \varepsilon, \theta' + \varepsilon]\) and \(\tau'' \in [\theta'' - \varepsilon, \theta'' + \varepsilon]\) such that
\[
j_y(0) = 2\varepsilon \left( -\pi_a(\tau', \hat{s}(y)) \frac{1}{2\varepsilon} - \hat{\nu}(\tau', y) \right) h(\tau') + (H(\theta'' - \varepsilon) - H(\theta' + \varepsilon))
\]
\[
+ 2\varepsilon \left( \pi_a(\tau'', \hat{s}(y)) \frac{1}{2\varepsilon} - \hat{\nu}(\tau'', 0) \right) h(\tau'').
\]
But then, since \(\hat{\nu}(\tau', 0)\) and \(\hat{\nu}(\tau'', 0)\) are bounded,
\[
\lim_{\varepsilon \to 0} j_y(0) = -\pi_a(\theta', \hat{s}) h(\theta') + H(\theta'') - H(\theta') + \pi_a(\theta'', \hat{s}) h(\theta'')
\]
\[
= \xi(\theta'') - \xi(\theta'),
\]
and so, if \(\xi(\theta'') - \xi(\theta') \neq 0\), then for \(\varepsilon\) sufficiently small, \(j_y(0) \neq 0\), and the firm has a profitable deviation, a contradiction. Thus, setting \(\kappa_0 = \xi(\hat{\theta})\) for any \(\hat{\theta} \in (\theta_1, \theta_h)\), and rearranging, we have that for all \(\theta \in (\theta_1, \theta_h)\), \(\pi_a(\theta, \hat{s}) = (\kappa_0 - H(\theta))/h(\theta)\), so that \(\hat{\alpha} = \gamma(\cdot, \kappa_0)\).

Let us show that \(\kappa_0 = \hat{\kappa}(\theta_1, \theta_h)\). Let \(v_l(\theta, \kappa) = \hat{\nu}(\theta_1) + \int_{\theta_1}^{\theta} \gamma(\tau, \kappa) d\tau\). Since \(v_l(\theta_1, \kappa) = \hat{\nu}(\theta_1)\) and so is independent of \(\kappa\), it follows that on \((\theta_1, \theta_h)\),
\[
\frac{d}{d\kappa} M(\theta, \gamma(\cdot, \kappa), v_l(\cdot, \kappa)) = \left( \pi_a(\theta, \gamma(\cdot, \kappa), v_l(\cdot, \kappa)) - \frac{H(\theta_h) - H(\theta)}{h(\theta)} \right) \gamma(\theta, \kappa)
\]
\[
= \left( \frac{\kappa - H(\theta_h)}{h(\theta)} \right) \gamma(\theta, \kappa) = s - (\kappa - H(\theta_h)),
\]
32
since $\gamma_\kappa < 0$. But then, by Lemma\[8\]
\[\frac{d}{d\kappa} \int_{\theta_1}^{\theta_h} \pi(\theta, \gamma(\cdot, \kappa), v_l(\cdot, \kappa))h(\theta)d\theta = -(\kappa - H(\theta_h)),\]
as well, so that $Y_l(\kappa) \equiv \int_{\theta_1}^{\theta_h} \pi(\theta, \gamma(\cdot, \kappa), v_l(\cdot, \kappa))h(\theta)d\theta$ is strictly single-peaked in $\kappa$ with maximum at $\kappa = H(\theta_h)$. Similarly, if we define $v_h(\theta, \kappa) = \tilde{v}(\theta_h) - \int_{\theta}^{\theta_h} \gamma(\tau, \kappa)d\tau$, then $Y_h(\kappa) \equiv \int_{\theta_1}^{\theta_h} \pi(\theta, \gamma(\cdot, \kappa), v_h(\cdot, \kappa))h(\theta)d\theta$ is strictly single-peaked in $\kappa$ with maximum at $\kappa = H(\theta_l)$ where to show this, one integrates
\[\int_{\theta_1}^{\theta_h} \pi(\theta, \alpha, v)h(\theta)d\theta = \int_{\theta_1}^{\theta_h} \left( B(\alpha(\theta)) - c(\alpha(\theta), \theta) - v(\theta_h) + \int_{\theta}^{\theta_h} \alpha(\tau)d\tau \right) h(\theta)d\theta\]
by parts to arrive at an analogue to $M$.

Note that one of $[5]$ and $[6]$ must bind, otherwise reducing $v$ by a small positive constant is profitable. Assume that $\tilde{v}(\theta_h) > v^{-n}(\theta_h)$. Then, $(\gamma(\cdot, \kappa), v_l(\cdot, \kappa))$ is feasible on a neighborhood of $\kappa_o$, and so, since $Y_l$ is strictly single-peaked with maximum at $H(\theta_h)$ we must have $\kappa_o = H(\theta_h)$. Since $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$, and since $(\gamma(\cdot, \kappa_o), \tilde{v})$ is feasible, we have
\[\tilde{v}(\theta_h) = v^{-n}(\theta_l) + \int_{\theta_l}^{\theta_h} \gamma(\tau, H(\theta_h))d\tau > v^{-n}(\theta_h),\]
and so $z(\theta_1, \theta_h, H(\theta_h)) < 0$, and thus by definition of $\tilde{\kappa}$, we have $\tilde{\kappa}(\theta_l, \theta_h) = H(\theta_h)$ as well. Similarly, if $\tilde{v}(\theta_l) > v^{-n}(\theta_l)$ then, using $Y_h$, we must have $\kappa_o = H(\theta_l) = \tilde{\kappa}(\theta_l, \theta_h)$.

Assume that $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$ and $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$. Then, by definition, we have $z(\theta_l, \theta_h, \kappa_o) = 0$. Assume $\kappa_o > H(\theta_h)$. Then,
\[v_l(\theta_h, H(\theta_h)) = \tilde{v}(\theta_l) + \int_{\theta_l}^{\theta_h} \gamma(\tau, H(\theta_h))d\tau > \tilde{v}(\theta_l) + \int_{\theta_l}^{\theta_h} \gamma(\tau, \kappa_o)d\tau = \tilde{v}(\theta_h) = v^{-n}(\theta_h),\]
so that $(\gamma(\cdot, H(\theta_h)), v_l(\cdot, H(\theta_h)))$ is feasible, which contradicts the optimality of $(\tilde{\alpha}, v_o)$ since $Y_l$ is uniquely maximized at $H(\theta_h)$. Hence $\kappa_o \leq H(\theta_h)$. Similarly, $\kappa_o \geq H(\theta_l)$, and thus $\kappa_o \in [H(\theta_l), H(\theta_h)]$, from which $\kappa_o = \tilde{\kappa}(\theta_l, \theta_h)$, again by definition of $\tilde{\kappa}$. \[\square\]

**Proposition 6** Let $s$ be Nash, fix $n$, and let $s^n = (\alpha, v)$. Then, there is $\kappa \in [H(\theta_l), H(\theta_h)]$ such that, for all $\theta \in [\theta_1, \theta_h]$,
\[\pi_\alpha(\theta, \alpha, v) = \frac{\kappa - H(\theta)}{h(\theta)},\]
where if $n = 1$, $\kappa = 0$ and if $n = N$, $\kappa = 1$.

**Proof** We will show that if $(\alpha, v)$ is not equal to $\tilde{s}$, then we can profitably perturb it in the direction of $\tilde{s}$ (other than respecting monotonicity of $\tilde{\alpha}$ and the integral condition, it is not
important that $\tilde{s}$ is defined outside of $(\theta_l, \theta_h)$). We need to respect monotonicity and the fact that workers both within and outside of $(\theta_l, \theta_h)$ may be affected.

Let $\tilde{s}(\delta)$ be given by $\tilde{\alpha}(\cdot, \delta) = (1 - \delta)\alpha + \delta \tilde{\alpha}$ and $\tilde{v}(\cdot, \delta) = (1 - \delta)v + \delta \tilde{v}$, so that $\tilde{s}(0) = (\alpha, v)$ and $\tilde{s}(1) = \tilde{s}$. Let $\tilde{v} = v^{-n}/2 + v/2$, so that $\tilde{v} > v^{-n}$ on $(\theta_l, \theta_h)$. Now, let $\tilde{v}(\cdot, \delta) = \max(\tilde{v}, \tilde{v}(\cdot, \delta))$, let $\tilde{\alpha}(\cdot, \delta)$ be a subgradient to $\tilde{v}(\cdot, \delta)$, and let $\tilde{s}(\delta) = (\tilde{\alpha}(\cdot, \delta), \tilde{v}(\cdot, \delta))$. By construction, $\tilde{s}$ always wins on $[\theta_l, \theta_h]$, and may win other workers as well. Finally, let $P(\delta)$ be the set upon which $\tilde{s}(\delta)$ is profitable, and construct $\tilde{s}(\delta) = (\tilde{\alpha}(\cdot, \delta), \tilde{v}(\cdot, \delta))$ from $\tilde{s}(\delta)$ as in Lemma 3. Note that

$$\Pi(\tilde{s}(\delta), s^{-n}) = \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) \varphi(\theta, \tilde{s}(\delta)) h(\theta) d\theta$$

$$\geq \int_{P(\delta) \cap [\theta_l, \theta_h]} \pi(\theta, \tilde{s}(\delta)) \varphi(\theta, \tilde{s}(\delta)) h(\theta) d\theta$$

$$= \int_{P(\delta) \cap [\theta_l, \theta_h]} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta$$

$$\geq \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta.$$

The first inequality follows since $\pi(\cdot, \tilde{s}(\delta)) \geq 0$, the second equality since $\tilde{s}(\delta)$ and $\tilde{s}(\delta)$ agree on $P(\delta)$ and $\varphi(\cdot, \tilde{s}(\delta)) = 1$ on $[\theta_l, \theta_h]$, and the second inequality since $\pi(\theta, \tilde{s}(\delta)) \leq 0$ outside of $P(\delta)$.

It is thus enough to show that for $\delta$ sufficiently small,

$$\int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta > \int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v) h(\theta) d\theta,$$

since by $PS$, $\pi(\theta, \alpha, v) \varphi(\theta, s) = 0$ outside of $[\theta_l, \theta_h]$. Because $\tilde{s}(0) = (\alpha, v)$, it is sufficient that

$$\frac{d}{d\delta} \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta \bigg|_{\delta=0} > 0.$$

But,

$$\frac{d}{d\delta} \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta \bigg|_{\delta=0} = \frac{d}{d\delta} \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta \bigg|_{\delta=0}.$$

since for each $\theta \in (\theta_l, \theta_h)$, $v(\theta) > \tilde{v}(\theta)$, and so at $\delta = 0$, $(\tilde{\alpha}(\theta, \delta))_\delta = (\alpha(\theta, \delta))_\delta$ and $(\tilde{v}(\theta, \delta))_\delta = (\tilde{v}(\theta, \delta))_\delta$. And, since $\tilde{s}$ is the unique solution on $(\theta_l, \theta_h)$ to the relaxed problem $P(\theta_l, \theta_h)$,

$$\int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}) h(\theta) d\theta = \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(1)) h(\theta) d\theta$$

$$> \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(0)) h(\theta) d\theta = \int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v) h(\theta) d\theta.$$
So, since \( \tilde{s} \) is linear in \( \delta \), and thus \( \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta \) is concave in \( \delta \), it must be that

\[
\frac{d}{d\delta} \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\delta)) h(\theta) d\theta \bigg|_{\delta=0} > 0
\]
as desired. \( \square \)

### 8.5 Proof of OB

**Proposition 7** Let \( s \) be Nash. Then, (9) and (10) hold.

**Proof** Fix \( n \). We will prove (10), with (9) analogous. If \( \theta_l^{n+1} > \theta_h^n = \theta_h \), then (10) is automatic, since by Proposition 1 \( \pi(\theta_h, \alpha, v) = 0 \) and \( \alpha(\theta_h) = \alpha^{n+1}(\theta_h) \). So, assume \( \theta_l^{n+1} = \theta_h \), and note that by Proposition 6 \( \alpha \) is strictly increasing to the left of \( \theta_h \), and \( a^{-n} = \alpha^{n+1} \) is strictly increasing to the right of \( \theta_h \).

We will consider deviations that add or subtract workers immediately to the right or left of \( \theta_h \). (This proof would be much easier if we knew that all crossing were strictly transverse, e.g., that we never had \( \alpha^n(\theta_h) = \alpha^{n+1}(\theta_h) \). Then, we could use \( \gamma(\cdot, \kappa) \) and vary \( \kappa \) holding fixed \( v(\theta_l) \).

Fix \( n \) and \( 0 < \epsilon < \theta_h - \theta_l \). For \( y \) positive or negative, define

\[
\hat{\alpha}(\theta, y) = \begin{cases} 
\alpha(\theta) & \text{if } \theta < \theta_h - \epsilon \\
\max\{\alpha(\theta_h - \epsilon), \min\{\alpha(\theta) + y, \alpha(\theta_h)\}\} & \text{if } \theta \in [\theta_h - \epsilon, \theta_h] \\
\alpha(\theta_h) & \text{if } \theta > \theta_h
\end{cases}
\]

and define \( \hat{v}(\theta, y) = v(\theta_l) + \int_{\theta_l}^\theta \hat{\alpha}(\tau, y) d\tau \). Because \( \hat{\alpha}(\tau, y) \) is bounded and for each \( y \), differentiable in \( y \) for almost all \( \tau \), with \( \hat{\alpha}_y(\tau, y) = 1 \) wherever it is defined, \( \hat{v} \) is differentiable on a neighborhood of \( (\theta_h, 0) \), with \( \hat{v}_y(\theta_h, 0) = \epsilon > 0 \). Hence, if we implicitly define \( \hat{y}(\theta') \) by \( \hat{v}(\theta', \hat{y}(\theta')) - v^{-n}(\theta') = 0 \), then \( \hat{y} \) is well defined, with

\[
\frac{\hat{y}_\theta'(\theta')}{\hat{y}_y(\theta', \hat{y}(\theta'))} = \frac{a^{-n}(\theta') - \hat{\alpha}(\theta', \hat{y}(\theta'))}{\hat{v}(\theta', \hat{y}(\theta'))} \geq 0. \tag{17}
\]

Further, when \( \hat{y}(\theta') > 0 \), then \( \hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta) > 0 \) for all \( \theta \in (\theta_l, \theta_h) \), and hence any crossing of zero by \( \hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta) \) above \( \theta_l \) occurs where \( \theta > \theta_h \), and thus where

\[
(\hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta))_\theta = \hat{\alpha}(\theta, \hat{y}(\theta')) - a^{-n}(\theta) = \alpha(\theta_h) - a^{-n}(\theta) < 0,
\]
since \( a^{-n}(\theta) > a^{-n}(\theta_h) \geq \alpha(\theta_h) \). Thus, indeed, \( \varphi = 1 \) for all \( \theta \in [\theta_l, \theta'] \), and \( \varphi = 0 \) elsewhere. Similarly, if \( \hat{y}(\theta') < 0 \), then any crossing of zero by \( \hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta) \) above \( \theta_l \) occurs where
\[ \theta < \theta_h, \text{ and thus where } \hat{\alpha}(\theta, \hat{y}(\theta')) \leq \alpha(\theta), \text{ and hence} \]
\[ (\hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta))_{\theta} = \hat{\alpha}(\theta, \hat{y}(\theta')) - a^{-n}(\theta) \leq \alpha(\theta) - a^{-n}(\theta_h) < 0, \]
by NEO, and so again \( \varphi = 1 \) for all \( \theta \in [\theta_l, \theta'] \), and \( \varphi = 0 \) elsewhere.

Let \( j(\theta') \) be the profit from this deviation. Then,
\[ j(\theta') = \int_{\theta_l}^{\theta_h - \varepsilon} \pi(\theta, \alpha, v)h(\theta)d\theta + \int_{\theta_h - \varepsilon}^{\theta'} \pi(\theta, \hat{\alpha}(\cdot, \hat{y}(\theta'))), \hat{v}(\cdot, \hat{y}(\theta'))h(\theta)d\theta, \]
since for \( \theta < \theta_h - \varepsilon, \hat{\alpha} = \alpha \) and \( \hat{v} = v \). Thus,
\[ j_{\theta'}(\theta') = \pi(\theta', \hat{\alpha}(\cdot, \hat{y}(\theta'))), \hat{v}(\cdot, \hat{y}(\theta'))h(\theta') \]
\[ + \int_{\theta_h - \varepsilon}^{\theta'} (\pi_a(\theta, \hat{\alpha}(\cdot, \hat{y}(\theta'))), \hat{v}(\cdot, \hat{y}(\theta')) - \hat{v}_y(\theta, \hat{y}(\theta')))h(\theta)d\theta. \]

To evaluate this at \( \theta' = \theta_h \), note that \( \hat{y}(\theta_h) = 0, \hat{\alpha}(\theta, 0) = \alpha(\theta), \hat{\alpha}_y(\theta, 0) = 1 \) for \( \theta \in (\theta_h - \varepsilon, \theta_h) \), and that \( \hat{v}(\cdot, 0) = v \), and so, using (17) and \( \hat{v}_y(\theta_h, 0) = \varepsilon \),
\[ j_{\theta'}(\theta_h) = \pi(\theta_h, \alpha, v)h(\theta_h) + \frac{a^{-n}(\theta_h) - \alpha(\theta_h)}{\varepsilon} \int_{\theta_h - \varepsilon}^{\theta_h} (\pi_a(\theta, \alpha, v) - \hat{v}_y(\theta, 0))h(\theta)d\theta \]
\[ = \pi(\theta_h, \alpha, v)h(\theta_h) + (a^{-n}(\theta_h) - \alpha(\theta_h)) \pi_a(\tau, \alpha, v)h(\tau) \]
for some \( \tau \in [\theta_h - \varepsilon, \theta_h] \) by the Mean Value Theorem, and where we note that \( \hat{v}_y(\tau, 0) = \tau - (\theta_h - \varepsilon) < \varepsilon \). But, for all \( \varepsilon > 0 \), this perturbation is feasible for all \( \theta' \) in a neighborhood of \( \theta_h \), and so since \( (\alpha, v) \) is optimal, we have \( j_{\theta'}(\theta_h) = 0 \). Taking \( \varepsilon \to 0 \), we have \( \tau \to \theta_h \), and hence, canceling \( h(\theta_h) \), we arrive at \( 0 = \pi(\theta_h, \alpha, v) + (a^{-n}(\theta_h) - \alpha(\theta_h)) \pi_a(\theta_h, \alpha, v) \) as desired. \( \square \)

8.6 Proofs for Section 5

Proof of Lemma 3 The first inequality follows from PP. Assume that \( \pi^n(\theta, \alpha^n, v^n) > B^n(\alpha^n(\theta)) - \max_{n'} B^{n'}(\alpha^n(\theta)) \). Then,
\[ v^n(\theta) = B^n(\alpha^n(\theta)) - c(\alpha^n(\theta), \theta) - \pi^n(\theta) \]
\[ < B^n(\alpha^n(\theta)) - c(\alpha^n(\theta), \theta) - (B^n(\alpha^n(\theta)) - \max_{n'} B^{n'}(\alpha^n(\theta))) \]
\[ = \max_{n'} B^{n'}(\alpha^n(\theta)) - c(\alpha^n(\theta), \theta), \]
contradicting NP. \( \square \)

Proof of Theorem 2 Let \( \hat{n} \) be the firm that serves \( \theta \) in an efficient allocation, let \( \hat{\theta} \) be any type that \( \hat{n} \) serves in equilibrium, and let \( \hat{\alpha} = \alpha^n(\hat{\theta}) \). By Lemma 3 \( \hat{n} \) is the most efficient firm at action
\( \hat{a} \). Hence, \( \hat{a} \in [a_{l}^{\hat{n}}, a_{h}^{\hat{n}}] \). Similarly, since \( \hat{n} \) efficiently serves \( \theta \), we also have \( a_{l}^{\hat{n}}(\theta) \in [a_{l}^{\hat{n}}, a_{h}^{\hat{n}}] \). Thus, \( |\hat{a} - a_{l}^{\hat{n}}(\theta)| \leq d_{2} \).

The payoff to \( \theta \) of imitating \( \hat{\theta} \) is

\[
\begin{align*}
v^{\hat{h}}(\hat{\theta}) + c(\hat{a}, \hat{\theta}) - c(\hat{a}, \theta) & \geq B^{\hat{h}}(\hat{a}) - c(\hat{a}, \theta) - d_{1} + c(\hat{a}, \theta) - c(\hat{a}, \theta) \\
& = B^{\hat{h}}(\hat{a}) - c(\hat{a}, \theta) - d_{1} \\
& = B^{\hat{h}}(a_{l}^{\hat{n}}(\theta)) - c(a_{l}^{\hat{n}}(\theta), \theta) + (B^{\hat{h}}(\hat{a}) - c(\hat{a}, \theta) - (B^{\hat{h}}(a_{l}^{\hat{n}}(\theta)) - c(a_{l}^{\hat{n}}(\theta), \theta))) - d_{1} \\
& = m_{\theta}(\theta) - d_{1} + (B^{\hat{h}}(\hat{a}) - c(\hat{a}, \theta) - (B^{\hat{h}}(a_{l}^{\hat{n}}(\theta)) - c(a_{l}^{\hat{n}}(\theta), \theta))) \\
& \geq m_{\theta}(\theta) - d_{1} - \frac{1}{2}d_{2}^2\delta,
\end{align*}
\]

where the first inequality uses Lemma 3, and the second inequality follows from \((B^{\hat{h}}(a_{l}^{\hat{n}}(\theta)) - c(a_{l}^{\hat{n}}(\theta), \theta))_{a} = 0\), the definition of \( \delta \), and \( |a_{l}^{\hat{n}}(\theta) - \hat{a}| < d_{2} \).

\[ \square \]

9 Appendix B: Proofs for Section 6

9.1 Proofs for Section 6.3

Remark 1 For this and the next three subsections, we assume stacking, and whenever we fix \( n \) and \( s^{-n} \), we assume \( s^{-n} \) satisfies C1 and C2.

Proof of Lemma 4 To see that \( s \) satisfies \( SPS \), note that \( v_{\theta}' = \alpha n' < \gamma(\cdot, 1) \) for each \( n' < n \), by C1 and stacking, and similarly \( v_{\theta}' = \alpha n' > \gamma(\cdot, 0) \) for each \( n' > n \). Thus, there is \( \theta_{x} \in [\theta, \bar{\theta}] \) such that \( a_{-n} < \gamma(\cdot, 1) \) for \( \theta < \theta_{x} \), and \( a_{-n} > \gamma(\cdot, 0) \) for \( \theta > \theta_{x} \). But then, \( v \), which by C1 has slope in \([\gamma(\cdot, 1), \gamma(\cdot, 0)]\), can cross \( v^{-n} \) at most twice, where if it crosses \( v^{-n} \) from below, it does so strictly and on \([\theta, \theta_{x}]\), and if it does so from above, it does so strictly and on \([\theta_{x}, \bar{\theta}]\). In all cases, \( v \) single dominates \( v^{-n} \) on some interval, which is non-empty since \( n \) sometimes wins, and we have \( SPS \).

To see that \( OB \) implies \( NP \), note that by C1 and stacking, \( a_{-n} > \alpha_{s} \) for all \( \theta \) above \( \theta_{h} \), and hence, using [11], the profits to poaching, \( \pi(\cdot, a_{-n}, v^{-n}) \), are falling everywhere above \( \theta_{h} \). (Similarly, poaching profits rise below \( \theta_{l} \).) It is thus enough to show that poaching just above \( \theta_{h} \) does not make sense. This follows from [10], since \( \pi \) is concave in \( a \), and so

\[
0 = \pi_{a}(\theta_{h}, \alpha, v) (a_{-n}(\theta_{h}) - \alpha(\theta_{h})) + \pi(\theta_{h}, \alpha, v) > \pi(\theta_{h}, a_{-n}(\theta_{h}), v) - \pi(\theta_{h}, \alpha, v) + \pi(\theta_{h}, \alpha, v) = \pi(\theta_{h}, a_{-n}, v).
\]

where the inequality follows since \( \pi \) is concave in \( a \).

\[ \square \]
9.2 Proofs for Section 6.4

Proof of Lemma [5] Part (i) follows since \( \tilde{\kappa} \in [0,1] \) using Lemma [7]. To see (ii), consider any maximizer \((\theta_l, \theta_h)\) of \( r \) at which \( \tilde{v}(\theta_h) > v^{-n}(\theta_h) \). Then, by the Envelope Theorem,

\[
r_{\theta_h}(\theta_l, \theta_h) = \left( \int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\theta_l, \theta_h)) h(\theta) d\theta \right)_{\theta_h}
\]

\[
= \pi(\theta_h, \tilde{s}(\theta_l, \theta_h)) h(\theta) + \int_{\theta_l}^{\theta_h} (\pi(\theta, \tilde{s}(\theta_l, \theta_h)))_{\theta_h} h(\theta) d\theta
\]

But, \( \tilde{v}(\theta_h) > v^{-n}(\theta_h) \) implies that \( \tilde{\kappa}(\theta_l, \theta_h) = H(\theta_h) \), and so by (11), \( \pi(\cdot, \tilde{s}(\theta_l, \theta_h)) \) is increasing on \((\theta_l, \theta_h)\). Thus, since \((\theta_l, \theta_h)\) is a maximum of \( r \), and so \( r(\theta_l, \theta_h) \) is strictly positive by C2, we have \( \pi(\theta_h, \tilde{s}(\theta_l, \theta_h)) > 0 \), and thus \( r_{\theta_h}(\theta_l, \theta_h) > 0 \). Since \((\theta_l, \theta_h)\) is optimal, it must thus be that \( \theta_h = \tilde{\theta} \).

Similarly, if \( \tilde{v}(\theta_l) > v^{-n}(\theta_l) \), then \( \theta_l = \theta \). But then, in all cases, \( \tilde{s}(\theta_l, \theta_h) \) is single dominant on \((\theta_l, \theta_h)\), using stacking and C1. Part (iii) follows immediately, as the relevant domains of integration agree.

Lemma 9 There exist \( m \) and \( \overline{m} \) with \( m \leq \theta^x \leq \overline{m} \) such that \( \pi(\theta, a^{-n}, v^{-n}) \) is strictly positive if \( \theta \in (m, \overline{m}) \), strictly negative and strictly increasing if \( \theta < m \), and strictly negative and strictly decreasing if \( \theta > \overline{m} \).

Proof By stacking and C1, for \( \theta > \theta^x \),

\[
a^{-n}(\theta) > \gamma(\theta, 0) \geq \gamma(\theta, H(\theta)) = \alpha_s(\theta),
\]

and so \( \pi_a(\theta, a^{-n}, v^{-n}) < 0 \). Hence, anywhere that \( a^{-n} \) is differentiable, we have by (11) that \( (\pi(\theta, a^{-n}, v^{-n}))_{\theta} < 0 \). Further, at any point where \( a^{-n} \) jumps, say from \( a_l \) to \( a_h \), we have, since \( v^{-n} \) is continuous, and since \( a_h > a_l > \alpha_s(\theta) \) that \( \pi(\theta, a_h, v^{-n}) - \pi(\theta, a_l, v^{-n}) < 0 \). Hence \( \pi(\cdot, a^{-n}, v^{-n}) \) is strictly decreasing and so single-crosses 0 from above at most once on \([\theta^x, \tilde{\theta}]\). If such a crossing exists, define \( \overline{m} \) as the crossing. If \( \pi(\tilde{\theta}, a^{-n}, v^{-n}) > 0 \), take \( \overline{m} = \tilde{\theta} \), and if \( \pi(\theta^x, a^{-n}, v^{-n}) < 0 \), take \( \overline{m} = \theta^x \). Construct \( m \) similarly.

Lemma 10 Let \((\alpha, \nu)\) be any feasible menu for \( n \), let \( \nu \) be dominant on \((\tau_l, \tau_h)\), and let

\[
\int_{\tau_l}^{\tau_h} \pi(\theta, \alpha, \nu) h(\theta) d\theta \geq 0. \tag{18}
\]

Then, \((\tau_l, \tau_h) \cap [m, \overline{m}] \neq \emptyset \).

Proof By Proposition [3] we can assume that \( \pi(\theta, \alpha, \nu) \geq 0 \) everywhere. Assume \( \tau_l \geq m \geq \theta^x \). Then, \( v(\tau_l) = v^{-n}(\tau_l) \) by definition of dominance and since \( v \) and \( v^{-n} \) are continuous. Since for
all $\theta \in (\tau_1, \tau_h)$

\[ v(\tau_1) + \int_{\tau_1}^{\theta} \alpha(\tau) d\tau = v(\theta) > v^{-n}(\theta) = v^{-n}(\tau_1) + \int_{\tau_1}^{\theta} a^{-n}(\tau) d\tau \]

it follows that there is $\tau \in (\tau_1, \tau_h)$ where $\alpha(\tau) > a^{-n}(\tau)$. But, since $\tau > m \geq \theta^x$, and using C1, it follows that $a^{-n}(\tau) > a_*(\tau)$, and so

\[ \pi(\tau, \alpha(\tau), v(\tau)) < \pi(\tau, a^{-n}(\tau), v(\tau)) < \pi(\tau, \alpha^{-n}(\tau), v^{-n}(\tau)) < 0, \]

a contradiction. Similarly, it cannot be that $\tau_h \leq m$.

\[ \square \]

**Proof of Proposition 2** Wlog, assume that $(\alpha, v)$ loses money nowhere, and recall that

\[ \Pi(s) = \int_\theta^\hat{\theta} \pi(\theta, \alpha, v) \varphi(\theta, s) h(\theta) d\theta. \quad (19) \]

Assume that $v$ dominates $v^{-n}$ on an interval $I_H$ with $\theta^x \leq L_H \leq m \leq \tilde{I}_H$. In this case, define $m^* = \tilde{I}_H$. If there is no such interval, define $m^* = m$. Similarly, if $v$ dominates $v^{-n}$ on an interval $I_L$ with $L_L \leq m \leq \tilde{I}_L \leq \theta^x$, then define $m^* = L_L$, and if there is no such interval, define $m^* = m$.

Consider first any positive lengthed interval $J \subseteq [\bar{m}^*, \bar{\theta}]$ on which $v = v^{-n}$, and such that $\int_J \varphi(\theta, s) d\theta > 0$. Then, $\alpha = a^{-n}$ on this interval, and so, $\bar{m}^* \geq \bar{m}$, $\pi(\theta, \alpha, v) < 0$ for all $\theta > \bar{m}^*$. Hence, excluding $J$ from the domain of the integral in (19) increases the integral.

Consider next any positive lengthed interval $J = (\bar{J}, J)$ with $J \geq \bar{m}^*$ and on which $v$ is dominant. Then, by Lemma 10

\[ \int_J \pi(\theta, \alpha, v) h(\theta) d\theta < 0, \]

and thus excluding $J$ again increases the integral in (19). The argument for intervals below $\bar{m}^*$ is the same. We thus have

\[ \Pi(s) \leq \int_{\bar{m}^*}^{m^*} \pi(\theta, \alpha, v) \varphi(\theta, s) h(\theta) d\theta. \quad (20) \]

Define $\hat{v} = \max(v, v^{-n})$, with associated $\hat{\alpha}$, where at all $\theta$ where $v(\theta) \geq v^{-n}(\theta)$, we can take $\hat{\alpha} = \alpha$, and at almost all $\theta$ where $v(\theta) \leq v^{-n}(\theta)$, we can take $\hat{\alpha} = a^{-n}$ (on any interval where $v(\theta) = v^{-n}(\theta)$, $\alpha = a^{-n}$ almost everywhere, and so there is a zero measure set where the two definitions might be in conflict). But then, everywhere that $\varphi(\theta, s)$ is positive (and so $v(\theta) \geq v^{-n}(\theta)$), we have $\pi(\theta, \hat{\alpha}, \hat{v}) = \pi(\theta, \alpha, v)$, and so, from (20),

\[ \Pi(s) \leq \int_{\bar{m}^*}^{m^*} \pi(\theta, \hat{\alpha}, \hat{v}) \varphi(\theta, s) h(\theta) d\theta. \]
Consider the set of all \( \theta \in (m^*, m^*) \) such that \( \varphi(\theta, s) < 1 \). Then, for each such \( \theta \), \( v(\theta) \leq v^{-n}(\theta) \), and so \( \hat{v}(\theta) = v^{-n}(\theta) \), and \( \hat{\alpha}(\theta) = a^{-n}(\theta) \) almost everywhere. And, since by construction, \( \varphi \) is 1 on \( I_H \) and \( I_L \) (if these sets exist), it follows that \( \theta \in [m, m] \), and so \( \pi(\theta, \hat{\alpha}, \hat{v}) = \pi(\theta, a^{-n}, v^{-n}) \geq 0 \). We thus have

\[
\Pi(s) \leq \int_{m^*}^{\bar{m}^*} \pi(\theta, \hat{\alpha}, \hat{v})h(\theta)d\theta.
\]

But, \( \hat{v} \geq v^{-n} \) by construction, and so \([5]\) and \([6]\) are satisfied in \( \mathcal{P}(m^*, m^*) \), while \( \hat{\alpha} \) was chosen to be a subgradient of the convex function \( \max(v, v^{-n}) \), and hence \([7]\) holds as well. Thus, \((\hat{\alpha}, \hat{v})\) is feasible in the relaxed problem \( \mathcal{P}(\bar{m}^*, \bar{m}^*) \), from which \( \Pi(s) \leq r(m^*, m^*) \).

\[ \square \]

### 9.3 Proofs for Section 6.5

In this section, we establish that the firm has an essentially unique best response. We begin with local properties of \( r \) and then use those properties to show that \( r \) has a unique maximum.

#### 9.3.1 Local Properties of \( r \)

We first show where \( r \) is positive and satisfies \( r_{\theta_l} = r_{\theta_h} = 0 \), we are at a strict local maximum of \( r \). Write \( f_x^+ \) and \( f_x^- \) for the right and left derivatives of \( f \) with respect to \( x \).

**Remark 2** Fix a maximal rectangle \( \tilde{R} = [\iota_l, \iota_h] \times [\iota_l', \iota_h'] \) as defined in Section 6.5. On \([\iota_l, \iota_h]\) and \([\iota_l', \iota_h']\), we can take \( \alpha^{-n} \) to be continuous by \( C1 \), and hence \( v^{-n} \) to be continuously differentiable.

**Lemma 11** Considered as a function on \( \tilde{R} \), \( r \) is continuously differentiable, with

\[
r_{\theta_h}(\theta_l, \theta_h) = \left( \pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \hat{v}) (a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})) + \pi(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \hat{v}) \right) h(\theta_h),
\]

and

\[
r_{\theta_l}(\theta_l, \theta_h) = \left( \pi_a(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \hat{v}) (\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l)) - \pi(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \hat{v}) \right) h(\theta_l).
\]

**Proof** The right side of \([21]\) has the same form as \([10]\). As in the analysis of \( OB \) in Section 8.5, this is the value of increasing \( \theta_h \) by increasing effort immediately to the left of \( \theta_h \), and, since \( \gamma(\cdot, \tilde{\kappa}(\theta_l, \theta_h)) \) solves the relaxed problem, this perturbation is as good as anything. Alternatively, differentiate \( r \), and manipulate, using integration by parts and \([16]\). The proof of \([22]\) is similar. On \( \tilde{R} \), all the terms of \( r_{\theta_h} \) and \( r_{\theta_l} \) are continuous. Hence, \( r \) is continuously differentiable. \( \square \)
As a coherence check, along the lower boundary of \( \tilde{R} \),

\[
(r|_{\tilde{R}})_{\theta h} (\theta_l, \theta_h) = \lim_{\varepsilon \downarrow 0} r_{\theta h}(\theta_l, \theta_h + \varepsilon) = \lim_{\varepsilon \downarrow 0} \frac{r(\theta_l, \theta_h + \varepsilon) - r(\theta_l, \theta_h)}{\varepsilon} = r_{\theta h}^+(\theta_l, \theta_h),
\]

(23)

where the second equality uses L'Hôpital's rule and continuity of \( r_{\theta h} \) on \((\varepsilon', \varepsilon'^{\prime})\). Things are similar on the other boundaries of \( \tilde{R} \).

Recall that \( \Theta \) is the subset of \( R \) on which \( z(\theta_l, \theta_h, \tilde{\kappa}(\theta_l, \theta_h)) = 0 \). Where there is no ambiguity, we will write \( \tilde{\kappa} \) for \( \tilde{\kappa}(\theta_l, \theta_h) \).

**Lemma 12** On \( \tilde{R} \cap \Theta \), we have

\[
(\gamma(\theta_l, \tilde{\kappa}(\theta_l, \theta_h)))_{\theta_l} > \gamma_{\kappa}(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l} \geq 0 \text{ and } (\gamma(\theta_h, \tilde{\kappa}(\theta_l, \theta_h)))_{\theta_h} > \gamma_{\kappa}(\theta_h, \tilde{\kappa})\tilde{\kappa}_{\theta_h} \geq 0,
\]

with

\[
\begin{vmatrix}
(\gamma(\theta_l, \tilde{\kappa}(\theta_l, \theta_h)))_{\theta_l} & (\gamma(\theta_l, \tilde{\kappa}(\theta_l, \theta_h)))_{\theta_h} \\
(\gamma(\theta_h, \tilde{\kappa}(\theta_l, \theta_h)))_{\theta_l} & (\gamma(\theta_h, \tilde{\kappa}(\theta_l, \theta_h)))_{\theta_h}
\end{vmatrix} > 0.
\]

(24)

**Proof** Note that

\[
(\gamma(\theta_l, \tilde{\kappa}))_{\theta_l} = \gamma_{\theta}(\theta_l, \tilde{\kappa}) + \gamma_{\kappa}(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l} > \gamma_{\kappa}(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l},
\]

since \( \gamma_{\theta} > 0 \) using that \( \tilde{\kappa} \in [0, 1] \). But, since \( z(\theta_l, \theta_h, \tilde{\kappa}) = 0 \), by C1, stacking, and \( \tilde{\kappa} \in [0, 1] \), we have \( z_{\theta_l} = \gamma(\theta_l, \tilde{\kappa}) - a^{-\alpha}(\theta_l) = s \theta^2 - \theta_l \geq 0 \) and so, since \( z_{\kappa} > 0 \), \( \tilde{\kappa}_{\theta_l} = -z_{\theta_l}/z_{\kappa} \leq 0 \). Thus \( \gamma_{\kappa}(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l} \geq 0 \), since \( \gamma_{\kappa} < 0 \). Similarly, \( \tilde{\kappa}_{\theta_h} \leq 0 \), and so \( (\gamma(\theta_h, \tilde{\kappa}))_{\theta_h} > \gamma_{\kappa}(\theta_h, \tilde{\kappa})\tilde{\kappa}_{\theta_h} \geq 0 \).

To see (24), note that

\[
\begin{vmatrix}
\gamma(\theta_l, \tilde{\kappa})_{\theta_l} & \gamma(\theta_l, \tilde{\kappa})_{\theta_h} \\
\gamma(\theta_h, \tilde{\kappa})_{\theta_l} & \gamma(\theta_h, \tilde{\kappa})_{\theta_h}
\end{vmatrix} > \begin{vmatrix}
\gamma_{\alpha}(\theta_l, \tilde{\kappa}) + \gamma_{\kappa}(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l} & \gamma_{\alpha}(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l} \\
\gamma_{\alpha}(\theta_h, \tilde{\kappa}) + \gamma_{\kappa}(\theta_h, \tilde{\kappa})\tilde{\kappa}_{\theta_h} & \gamma_{\alpha}(\theta_h, \tilde{\kappa})\tilde{\kappa}_{\theta_h}
\end{vmatrix} = 0,
\]

since each of the four terms on the main diagonal in the second expression is positive, with the terms eliminated in moving to the third expression strictly so.

**Lemma 13** For any \( v, \pi_a(\cdot, \gamma(\cdot, \kappa), v) \) strictly single-crosses zero from above. So does \( \pi_a(\cdot, \gamma(\cdot, \tilde{\kappa}(\cdot, \theta_h)), v) \).

**Proof** We have

\[
\pi_a(\theta, \gamma(\theta, \kappa), v) = \frac{\kappa - H(\theta)}{h(\theta)},
\]

41
and hence,

\( (\pi_a(\theta, \gamma(\theta, \kappa), v))_\theta = -1 - \frac{\kappa - H(\theta)}{h^2(\theta)} h_\theta(\theta), \)

where when \( \pi_a = 0 \) the second term on the right side disappears, and so \( (\pi_a)_\theta = -1 \). The proof of the second claim is similar, replacing \( \kappa \) by \( \tilde{\kappa}(\theta_1, \theta_h) \), and recalling that \( \tilde{\kappa}_{\theta_i} \leq 0 \). \( \square \)

**Lemma 14** Consider \( r \) as a function on \( \tilde{R} \cap \Theta \). Then, \( r_{\theta_h} \theta_h < 0 \). Assume that \( r(\theta_1, \theta_h) > 0 \). Then, if \( r_{\theta_1}(\theta_1, \theta_h) = 0 \), then \( r_{\theta_1 \theta_h}(\theta_1, \theta_h) < 0 \), if \( r_{\theta_1}(\theta_1, \theta_h) = 0 \), then \( r_{\theta_1 \theta_h}(\theta_1, \theta_h) < 0 \), and if \( r_{\theta_1}(\theta_1, \theta_h) = r_{\theta_1}(\theta_1, \theta_h) = 0 \), then \( r \) is locally strictly concave at \((\theta_1, \theta_h)\).

**Proof** Note first that since \( z(\theta_1, \theta_h, \tilde{\kappa}(\theta_1, \theta_h)) = 0 \), \( \tilde{v}(\theta_1) = v^{-\eta}(\theta_1) \) and \( \tilde{v}(\theta_h) = v^{-\eta}(\theta_h) \). Assume that \( r_{\theta_h}(\theta_1, \theta_h) = 0 \). We claim that \( \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) < 0 \). To see this, note that by Lemma 13 if \( \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) \geq 0 \), then \( \pi_a > 0 \) on all of \([\theta_1, \theta_h)\). But then, by \([11] \), and since \( r(\theta_1, \theta_h) > 0 \), it follows that \( \pi(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) > 0 \), and so using \([21] \), \( r_{\theta_h}(\theta_1, \theta_h) > 0 \), contradicting the premise.

Using \([21] \),

\[
\frac{r_{\theta_h \theta_h}(\theta_1, \theta_h)}{h(\theta_h)} = \left(1 + \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) (\gamma(\theta_1, \kappa))_\theta (a^{-\eta}(\theta_h) - \gamma(\theta_1, \kappa)) \right) + \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) (a^{-\eta}(\theta_h) - \gamma(\theta_1, \kappa))_\theta + \gamma(\theta_1, \kappa) + \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) (\gamma(\theta_1, \kappa))_\theta - a^{-\eta}(\theta_h)
\]

where the term involving \( h_\theta \) disappears since \( r_{\theta_1} = 0 \), and where we use that \( \tilde{v}(\theta_h) = v^{-\eta}(\theta_h) \), and hence \( (\tilde{v}(\theta_h))_\theta = (v^{-\eta}(\theta_h))_\theta = a^{-\eta}(\theta_h) \). Cancelling,

\[
\frac{r_{\theta_h \theta_h}(\theta_1, \theta_h)}{h(\theta_h)} = \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) (\gamma(\theta_1, \kappa))_\theta (a^{-\eta}(\theta_h) - \gamma(\theta_1, \kappa)) \leq \frac{\pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) (\gamma(\theta_1, \kappa))_\theta}{a^{-\eta}(\theta_h)}
\]

where the first inequality uses that \( \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) < 0 \) and the second uses that \( \pi_{aa} < 0 \), that by Lemma 12 \( (\gamma(\theta_1, \kappa))_\theta > 0 \), and that by stacking, C1, and \( \tilde{\kappa} \in [0, 1], a^{-\eta}(\theta_h) - \gamma(\theta_1, \kappa) > 0 \).

Similarly, using \([22] \), if \( r(\theta_1, \theta_h) > 0 \) and \( r_{\theta_1}(\theta_1, \theta_h) = 0 \) then \( \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) > 0 \). Thus, taking cancelations as before, we obtain

\[\text{To be careful, } a_g^{-\eta}, \text{ and hence } r_{\theta_h \theta_h} \text{ may not be everywhere defined. But, since } a^{-\eta} \text{ is increasing, } \lim_{\varepsilon \downarrow 0} a_g^{-\eta}(\theta_h + \varepsilon) \geq 0 \text{ and } \lim \inf_{\varepsilon \downarrow 0} a_g^{-\eta}(\theta_h - \varepsilon) \geq 0, \text{ and so, since } \pi_a(\theta_1, \gamma(\theta_1, \kappa), \tilde{v}) \leq 0, \text{ we have } \lim \sup_{\varepsilon \downarrow 0} r_{\theta_h \theta_h}(\theta_1, \theta_h + \varepsilon) > 0, \text{ and } \lim \sup_{\varepsilon \downarrow 0} r_{\theta_h \theta_h}(\theta_1, \theta_h - \varepsilon) < 0. \text{ We henceforth ignore this technical detail.} \]
\[
\frac{r_{\theta_1 \theta_1}}{h(\theta_1)}(\theta_1, \theta_h) = \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( \gamma(\theta_1, \kappa) - a^{-n}(\theta_1) \right) - \pi_a(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) a_g^{-n}(\theta_1)
\]
\[
\leq \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( \gamma(\theta_1, \kappa) - a^{-n}(\theta_1) \right)
\]
\[
< 0.
\]

We turn next to the cross derivatives. From (21), whether or not \(r_{\theta_h}(\theta_1, \theta_h) = 0\),
\[
\frac{r_{\theta_1 \theta_h}}{h(\theta_1)}(\theta_1, \theta_h) = \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( a^{-n}(\theta_1) - \gamma(\theta_1, \kappa) \right) + \pi_a(\theta_1, \gamma(\theta_1, \kappa), \bar{v})(\gamma(\theta_1, \kappa))_{\theta_1},
\]
and similarly, from (22),
\[
\frac{r_{\theta_1 \theta_h}}{h(\theta_1)}(\theta_1, \theta_h) = \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( \gamma(\theta_1, \kappa) - a^{-n}(\theta_1) \right) 15
\]

To see that \(r_{\theta_1 \theta_h}(\theta_1, \theta_h) < 0\), start from (28), and note that \(\pi_{aa} < 0\), that \((\gamma^n(\theta_1, \kappa))_{\theta_h} > 0\), and that \((\gamma(\theta_1, \kappa) - a^{-n}(\theta_1)) > 0\).

For strict local concavity, it remains to show that where \(r_{\theta_1} = r_{\theta_h} = 0\), we have \(d \equiv r_{\theta_1 \theta_1} r_{\theta_1 \theta_h} - r_{\theta_1 \theta_h}^2 > 0\). From (25) and (26),
\[
\frac{d}{h(\theta_1)h(\theta_h)} \geq \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( a^{-n}(\theta_1) - \gamma(\theta_1, \kappa) \right) \times \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( \gamma(\theta_1, \kappa) - a^{-n}(\theta_1) \right),
\]
while from (27) and (28),
\[
\frac{d}{h(\theta_1)h(\theta_h)} = \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( \gamma(\theta_1, \kappa) - a^{-n}(\theta_1) \right) \times \pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}) (\gamma(\theta_1, \kappa))_{\theta_1} \left( a^{-n}(\theta_1) - \gamma(\theta_1, \kappa) \right).
\]

Cancelling the three positive terms \(h(\theta_1)h(\theta_h), (\gamma(\theta_1, \kappa) - a^{-n}(\theta_1))(a^{-n}(\theta_h) - \gamma(\theta_1, \kappa)),\) and \(\pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v})\pi_{aa}(\theta_1, \gamma(\theta_1, \kappa), \bar{v}),\) it suffices that
\[
(\gamma(\theta_1, \kappa))_{\theta_1} (\gamma(\theta_1, \kappa))_{\theta_1} - (\gamma(\theta_1, \kappa))_{\theta_1} (\gamma(\theta_1, \kappa))_{\theta_1} \theta_1 > 0,
\]
\[\text{These two expressions must of course be equal, but it is convenient to express them in these two different ways.}\]
which follows from Lemma 15.

9.3.2 Essentially Unique Optimality

Say that \( f : \mathbb{R} \to \mathbb{R} \) has a critical point at \( x \) if \( f_x^-(x)f_x^+(x) \leq 0 \), so that \( f_x \) at least weakly changes sign at \( x \). This includes the case where \( f \) is differentiable at \( x \) and \( f_x(x) = 0 \). Let \( K \) be the set of kink points of \( v^{-n} \). By C1, \( |K| \leq N - 1 \).

**Lemma 15** Any maximum of \( r \) is in \( R = [\bar{\theta}, \theta^x] \times [\theta^x, \bar{\theta}] \).

**Proof** Consider any \((\theta_1, \theta_h)\) with \( \theta_h < \theta_x \), and assume \((\theta_1, \theta_h)\) is a maximum of \( r \). Then, \( \bar{v}(\theta_h) = v^{-n}(\theta_h) \) by Lemma 5, and so, since \( \kappa < 1 \), it follows from using stacking and the definition of \( \theta^x \) that \( a^{-n} < \gamma(\cdot, 1) \) for \( \theta < \theta^x \), and so \( v \) crosses \( v^{-n} \) from below at \( \theta_h \), contradicting the definition of \( \theta_h \). Thus, \( \theta_h \geq \theta_x \), and similarly, \( \theta_1 \leq \theta_x \).

**Lemma 16** For each \( \kappa \in [0, 1] \), \( z(\cdot, \cdot, \kappa) \) is strictly increasing on \( R \), with \( z(\theta_1, \theta^x, \kappa) < 0 \) for all \( \theta_1 < \theta^x \).

**Proof** On each \( \tilde{R} \), \( z_{\theta_h}(\theta_1, \theta_h, \kappa) = a^{-n}(\theta_h) - \gamma(\theta_h, \kappa) = \theta_h - \theta^x \), and so \( z(\theta_1, \cdot, \kappa) \) is strictly single-peaked with minimum at \( \theta^x \). But, \( z(\theta_1, \theta_1, \kappa) = 0 \), and hence \( z(\theta_1, \theta^x, \kappa) < 0 \) for all \( \theta_1 < \theta^x \). Similarly \( z_{\theta}(\theta_1, \theta_h, \kappa) = -a^{-n}(\theta_1) + \gamma(\theta_1, \kappa) > 0 \) for \( \theta_1 < \theta^x \).

Note that \( z \) is differentiable on each \( \tilde{R} \). Hence, since \( z_{\theta_1}, z_{\theta_h}, \) and \( z_{\kappa} \) are strictly positive, \( L_N \) is continuous, strictly decreasing, and by definition of \( z \), goes through \((\theta^x, \theta^x)\). The locus \( L_S \), which lies below \( L_N \), has the same properties.

**Assumption 1 (Binding at the Top)** \( z(\bar{\theta}, \bar{\theta}, 1) \geq 0 \).

This formalizes that \( L_S \) hits the western boundary of \( R \). Under Assumption 1 at any maximum of \( r \), we have \( v(\theta_h) = v^{-n}(\theta_h) \). Define \( \theta_T \) by \( z(\theta, \theta_T, 0) = 0 \) if there is such a \( \theta_T \geq \theta^x \), and by \( \theta_T = \bar{\theta} \) otherwise. For Firm 1, \( \theta^x = \bar{\theta} \), and hence \( \theta_T = \bar{\theta} \). Let \( A = \{ (\bar{\theta}, \theta_h) | \theta_h \geq \theta_T \} \). By Lemmas 1 and 5, any maximum of \( r \) occurs in \( \Theta \cup A \), and so \( \max_{\theta_h} \psi(\theta_h) = \max_{\Theta \cup A} r(\theta_1, \theta_h) \).

Let us begin by considering the maximization problem as one moves east to west.

**Lemma 17** Fix \( \theta_h \in D \). Then on \( \Theta(\theta_h) \), \( r(\cdot, \theta_h) \) is strictly single-peaked where it is positive and has a unique maximum \( \lambda(\theta_h) \).

**Proof** This is trivial for \( \theta_h > \theta_T \), since \( \Theta(\theta_h) = \{ \bar{\theta} \} \). Fix \( \theta_h \leq \theta_T \). Let the (closed) interval \( \Theta(\theta_h) \) be denoted \([\tau_1, \tau_h] \). Existence of a maximum follows since \( r(\cdot, \theta_h) \) is continuous in \( \theta_1 \). Let

\[ 16 \text{If } \theta_h = \bar{\theta} \text{ and } v(\bar{\theta}) > v^{-n}(\bar{\theta}), \text{ then by Assumption 1 } v(\bar{\theta}) > v^{-n}(\theta_1) \text{ as well, contradicting Lemma 5.} \]
\( \tilde{\theta} \in [\tau, \tau_h] \) be such that \( \pi_a(\theta_l, \gamma(\theta_l, \kappa(\theta_l, \theta_h)), \tilde{v}) \) is strictly positive for \( \theta_l \in [\tau, \tilde{\theta}) \) and strictly negative for \( \theta_l \in (\tilde{\theta}, \tau_h] \). This is well-defined using the second part of Lemma 13.

Consider any \( \theta_l \in (\tilde{\theta}, \tau_h] \) at which \( r(\theta_l, \theta_h) > 0 \). By Lemma 13, \( \pi_a(\theta, \gamma(\theta, \kappa), \tilde{v}) < 0 \) for all \( \theta \in (\theta_l, \theta_h] \). Since \( \kappa \in [0, 1], \gamma_\theta > 0 \), and so, by (11), \( \pi(\cdot, \gamma(\cdot, \kappa), \tilde{v}) \) is strictly decreasing, and so since \( r(\theta_l, \theta_h) > 0 \), we have \( \pi(\theta_l, \gamma(\theta_l, \kappa), \tilde{v}) > 0 \). Thus, from (22), if \( \theta_l \notin K \), then

\[
r_{\theta_l}(\theta_l, \theta_h) \leq -\pi(\theta_l, \gamma(\theta_l, \kappa), \tilde{v}) < 0,
\]

and if \( \theta_l \in K \), then the same is true for each of \( r_{\theta_l}^-(\theta_l, \theta_h) \) and \( r_{\theta_l}^+(\theta_l, \theta_h) \). Thus, to the right of \( \tilde{\theta} \), \( r(\cdot, \theta_h) \) is decreasing anywhere it is positive.

Consider \( \theta_l \in [\tau, \tilde{\theta}) \). If \( \theta_l \notin K \) and \( \theta_l \) is a critical point, then \( r_{\theta_l} = 0 \) and we claim \( r_{\theta_l, \theta_h} < 0 \) whether or not \( r \) is positive. This holds since \( \theta_l \leq \tilde{\theta} \) and thus \( \pi_a(\theta_l, \gamma(\theta_l, \kappa), \tilde{v}) \geq 0 \), and so the first inequality in (26) holds regardless of the sign of \( r \). Thus \( \theta_l \) is a local maximum.

Assume \( \theta_l \in K \), and that \( \theta_l \) is a critical point. Then, since \( \pi_a(\theta_l, \gamma(\theta_l, \kappa), \tilde{v}) \geq 0 \) and since \( a^{-n} \) jumps upwards at \( \theta_l \), we have that \( r_{\theta_l}^-(\theta_l, \theta_h) \geq r_{\theta_l}^+(\theta_l, \theta_h) \). If \( r_{\theta_l}^- < 0 \) then \( r_{\theta_l}^+ < 0 \), contradicting that \( \theta_l \) is a critical point. Thus, \( r_{\theta_l}^- \geq 0 \). Assume first that \( r_{\theta_l}^- > 0 \). Then, \( r(\theta_l, \theta_h) > r(\theta_l', \theta_h) \) for all \( \theta_l' \) in a neighborhood to the left of \( \theta_l \). Assume instead that \( r_{\theta_l}^- = 0 \). Then, by (26) applied to the rectangle \( \tilde{R} \) to the left of \( (\theta_l, \theta_h) \), \( r(\cdot, \theta_h) \) is strictly concave on a neighborhood to the left of \( \theta_l \), and this again holds independent of the sign of \( r \), and so, since \( (r|_{\tilde{R}})_{\theta_l} = r_{\theta_l}^- = 0 \), \( r(\cdot, \theta_h) \) is strictly increasing on that neighborhood. Thus, again, \( r(\theta_l, \theta_h) > r(\theta_l', \theta_h) \) for all \( \theta_l' \) in a neighborhood to the left of \( \theta_l \). Arguing similarly, \( r(\theta_l, \theta_h) > r(\theta_l', \theta_h) \) for all \( \theta_l' \) in a neighborhood to the right of \( \theta_l \). Thus \( \theta_l \) is again a local maximum.

It follows that \( r(\cdot, \theta_h) \) is strictly single-peaked on \( [\tau, \tilde{\theta}] \), and hence has a single optimum on \( [\tau, \tilde{\theta}] \). Since to the right of \( \tilde{\theta} \), \( r(\cdot, \theta_h) \) is strictly decreasing anywhere that it is positive, this optimum remains the optimum on \( [\tau, \tau_h] \), and \( r(\cdot, \theta_h) \) is single-peaked where it is positive.

**Lemma 18** On or below \( L_S \), \( r_{\theta_h}(\theta_l, \theta_h) \) exists and equals \( \pi(\theta_h, \gamma(\cdot, H(\theta_h)), v^{-n})h(\theta_h) \), and if \( r(\theta_l, \theta_h) > 0 \), then \( r_{\theta_h}(\theta_l, \theta_h) > 0 \). Similarly, on or above \( L_N \), \( r_{\theta_h}(\theta_l, \theta_h) \) exists and equals \( -\pi(\theta_l, \gamma(\cdot, H(\theta_l)), v^{-n})h(\theta_l) \), and if \( r(\theta_l, \theta_h) > 0 \), then \( r_{\theta_h}(\theta_l, \theta_h) < 0 \).

**Proof** Fix \( (\theta_l, \theta_h) \) below \( L_S \). Then, \( z(\theta_l, \theta_h, H(\theta_h)) < 0 \), and so by definition, \( \kappa(\theta_l, \theta_h) = H(\theta_h) \), and by Lemma 7, \( \tilde{v}(\theta_l) = v^{-n}(\theta_l) \), and thus \( \tilde{v}(\theta_h) > v^{-n}(\theta_h) \). But then, by the Envelope Theorem (or by manipulation involving Lemma 8), it follows that \( r_{\theta_h}(\theta_l, \theta_h) = \pi(\theta_l, \gamma(\cdot, H(\theta_h)), \tilde{v})h(\theta_h) \).

Consider next \( (\theta_l, \theta_h) \in L_S \). Since for each \( \varepsilon > 0 \), \( (\theta_l, \theta_h - \varepsilon) \) is below \( L_S \), \( r_{\theta_h}(\theta_l, \theta_h - \varepsilon) = \pi(\theta_l, \gamma(\cdot, H(\theta_h - \varepsilon)), \tilde{v})h(\theta_h - \varepsilon) \), it follows as in (23) that \( r_{\theta_h}(\theta_l, \theta_h) = \pi(\theta_l, \gamma(\cdot, H(\theta_h)), v^{-n})h(\theta_h) \), where we note that on \( L_S \), \( \tilde{v}(\theta_h) = v^{-n}(\theta_h) \). Finally, from (21) and the discussion immediately.
following Remark\textsuperscript{2} and again exploiting that immediately above \( L_S \), \( \hat{\nu}(\theta_h) = v^{-n}(\theta_h) \),

\[
r^+_{\theta_h}(\theta_l, \theta_h) = \lim_{\varepsilon \downarrow 0} r_{\theta_h}(\theta_l, \theta_h + \varepsilon)
= \lim_{\varepsilon \downarrow 0} (\pi_a(\theta_h + \varepsilon, \gamma(\cdot, H(\theta_h + \varepsilon)), v^{-n}) (a^{-n}(\theta_h + \varepsilon) - \gamma(\theta_h + \varepsilon, H(\theta_h + \varepsilon))) h(\theta_h + \varepsilon)
+ \lim_{\varepsilon \downarrow 0} (\pi(\theta_h + \varepsilon, \gamma(\cdot, H(\theta_h + \varepsilon)), v^{-n})) h(\theta_h + \varepsilon)
= \pi(\theta_h, \gamma(\cdot, H(\theta_h)), v^{-n}) h(\theta_h).
\]

This follows since \( a^{-n}(\cdot) - \gamma(\cdot, H(\theta_h)) \) is bounded and since \( \lim_{\varepsilon \downarrow 0} \pi_a(\theta_h + \varepsilon, \gamma(\cdot, H(\theta_h)), v^{-n}) = 0 \) using that \( \gamma \) and \( v^{-n} \) are continuous and that on \( L_S \), \( \hat{\kappa} = H(\theta_h) \), and hence

\[
\pi_a(\theta_h, \gamma(\cdot, H(\theta_h)), v^{-n}) = \frac{H(\theta_h) - H(\theta_h)}{h(\theta_h)} = 0
\]

by definition of \( \gamma \). But then, since \( r^+_{\theta_h}(\theta_l, \theta_h) = r^-_{\theta_h}(\theta_l, \theta_h), r_{\theta_h}(\theta_l, \theta_h) \) exists and has the claimed value. If \( r(\theta_l, \theta_h) > 0 \), then, since \( \hat{\kappa} = H(\theta_h) \), we have by the usual argument using (11) that \( \pi(\theta_h, \gamma(\cdot, H(\theta_h)), \hat{\nu}) > 0 \), and hence \( r_{\theta_h}(\theta_l, \theta_h) > 0 \). The proof for \( (\theta_l, \theta_h) \) above \( L_N \) is similar. \( \Box \)

Our next lemma shows that the set of \( \theta_h \) where profits are strictly positive is an interval.

**Lemma 19** The set \( D = \{ \theta_h > \theta^x | \psi(\theta_h) > 0 \} \) is an interval.

**Proof** Let \( E \) be the set on which \( \theta > \theta^x \) and \( v_*(\theta) = B(\alpha_*(\theta)) - c(\alpha_*(\theta), \theta) > v^{-n}(\theta) \). Our first step is to show that \( E \subseteq D \). Choose \( \theta_h \in E \), set \( \alpha \) constant at \( \alpha_*(\theta_h) \) and set \( v(\theta_h) = v^{-n}(\theta_h) \). Then, by stacking and C1, \( (\alpha, v) \) hires some non-empty interval of types \( (\hat{\theta}_l, \theta_h) \), and earns \( v_*(\theta_h) - v^{-n}(\theta_h) > 0 \) on each type, since profits are constant by (11). A fortiori, \( r(\hat{\theta}_l, \theta_h) > 0 \).

Let us first see that \( \hat{\theta}_l \geq \tau_l \), where \( \tau_l \) is the left endpoint of the interval \( \Theta(\theta_h) \). If \( \tau_l = \underline{\theta}_l \), this is trivial, and if \( \tau_l > \underline{\theta}_l \), then, by definition of \( \Theta(\theta_h) \),

\[
0 = z(\tau_l, \theta_h, H(\theta_h))
= v^{-n}(\theta_h) - v^{-n}(\tau_l) - \int_{\tau_l}^{\theta_h} \gamma(\tau, H(\theta_h)) d\tau
> v^{-n}(\theta_h) - v^{-n}(\tau_l) - \int_{\tau_l}^{\theta_h} \gamma(\theta_h, H(\theta_h)) d\tau.
\]

But, the last expression is increasing in \( \tau_l \) using C1 and stacking, and since \( \gamma(\theta_h, H(\theta_h)) = \alpha_*(\theta_h) \), it is equal to 0 when \( \tau_l \) is replaced by \( \hat{\theta}_l \). Hence we have \( \hat{\theta}_l \geq \tau_l \).

Assume that \( \hat{\theta}_l > \tau_h \), where \( \tau_h \) is the right endpoint of the interval \( \Theta(\theta_h) \). Since for any \( \theta_l > \tau_h \) with \( r(\theta_l, \theta_h) > 0 \), we have \( r_{\theta_h}(\theta_l, \theta_h) < 0 \) by Lemma\textsuperscript{18} it follows by transfinite induction that \( r(\tau_h, \theta_h) > 0 \), and hence \( \theta_h \in D \). Hence \( E \subseteq D \).
By stacking and $C1$, $v^{-n}$ is strictly steeper than $v_{\lambda}$ above $\theta$. Hence $E$ is an interval $(\theta, \bar{E})$. Consider any $\theta_h \in D$, with $\theta_h > \bar{E}$. Then, $\pi(\theta_h, \gamma(\cdot, \tilde{\kappa}), \tilde{v}) < 0$, and so, since $\psi(\theta_h) > 0$, there is $\theta < \theta_h$ such that

$$0 > (\pi(\theta, \gamma(\cdot, \tilde{\kappa}), \tilde{v})) = \pi(\theta_h, \gamma(\cdot, \tilde{\kappa}), \tilde{v})$$

where the equality of sign uses (11). Thus $\tilde{c} < H(\theta_h)$ and so $(\lambda(\theta_h), \theta_h)$ is strictly above $L_S$. But then, for small $\epsilon > 0$, $(\lambda(\theta_h) - \epsilon)$, and thus

$$\psi_{\theta_h}(\theta_h) = \lim_{\epsilon \downarrow 0} \frac{r(\lambda(\theta_h), \theta_h) - r(\lambda(\theta_h) - \epsilon, \theta_h - \epsilon)}{\epsilon}$$

$$\leq \lim_{\epsilon \downarrow 0} \frac{r(\lambda(\theta_h), \theta_h) - r(\lambda(\theta_h), \theta_h - \epsilon)}{\epsilon}$$

$$= r_{\theta_h}(\lambda(\theta_h), \theta_h).$$

But, since $\tilde{c} < H(\theta_h)$, $\pi(\theta_h, \gamma(\cdot, \tilde{\kappa}), \tilde{v}) < 0$, and thus by (21), $r_{\theta_h} < 0$ and hence $\psi_{\theta_h}(\theta_h) < 0$. Thus an interval to the left of $\theta_h$ is also in $D$. Since $\theta_h$ was an arbitrary point above $\bar{E}$ at which $\psi > 0$, it follows by transfinite induction that $D$ is an interval as claimed. \hfill \Box

**Lemma 20** On $D$, $\lambda$ and $\psi$ are continuous.

**Proof** Since $L_N$ and $L_S$ are strictly decreasing and continuous, the correspondence $\Theta(\cdot)$ is nonempty, compact-valued, and continuous, and so by the Theorem of the Maximum, $\psi$ is continuous, and the set of maximizers of $r(\cdot, \theta_h)$ is upper hemicontinuous in $\theta_h$. But then, since $\lambda$ is single-valued on $D$ by Lemma 17, it is continuous as a function. \hfill \Box

Since $r_{\theta_l}(\theta_l, \theta_h) < 0$ on $L_N$, $(\lambda(\theta_h), \theta_h)$ is never on $L_N$. But, because we maximize first with respect to $\theta_l$, it may well be that $(\lambda(\theta_h), \theta_h) \in L_S$. Our next result shows that if $\lambda$ is on $L_S$, then $\psi$ is strictly increasing. Hence, no such point is a local maximum or minimum of $\psi$.

**Lemma 21** Let $(\lambda(\theta_h), \theta_h) \in L_S$ with $\theta_h \in D$. Then, $\psi_{\theta_h}(\lambda(\theta_h), \theta_h) > 0$, and $\psi_{\theta_l}(\lambda(\theta_h), \theta_h) > 0$, and hence $\psi$ is not critical at $\theta_h$.

**Proof** Let $(\lambda(\theta_h), \theta_h) \in L_S$ with $\theta_h \in D$. Since $K$ is finite, there is $\delta > 0$ such that $(\theta_h - \delta, \theta_h) \cap K = \emptyset$, such that $(\theta_l, \theta_l + \delta) \cap K = \emptyset$, and such that $\tilde{c}(\theta_l + \delta, \theta_h) > H(\theta_l)$, so that all of $(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)$ lies below $L_N$. From Lemma 11, $r_{\theta_h}$ is continuous on $(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)$, and by examination of (21) and by Lemma 18, it follows that $r_{\theta_h}$ is continuous on $X = \{(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)\} \cup \{(\theta_l, \theta_h)\}$. Further, since $\theta_h \in D$, $r(\lambda(\theta_h), \theta_h) > 0$, and so by Lemma 18, $r_{\theta_h}(\lambda(\theta_h), \theta_h) > 0$.  

47
Note next that for each \((\lambda(\theta'_h), \theta'_h) \in X \cap \Theta,\)

\[
\psi^+_{\theta_h}(\theta'_h) = \lim_{\varepsilon \downarrow 0} \frac{\psi(\theta'_h + \varepsilon) - \psi(\theta'_h)}{\varepsilon} \\
\geq \lim_{\varepsilon \downarrow 0} \frac{r(\lambda(\theta'_h), \theta'_h + \varepsilon) - r(\lambda(\theta'_h), \theta'_h)}{\varepsilon} = r_{\theta_h}(\lambda(\theta'_h), \theta'_h),
\]

where the inequality follows since \(\lambda(\theta'_h)\) is feasible at \(\theta'_h + \varepsilon\). In particular, \(\psi^+_{\theta_h}(\theta'_h) > 0\).

Finally, consider \(\psi^-_{\theta_h}(\theta'_h).\) Fix \(0 < \rho < r_{\theta_h}(\lambda(\theta_h), \theta_h)\). Since \(r_{\theta_h}\) is continuous on \(X\), and using (29), there is \(\delta > 0\) such that for all \(\varepsilon \in (0, \delta), \psi^-_{\theta_h}(\theta_h - \varepsilon) > \rho.\) Fix \(\varepsilon < \delta.\) Then, by transfinite induction, \(\psi(\theta_h) - \psi(\theta_h - \varepsilon) \geq \varepsilon \rho,\) since \(\psi(\theta_h - \varepsilon) - \psi(\theta_h - \delta) \geq 0,\) and since for any \(\tau < \theta_h\) where \(\psi(\tau) - \psi(\theta_h - \varepsilon) \geq (\tau - (\theta_h - \varepsilon))\rho,\) the same holds on an interval to the right of \(\tau\) by the definition of a right derivative. Since \(\psi(\theta_h) - \psi(\theta_h - \varepsilon) \geq \varepsilon \rho\) and \(\varepsilon < \delta\) was arbitrary, \(\psi^-_{\theta_h} > \rho > 0.\)

Given this result, we can turn attention away from \(L_S.\) Let \(D' = \{\theta_h \in D|(\lambda(\theta_h), \theta_h) \notin L_S\}.\)

**Lemma 22** For all \(\theta_h \in D',\) \(\psi^+_{\theta_h}(\theta'_h) = r^+_{\theta_h}(\lambda(\theta_h), \theta_h)\) and \(\psi^-_{\theta_h}(\theta'_h) = r^-_{\theta_h}(\lambda(\theta_h), \theta_h).\)

**Proof** Let \(K_1 = (K \cap \lambda(D')) \cup \{\emptyset\}\) and \(K_2 = K \cap D'.\) Thus, \(r(\lambda(\cdot), \cdot)\) may be non-differentiable either because \(\theta_h \in K_2\) or \(\lambda(\theta_h) \in K_1.\) In \(K,\) \(K_1 = \{\emptyset, K_2\}\) and \(K_2 = \{k_2\}.\)

**Case 1** Consider first \(\theta_h \in D'\) such that \(\lambda(\theta_h) \notin K_1\) and \(\theta_h \notin K_2.\) Then, we are not on \(L_S\) by definition of \(D',\) and we are not on \(L_N\) since by Lemma 18 \(r_{\theta_h}(\theta_1, \theta_h) < 0\) on \(L_N.\) Thus, since \(\psi(\theta_h) = r(\lambda(\theta_h), \theta_h),\)

\[
\psi_{\theta_h}(\theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h),\]

(30)

by the Envelope Theorem.

**Case 2** For any given \(\theta_1 \in K_1,\) let \(J(\theta_1) = \min\{\theta_h|\lambda(\theta_h) = \theta_1\},\) and let \(\tilde{J}(\theta_1) = \max\{\theta_h|\lambda(\theta_h) = \theta_1\}.\) Let \(J(\theta_1) = (\tilde{J}(\theta_1), \tilde{J}(\theta_1)).\) Since \(\lambda\) is constant on \(J(\theta_1),\) if \(J(\theta_1)\) is non-empty, then for all \(\theta_h \in J(\theta_1)\backslash K_2,\) we have

\[
\psi_{\theta_h}(\theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h).\]

(31)

**Case 3** Consider next \(\theta_h \in ((\tilde{J}(\theta_1))_{\theta_1 \in K_1} \cup \{\tilde{J}(\theta_1)\}_{\theta_1 \in K_1}) \backslash K.\) Assume that \(\theta_h = \tilde{J}(\theta_1)\) for some \(\theta_t \in K_1\) (the other case is similar). Then, \(\psi_{\theta_h}(\theta'_h) = r_{\theta_h}(\lambda(\theta'_h), \theta'_h)\) for \(\theta'_h\) on a neighborhood above \(\theta_h\) by (31). For a neighborhood below \(\theta_h, \theta'_h \notin K,\) since \(K\) is finite, and \(\lambda(\theta'_h) \notin K\) by definition of \(\tilde{J}(\theta_1)\) and again since \(K\) is finite. Hence \(\psi_{\theta_h}(\theta'_h) = r_{\theta_h}(\lambda(\theta'_h), \theta'_h)\) by (30). But then, by continuity of \(r_{\theta_h}\) on these neighborhoods, and by continuity of \(\lambda, \psi_{\theta_h}(\theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h)\) as well.

---

17That is, \(\lambda(\theta'_h) \in \Theta(\theta'_h + \varepsilon),\) since \(L_S\) is decreasing and \((\theta_1, \theta_1 + \delta) \times (\theta_h - \delta, \theta_h)\) lies strictly below \(L_N.\)

18These correspond to the bottoms and tops of the vertical segments of the path in Fig. 3.
Case 4 Finally, consider \( \theta_h \in K_2 \). Since \( K \) is finite, on some neighborhood above \( \theta_h \), \( \psi_{\theta_h} = r_{\theta_h} \) by the previous cases, and \( \lambda \) is continuous, and so
\[
\psi_{\theta_h}^+(\theta_h) = \lim_{\varepsilon \downarrow 0} \frac{\psi(\theta_h + \varepsilon) - \psi(\theta_h)}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \psi_{\theta_h}(\theta_h + \varepsilon) = \lim_{\varepsilon \downarrow 0} r_{\theta_h}(\lambda(\theta_h + \varepsilon), \theta_h + \varepsilon) = r_{\theta_h}^+(\lambda(\theta_h), \theta_h),
\]
using L'Hôpital’s rule at the second inequality, and where to justify the last equality, we note from \[21\] that \( r_{\theta_h} \) does not depend on \( a^{-n}(\theta_l) \), and so it does not matter whether or not \( \lambda(\theta_h) \in K_1 \). Similarly, \( \psi_{\theta_h}^-(\theta_h) = r_{\theta_h}^-(\lambda(\theta_h), \theta_h) \).

Lemma 23 The function \( \psi \) is strictly single-peaked and thus has a unique maximum on \( D \).

Proof We will show first that if \( \theta_h \) is a critical point, then it is a strict local maximum of \( \psi \). By Lemma \[21\] any critical point of \( \psi \) is in \( D' \). We go through the same four cases as in Lemma \[22\].

Case 1 Consider first \( \theta_h \in D' \) such that \( \lambda(\theta_h) \notin K_1 \) and \( \theta_h \notin K_2 \). Then, since \( \bar{\theta} \in K_1 \), \( (\lambda(\theta_h), \theta_h) \in \Theta \), and so Lemma \[14\] applies, and thus, by the Implicit Function Theorem, \( \lambda_{\theta_h} = -r_{\theta_l \theta_h}/r_{\theta_l \theta_l} \), where by Lemma \[14\] \( r_{\theta_l \theta_l} < 0 \). Since \[30\] holds on a neighborhood of \( \theta_h \),
\[
\psi_{\theta_l \theta_h}(\theta_h) = r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h) + r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h) \\
= -\frac{(r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h))^2}{r_{\theta_l \theta_l}(\lambda(\theta_h), \theta_h)} + r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h) \\
= \frac{1}{r_{\theta_l \theta_l}(\lambda(\theta_h), \theta_h)} (r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h) r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h) - (r_{\theta_l \theta_h}(\lambda(\theta_h), \theta_h))^2).
\]
But then, if \( \theta_h \) is a critical point, so that \( \psi_{\theta_l}(\theta_h) = 0 \), then by Lemma \[14\] \( r_{\theta_l \theta_l} r_{\theta_l \theta_h} - r_{\theta_l \theta_l}^2 \theta_h > 0 \). Hence, \( \psi_{\theta_l \theta_h}(\theta_h) < 0 \), and \( \theta_h \) is a strict local maximum of \( \psi \).

Case 2 Consider next \( \theta_l \notin K_2 \) such that for some \( \theta_l \in K_1 \), \( \lambda(\theta_l) \in J(\theta_l) \). Then, since \( J(\theta_l) \backslash K_2 \) is open, \[31\] holds on a neighborhood of \( \theta_h \), and so, since \( \lambda \) is constant on \( J(\theta_l) \), \( \psi_{\theta_l \theta_h}(\theta_h) = r_{\theta_l \theta_h}(\lambda(\theta_l), \theta_h) \). If \( \theta_h \leq \theta_T \), so that \( (\lambda(\theta_h), \theta_h) \in \Theta \), then by Lemma \[14\] if \( \psi_{\theta_h}(\theta_h) = 0 \), then \( \psi_{\theta_l \theta_h}(\theta_h) < 0 \), so \( \theta_h \) is a strict local maximum of \( \psi \). Assume that \( \theta_h \geq \theta_T \), so that \( \lambda(\theta_h) = \bar{\theta} \) and \( \kappa(\lambda(\theta_h), \theta_h) = 0 \). Trace the proof of Lemma \[14\] up through \[25\] with \( \bar{\theta} \) replaced by \( 0 \), and note that this part of the proof relies on \( \hat{v}(\theta_l) = c^{-n}(\theta_l) \) but not on \( \hat{v}(\theta_l) = c^{-n}(\theta_l) \). It follows that where \( r_{\theta_h}(\bar{\theta}, \theta_h) = 0 \), \( \psi(\bar{\theta}, \theta_h) = r_{\theta_l \theta_h}(\bar{\theta}, \theta_h) < 0 \), and again, \( \theta_h \) is a strict local maximum of \( \psi \).

Case 3 Consider next \( \theta_h \in (\{J(\theta_l)\}_{\theta_l \in K_1} \cup \{J(\theta_l)\}_{\theta_l \in K_1}) \backslash K \). Assume that \( \theta_h = J(\theta_l) \) for some \( \theta_l \in K_1 \) (the other case is similar), and assume that \( \psi_{\theta_h}(\theta_h) = 0 \). Then by Case 2, \( \psi \) is strictly concave on a neighborhood just above \( \theta_h \), while by Case 1, \( \psi \) is strictly concave on a neighborhood just below \( \theta_h \). Hence, again, \( \theta_h \) is a strict local maximum of \( \psi \).

Case 4 Finally, consider \( \theta_h \in K_2 = K \cap D' \). Since \( \kappa \in [0, 1] \), and since \( \theta_h > \theta^r \), we have \( a^{-n} - \gamma > 0 \) on a neighborhood of \( \theta_h \) by stacking and \( C1 \). At any point \( \theta_h \) of continuity of \( a^{-n} \),
and repeating (21) for convenience,

\[ \frac{\psi_{\theta_h}(\theta'_h)}{h(\theta'_h)} = r_{\theta_h}(\lambda(\theta'_h), \theta'_h) = \pi(\theta'_h, \gamma(\cdot, \kappa), v^{-n}(\theta'_h)) + \pi_a(\theta'_h, \gamma(\cdot, \kappa), v^{-n}(\theta'_h)) \left( a^{-n}(\theta'_h) - \gamma(\theta'_h, \kappa) \right), \]

where we recall that \( \kappa \) is continuous, and hence so is \( \gamma^n(\cdot, \kappa) \), and that \( v^{-n} \) is also continuous.

Thus any discontinuity in \( \psi_{\theta_h} \) at \( \theta_h \) is driven by an upward jump of \( a^{-n} \).

There are four cases to consider.

1. \( \pi \leq 0 \) and \( \pi_a < 0 \) at \( \theta_h \). Then, by (21) both \( \psi^+_{\theta_h}(\theta_h) \) and \( \psi^-_{\theta_h}(\theta_h) \) are strictly negative, and \( \theta_h \) is not a critical point.

2. \( \pi \leq 0 \) and \( \pi_a \geq 0 \) at \( \theta_h \). Then, since by Lemma 13 \( \pi_a \) strictly single-crosses zero from above, \( \pi_a \) is strictly positive everywhere on \( [\lambda(\theta_h), \theta_h) \) and so, \( \pi < 0 \) on \( [\lambda(\theta_h), \theta_h) \). Thus, \( \kappa(\lambda(\theta_h), \theta_h) < 0 \), and so \( \theta_h \notin D \), a contradiction.

3. \( \pi > 0 \) and \( \pi_a \geq 0 \) at \( \theta_h \). Then, by (21) both \( \psi^+_{\theta_h} \) and \( \psi^-_{\theta_h} \) are strictly positive, and so again \( \theta_h \) is not critical.

4. \( \pi > 0 \) and \( \pi_a < 0 \) at \( \theta_h \). Then, since \( a^{-n} \) jumps up at \( \theta_h \), we have \( \psi^-_{\theta_h} > \psi^+_{\theta_h} \). Assume that \( \theta_h \) is a critical point, so that \( \psi^-_{\theta_h} = \psi^+_{\theta_h} = 0 \). If \( \psi^-_{\theta_h} > 0 > \psi^+_{\theta_h} \), then \( \theta_h \) is a strict local maximum of \( \psi \). If \( \psi^+_{\theta_h} = 0 \), then, first, \( \psi^-_{\theta_h} > 0 \), and, second, from the previous cases, \( \psi_{\theta_h, \theta_h} < 0 \) for all \( \theta \) on a neighborhood to the right of \( \theta_h \). Similarly if \( \psi^-_{\theta_h} = 0 \), then \( \psi^+_{\theta_h} < 0 \), and \( \psi_{\theta, \theta_h} < 0 \) for all \( \theta \) on a neighborhood to the left of \( \theta_h \). In each case \( \theta_h \) is again a strict local maximum of \( \psi \).

Thus, if \( \theta_h \in D \) is a critical point of \( \psi \), then \( \theta_h \) is a strict local maximum of \( \psi \). Since \( D \) is an interval, \( \psi \) is strictly single-peaked on \( D \), and so, since \( D \) is non-empty, \( \psi \) has a unique maximum, and any critical point of \( \psi \) is that maximum. \( \square \)

**Lemma 24** Let \( \theta^*_h \) be the unique maximizer of \( \psi \). Then, the unique maximizer of \( r \) is \( (\lambda(\theta^*_h), \theta^*_h) \).

**Proof** Let \( (\theta^{**}_l, \theta^{**}_h) \in \arg\max_{\{\theta_l, \theta_h\); \theta_h \geq \theta_l\}} r(\theta_l, \theta_h) \). Since \( D \) is non-empty, \( r(\theta^{**}_l, \theta^*_h) > 0 \), and hence \( \theta^{**}_l < \theta^*_l \) and \( \theta^*_h \in D \). By Lemma 18 \( (\theta^{**}_l, \theta^{**}_h) \in \Theta \cup A \), and so \( \theta^{**}_l \in (\theta^*_l, \theta^{**}_h) \) is a local maximum of \( \lambda(\theta^*_h) \). Hence by Lemma 17 \( \theta^{**}_l = \lambda(\theta^*_h) \). Since \( (\theta^{**}_l, \theta^{**}_h) \) is optimal and since the constraint \( \theta_h \geq \theta_l \) is slack, we must have \( r^{-n}_{\theta_h}(\lambda(\theta^*_h), \theta^*_h) \leq 0 \) and \( r^{-n}_{\theta_l}(\lambda(\theta^*_h), \theta^*_h) \geq 0 \). But then, by Lemma 22 \( \psi^+_{\theta_h}(\theta^*_h) \leq 0 \) and \( \psi^-_{\theta_h}(\theta^*_h) \geq 0 \), and so by Lemma 23 \( \theta^{**}_h = \theta^*_h \), and we are done. \( \square \)

**9.4 Proofs for Section 6.6**

**Proof of Theorem 4** Let \( \hat{s} \) satisfy stacking, \( PS \), \( IO \) and \( OB \). Fix \( n \) and let \( \hat{s}^n = (\hat{\alpha}, \hat{\nu}) \). By \( IO \), \( (\hat{\alpha}, \hat{\nu}) \) satisfies \( C1 \) on \( (\theta_l, \theta_h) \). But then, by \( IO \), if \( n < N \), then \( \pi_a(\theta_l, \hat{\alpha}, \hat{\nu}) < 0 \), and by \( C1 \) and stacking, \( a^{-n}(\theta_h) - \hat{\alpha}(\theta_h) > 0 \). Hence, by \( [10] \), \( \pi(\theta, \hat{\alpha}, \hat{\nu}) > 0 \). Similarly, \( \pi(\theta_l, \hat{\alpha}, \hat{\nu}) > 0 \) if \( n > 1 \). But then, since profits are strictly single-peaked with maximum at \( \theta_0 \) solving \( H(\theta_0) = \hat{\kappa}, \pi(\theta, \hat{\alpha}, \hat{\nu}) > 0 \) for all \( \theta \in [\theta_l, \theta_h] \). Thus \( \psi(\theta) < v_*(\theta) \), so that \( (\hat{\alpha}, \hat{\nu}) \) satisfies \( C2 \) on \( [\theta_l, \theta_h] \).
Let us first re-define \((\hat{\alpha}, \hat{\nu})\) outside of \([\theta_l, \theta_h]\) to satisfy C1 and C2 there as well. Set

\[
\alpha(\theta) = \begin{cases} 
\min \{\gamma(\theta, 0), \hat{\alpha}(\theta)\} & \theta < \theta_l \\
\hat{\alpha}(\theta) & \theta \in [\theta_l, \theta_h] \\
\max \{\gamma(\theta, 1), \hat{\alpha}(\theta)\} & \theta > \theta_h
\end{cases}
\]

and set

\[
v(\theta) = \hat{\nu}(\theta) + \int_{\theta_l}^{\theta} \alpha(\tau)d\tau
\]

for all \(\theta\). That is, modify \((\hat{\alpha}, \hat{\nu})\) such that actions and surplus are unchanged in \([\theta_l, \theta_h]\), and modified outside of \([\theta_l, \theta_h]\) to ensure that C1 holds. Note that \(\hat{\alpha}(\theta_h) = \gamma(\theta_h, \hat{\gamma}) \geq \gamma(\theta_h, 1)\), and so no discontinuity is introduced at \(\theta_h\), and similarly at \(\theta_l\). By stacking, it remains the case that \((\alpha, \nu)\) is single-dominant on \([\theta_l, \theta_h]\), and so, since \((\alpha, \nu)\) and \((\hat{\alpha}, \hat{\nu})\) agree on \([\theta_l, \theta_h]\), \((\alpha, \nu)\) and \((\hat{\alpha}, \hat{\nu})\) are essentially equivalent.

To show that C2 is now satisfied for \(\theta \notin [\theta_l, \theta_h]\), assume \((\theta_h, \bar{\theta})\) is non-empty (the argument for \([\theta, \theta_l]\) non-empty is the same). Where \(\alpha(\cdot) = \hat{\alpha}(\theta_h)\), \((\pi(\theta, \alpha, v))_\theta = \pi_a(\theta, \alpha, v)\alpha(\theta) = 0\) by \([11]\). Where \(\alpha(\cdot) = \gamma(\cdot, 1), (\pi(\theta, \alpha, v))_\theta = \pi_a(\theta, \gamma(\cdot, 1), v)\gamma_\theta(\theta, 1) \geq 0\), using that \(\gamma_\theta(\theta, 1) > 0\), that \(\gamma(\theta, 1) \leq \gamma(\theta, H(\theta)) = a_s(\theta)\), and that \(\pi\) is strictly concave in \(a\), and so \(\pi_a(\theta, \gamma(\cdot, 1), v) \geq 0\). Thus, \(\pi(\theta, \alpha, v) \geq \pi(\theta_h, \alpha, v) > 0\) for all \(\theta > \theta_h\), and so \(v(\theta) < v_s(\theta)\) and C2 is satisfied.

Construct the strategy profile \(s\) by performing the above process for each \(n\). Then \(OB\) continues to hold for all \(n\), since for each of \(n\)'s opponents, \(\hat{\alpha}\) and \(\alpha\) agree on \([\theta_l, \theta_h]\), and by the fact that both the modified action profiles and the original action profiles of \(n\)'s opponents are continuous. Let us show that \(s\) is a Nash equilibrium. Fix \(n \notin \{1, N\}\). Assume first that Assumption \([11]\) holds. By the argument in the first paragraph of this proof, \(\theta_h \in D\). Since by PS, \(\bar{\theta} < \theta_l < \theta_h < \bar{\theta}\), it follows that \(z(\theta_l, \theta_h, \bar{\kappa}(\theta_l, \theta_h)) = 0\), where \(\bar{\kappa}(\theta_l, \theta_h) \in (H(\theta_l), H(\theta_h))\) by IO, and so we have \(\theta_l \in \Theta(\theta_h)\). But then, since \(r_{\theta_l}(\theta_l, \theta_h) = 0\) by \(OB\), we must have \(\theta_l = \lambda(\theta_h)\) by Lemma \([17]\). But then, again by \(OB\), \(0 = r_{\theta_h}(\theta_l, \theta_h) = r_{\theta_h}(\theta_l, \lambda(\theta_h), \theta_h) = \psi_{\theta_h}(\theta_l), \) where the third equality is by Lemma \([22]\). Finally, since by Lemma \([23]\) \(\psi\) is strictly single-peaked on the interval \(D\), we have \(\theta_h = \theta_h^*\) by Lemma \([24]\). Thus, \(s^n\) is a best response to \(s^{-n}\) by Corollary \([2]\).

If Assumption \([1]\) fails, then recall from the end of Section 6.5 that \(\hat{\lambda}\) is the analogue to \(\lambda\). So, we argue first that \(\theta_h \in \hat{\Theta}(\theta_l)\), then by the analogue to Lemma \([17]\) that \(\theta_h = \hat{\lambda}(\theta_l)\), and then by the analogue to Lemma \([22]\) that \(0 = r_{\theta_l}(\theta_l, \theta_h) = r_{\theta_l}(\theta_l, \hat{\lambda}(\theta_l)) = \hat{\psi}_{\theta_l}(\theta_l)\). But then, since \(\hat{\psi}\) is strictly single-peaked on \(\hat{D}\), we have \(\theta_l = \hat{\theta}_l^*\), and again \(s^n\) is a best response to \(s^{-n}\).

Consider \(n = 1\). Then, \(\kappa^1 = 0\) by IO, and \(\theta_l = \bar{\theta}\) by PS. But, since \(\kappa^1 = 0\), and since \(\pi(\theta_l, \hat{\alpha}, \hat{\nu}) > 0\), it follows by the usual argument that \(\pi(\theta, \hat{\alpha}, \hat{\nu}) > 0\) for all \(\theta < \theta_h\). Hence, since \(\pi_a(\theta, \hat{\alpha}, \hat{\nu}) < 0\) for all \(\theta < \theta_h\), we have \(r_{\theta_h} < 0\), and so \(\bar{\theta} = \lambda(\theta_h)\). But then, by \(OB\),

\[
0 = r_{\theta_h}(\bar{\theta}, \theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h) = \psi_{\theta_h}(\theta_h) ,
\]
and again \( \theta_h = \theta_h^\ast \), and \( s^1 \) is a best response to \( s^{-1} \).

Finally, consider \( n = N \). Then, since \( \kappa^N = 1 \) by \( IO \), it similarly follows that \( \theta_h = \tilde{\theta} = \tilde{\lambda}(\theta_l) \).

Thus, by \( OB \),
\[
0 = r_{\theta_l}(\theta_l, \tilde{\theta}) = r_{\theta_l}(\theta_l, \tilde{\lambda}(\theta_l)) = \psi_{\theta_l}(\theta_l),
\]
and so \( \theta_l = \tilde{\theta}_l^\ast \), and \( s^N \) is a best response to \( s^{-N} \).

\[ \square \]

9.5 Proofs For Section 6.7

We begin by defining three further restrictions on strategies that will turn out not to bind in equilibrium, but that help us towards compactness and continuity.

Let \( BR(s^{-n}) = \arg \max_{s^n \in S^n} \Pi^n(s^n, s^{-n}) \). Let

\[
\eta = \max\{\gamma^N(\bar{\theta}, 0), \max_{n, \theta, \kappa \in [0, 1]} \gamma^0_n(\theta, \kappa)\}.
\]

Since \( \eta > -\infty \) since each relevant object is continuous and hence bounded on the compact choice set. We will see that anywhere that (10) holds, \( v^n(\bar{\theta}) \geq \beta \), motivating our next restriction.

\[ C3 \quad 0 \leq \alpha^n(\theta') - \alpha^n(\theta) \leq \eta(\theta' - \theta) \text{ for all } \theta, \theta' \text{ with } \theta' > \theta. \]

Next, let
\[
\beta = \min_{n, \theta, \kappa \in [0, 1]} \left( \pi^a_n(\theta_h, \gamma^0_n(\cdot, \kappa), 0) \gamma^N(\theta_h, 0) + \pi^a_n(\theta_h, \gamma^0_n(\cdot, \kappa), 0) \right)
\]
where \( \beta > -\infty \) since each relevant object is continuous and hence bounded on the compact choice set. We will see that anywhere that (10) holds, \( v^n(\bar{\theta}) \geq \beta \), motivating our next restriction.

\[ C4 \quad v^n(\bar{\theta}) \geq \beta. \]

Finally, we impose the innocuous condition that \( v \) and \( \alpha \) relate via the standard integral.

\[ C5 \quad v^n(\theta) = v^n(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} \alpha^n(\tau) d\tau \text{ for all } \theta. \]

For each \( n \), let \( S^n_R \) be the subset of \( S^n \) such that \( C1-C5 \) hold. Let \( S_R = \times_n S^n_R \), and \( S^{-n}_R = \times_{n' \neq n} S^n_{R \cdot} \).
Lemma 25  Fix $s^{-n} \in S_{R}^{-n}$. Then $BR^{n}(s^{-n}) \cap S_{R}^{n}$ is nonempty.

Proof of Lemma [25] Fix $n$. By Corollary [2] and using stacking, any $\hat{s}^{n} \in BR^{n}(s^{-n})$ is single-dominant on some region $[\theta_{1}, \theta_{2}]$, and has the form $(\hat{\alpha}, \hat{v})$, where $\hat{\alpha} = \gamma(\cdot, \kappa)$ on $[\theta_{1}, \theta_{2}]$, where $\kappa \in [H(\theta_{1}), H(\theta_{2})]$, and where $C1$ and $C2$ are satisfied on $[\theta_{1}, \theta_{2}]$. Let $(\alpha, v)$ be defined from $(\hat{\alpha}, \hat{v})$ as in the proof of Theorem [4] so that as shown there, $C1$ and $C2$ are satisfied on $[\theta_{1}, \theta_{2}]$. By the stacking condition and using that for $n' \neq n$, $C1$ and $C2$ are satisfied by assumption, it remains the case that $(\alpha, v)$ is single-dominant on $[\theta_{1}, \theta_{2}]$, and since $(\alpha, v)$ and $(\hat{\alpha}, \hat{v})$ agree on $[\theta_{1}, \theta_{2}]$, it follows that $(\alpha, v) \in BR(s^{-n})$. Conditions $C3$ and $C5$ hold by construction.

To show that $C4$ holds, assume by way of contradiction, that $v(\overline{\theta}) < \beta$. Then, since $v^{n'}(\overline{\theta}) \geq \beta$ for each $n' \neq n$, $\theta_{h} < \overline{\theta}$, and so by [21], if we let $\overline{a} = \lim_{\theta_{h} \uparrow \theta_{h}} a^{-n}(\theta_{h}^{'})$, then

$$0 \geq \frac{r_{\theta_{h}}^{n}(\theta_{h}, \theta_{h})}{h(\theta_{h})} = \pi_{a}(\theta_{h}, \gamma(\cdot, \kappa), v)(\overline{a} - \gamma(\theta_{h}, \kappa)) + \pi(\theta_{h}, \gamma(\cdot, \kappa), v).$$

But, since $s^{n}$ is a best response, it follows from Proposition [3] and continuity of $\pi$, $\gamma$, and $v$ that $\pi(\theta_{h}, \gamma(\cdot, \kappa), v) \geq 0$. By $C1$ and $C2$ for $n' \neq n$, and stacking, $\overline{a} - \gamma(\theta_{h}, \kappa) > 0$. Hence $\pi_{a}(\theta_{h}, \gamma(\cdot, \kappa), v) \leq 0$, and so we have

$$0 \geq \pi_{a}(\theta_{h}, \gamma(\cdot, \kappa), v) \gamma^{N}(\theta_{h}, 0) + \pi(\theta_{h}, \gamma(\cdot, \kappa), v)$$

$$> \pi_{a}(\theta_{h}, \gamma(\cdot, \kappa), 0) \gamma^{N}(\theta_{h}, 0) + \pi(\theta_{h}, \gamma(\cdot, \kappa), 0) - \beta$$

$$\geq 0,$$

where the first inequality uses $\overline{a} - \gamma(\theta_{h}, \kappa) \leq \overline{a} \leq \gamma^{N}(\theta_{h}, 0)$, the second uses the monotonicity of $v$, and the premise that $v(\overline{\theta}) < \beta$, and the last inequality follows from the definition of $\beta$. This is a contradiction, and hence $v(\overline{\theta}) \geq \beta$ as required. \hfill \Box

Proof of Theorem [5] Let us now prove that the game $(S^{n}, \Pi^{n})_{n=1}^{N}$ has a pure-strategy equilibrium. It is enough to show that $(S_{R}^{n}, \Pi^{n})_{n=1}^{N}$ has a pure-strategy equilibrium: By Lemma [25] $BR^{n}(s^{-n}) \cap S_{R}^{n}$ is nonempty, and so in a Nash equilibrium of $(S_{R}^{n}, \Pi^{n})_{n=1}^{N}$, each player is playing an element of $BR^{n}(s^{-n})$, and we have a Nash equilibrium of $(S_{R}^{n}, \Pi^{n})_{n=1}^{N}$.

The set of continuous functions from $[\theta, \overline{\theta}]$ to $\mathbb{R}$, endowed with the sup norm $\| \cdot \|_{\infty}$, is a Banach space, and thus $S_{R}^{n}$, with norm $\| (\alpha^{n}, v^{n}) \| = \| \alpha^{n} \|_{\infty} + \| v^{n} \|_{\infty}$ is a subset of a Banach space. Similarly $S_{R}$ with norm $\sum_{n} \| (\alpha^{n}, v^{n}) \|$ is a subset of a Banach space.

Let us show that for each $n$, the set $S_{R}^{n}$ is nonempty, convex, and compact. To see that $S_{R}^{n}$ is nonempty, we will argue that $(\alpha_{n}^{n}, v_{n}^{n}) \in S_{R}^{n}$. Note that $C5$ and $C2$ are immediate, and that $C1$ follows because $\alpha_{n}^{n}(\theta) = \gamma^{n}(\theta, H(\theta))$. But then,

$$\gamma_{\theta}^{n}(\theta, H(\theta)) + \gamma_{\theta}^{n}(\theta, H(\theta))h(\theta) < \gamma_{\theta}^{n}(\theta, H(\theta)) \leq \eta,$$
using that \(\gamma^n_k < 0\), and so C3 follows. To see C4, note that since \(\alpha^n_n(\bar{\theta}) = \gamma^n(\bar{\theta}, 1)\), it follows that 
\[
\pi^n_\alpha(\bar{\theta}, \gamma^n(\bar{\theta}, 1), 0) = B^n_\alpha(\gamma^n(\bar{\theta}, 1)) - c_\alpha(\gamma(\bar{\theta}, 1), \bar{\theta}) = 0, \text{ and hence }
\]
\[
\pi^n_\alpha(\bar{\theta}, \gamma^n(\bar{\theta}, 1), 0)\gamma^N(\theta_h, 0) + \pi^n(\bar{\theta}, \gamma^n(\bar{\theta}, 1), 0) = v^n_\alpha(\bar{\theta}),
\]
and thus \(\beta \leq v^n_\alpha(\bar{\theta})\). Thus, \(S^n_R\) is nonempty.

To prove convexity of \(S^n_R\), let \((\alpha^n_1, v^n_1)\) and \((\alpha^n_2, v^n_2)\) \(\in S^n_R\), let \(\delta \in [0, 1]\), and let \((\alpha^n_3, v^n_3) = (\delta \alpha^n_1 + (1 - \delta)\alpha^n_2, \delta v^n_1 + (1 - \delta)v^n_2)\). Then, \((\alpha^n_3, v^n_3)\) satisfies C5 since integration is a linear operator, and it is direct that \((\alpha^n_3, v^n_3)\) satisfies the inequalities required for C1–C4.

To prove compactness, let \((\alpha^n_1, v^n_1)\) \(\subset S^n_R\) be a sequence of elements of \(S^n_R\). Then, by C1 and the definition of \(\eta\), we have \(\alpha^n_\eta(\bar{\theta}) \geq 0\) and \(\alpha^n_\eta(\bar{\theta}) \leq \eta\). Hence, since C3 is satisfied by \(\alpha^n_\eta\) for each \(k\), it follows by the Arzela-Ascoli Theorem (e.g., [Rudin (1987), Theorem 11.28, p. 245]) that there exists \(\alpha^n\) satisfying C1 and C3 and a subsequence along which \(\|\alpha^n_k - \alpha^n\|_\infty \to 0\). Note that \(\alpha^n\) is increasing and has range contained in \([0, \eta]\), and so is integrable. Since \(v^n_k(\bar{\theta})\) lies in a compact set by C2 and C4, we can also, wlog, assume that along the chosen subsequence \(v^n_k(\bar{\theta}) \to \bar{v}\), for some \(\bar{v}\). Define \(v^n(\theta) = \bar{v} - \int_\theta^\bar{\theta} \alpha^n(\tau)d\tau\), so that C5 is satisfied by construction. We claim that (i) along the same subsequence, \(\|v^n_k - v^n\|_\infty \to 0\), and (ii) \((\alpha^n, v^n)\) \(\in S^n_R\). To see (i), note that by C5, for each \(\theta\) and \(k\), \(v^n_k(\theta) = v^n_k(\bar{\theta}) - \int_\theta^\bar{\theta} \alpha^n_\eta(\tau)d\tau\), and hence

\[
|v^n(\theta) - v^n_k(\theta)| = |v^n_k(\bar{\theta}) - \bar{v}| + \int_\theta^\bar{\theta} |\alpha^n_k(\tau) - \alpha^n(\tau)| d\tau
\]

and thus, since the last expression is independent of \(\theta\), \(\|v^n_k - v^n\|_\infty \to 0\). To see (ii), note that we have already checked C1, C3, and C5, and that weak inequalities are preserved under limits, and so C2 and C4 hold as well. Thus, \(S^n_R\) is sequentially compact, and since it is also a metric space, it is compact.

Since \(N\) is finite and, for each \(n\), \(S^n_R\) is nonempty, convex, and compact, so is the product \(S_R = \times_{i=1}^N S^n_i\). Fix \(s \in S_R\), let \(s_k \to s\), and fix \(n\). Then, by stacking and since \(s \in S_R\), there exist \(\theta_t\) and \(\theta_h\) such that \(\varphi(\theta, s) = 1\) on \((\theta_t, \theta_h)\) and \(\varphi(\theta, s) = 0\) for \(\theta \notin [\theta_t, \theta_h]\). But then, since for each \(n', \|v^n_\alpha' - v^n_s\| \to 0\), and again using stacking, for any given \(\delta > 0\), and for \(s'\) close enough to \(s\), \(\varphi(\theta, s') = 1\) on \([\theta + \delta, \theta - \delta]\) and \(\varphi(\theta, s') = 0\) for \(\theta \notin (\theta - \delta, \theta + \delta)\). Since \(\|\alpha^n_k - \alpha^n\| \to 0\) as well, and since \(\pi\) is bounded and continuous, it follows that \(\Pi^n(s_k) \to \Pi^n(s)\), and thus that \(\Pi^n\) is continuous on \(S_R\).\(^{[19]}\)

Fix \(n\). Since \(\Pi^n\) is continuous on \(S_R\), and since \(S^n_R\) is non-empty, compact, and independent of \(v^{−n}\) (and hence trivially continuous as a correspondence) the Theorem of the Maximum implies that \(BR^n_R(s^{−n}) = \arg\max_{s \in S^n_R} \Pi^n(s, s^{−n})\) is non-empty and compact valued for each \(s^{−n}\), and is

\(^{[19]}\)Recall that without stacking, and outside of \(S_R\), it is easy to construct examples where payoffs are discontinuous.
upper hemi-continuous in $s^{-n}$.

Finally, let us show that $BR_R^n(s^{-n})$ is convex. Let $s^n \in BR_R^n(s^{-n})$, with single-dominance region $(\hat{\theta}_l, \hat{\theta}_h)$. Then, by Corollary 2, $(\hat{\theta}_l, \hat{\theta}_h)$ maximizes $r$, and on $(\hat{\theta}_l, \hat{\theta}_h)$, $s^n = \hat{s}(\hat{\theta}_l, \hat{\theta}_h)$, and by Lemma 24, $(\hat{\theta}_l, \hat{\theta}_h) = (\lambda(\theta^*_h), \theta^*_h)$. Thus, any two elements of $BR_R^n(s^{-n})$ win for sure on $(\lambda(\theta^*_h), \theta^*_h)$ and agree with $\hat{s}(\lambda(\theta^*_h), \theta^*_h)$ on $(\lambda(\theta^*_h), \theta^*_h)$, and lose for sure for $\theta \notin (\lambda(\theta^*_h), \theta^*_h)$. But then, their convex combination does the same, and so is also a best response.

We have established that $S_R$ is a non-empty, compact, convex subset of a Banach space, and that the correspondence $BR_R(s) \equiv BR_R^1(s^{-1}) \times \cdots \times BR_R^N(s^{-N})$ from $S_R$ to $S_R$ has a closed graph and nonempty convex values. Thus, by the Kakutani-Fan-Glicksberg Theorem (Aliprantis and Border 2006, Corollary 17.55, p. 583) $BR_R$ has a fixed-point on $S_R$, and we are done. □
References


56


