Aggregative Oligopoly Games with Entry[1]

Simon P. Anderson[2], Nisvan Erkal[3] and Daniel Piccinin[4]

First version: November 2009
This version: June 2015

[1]We thank David Byrne, Chris Edmond, Maxim Engers, Daniel Halbheer, Joe Harrington, Simon Loertscher, Phil McCalman, Claudio Mezzetti, Volker Nocke, Martin Peitz, Frank Stähler, Jidong Zhou, and especially Richard Cornes for comments and discussion. We also thank seminar participants at the Federal Trade Commission, Johns Hopkins University, New York University, National University of Singapore, University of Mannheim, and conference participants at Université Catholique de Louvain (Conference in Honor of Jacques Thisse), EARIE (2010), Australasian Economic Theory Workshop (2011), North American Winter Meeting of the Econometric Society (2013) for their comments. Imogen Halstead, Boon Han Koh, Charles Murry, and Yiyi Zhou have provided excellent research assistance. The first author thanks National Science Foundation (0752923) for financial support and the Department of Economics at the University of Melbourne for its hospitality. The second author gratefully acknowledges funding from the Australian Research Council (DP0987070).

[2]Department of Economics, University of Virginia P.O. Box 400182, Charlottesville, VA 22904-4182, USA. sa9w@virginia.edu.

[3]Department of Economics, University of Melbourne, VIC 3010, Australia. n.erkal@unimelb.edu.au.

[4]Brick Court Chambers, 7-8 Essex Street, London, WC2R 3LD, United Kingdom. daniel.piccinin@brickcourt.co.uk.

Abstract

We compile an IO toolkit for aggregative games and use inclusive best reply functions to compare long-run market structures. We show strong neutrality properties across market structures. The IIA property of demands (CES and logit) implies that consumer surplus depends on the aggregate alone, and the Bertrand pricing game is aggregative. We link together the following results: merging parties' profits fall but consumer surplus is unchanged, monopolistic competition is the market structure with the highest surplus, consumer gains from trade are higher under oligopoly than monopolistic competition. The basic results are shown to extend to games with a sub-aggregative structure.

JEL Classifications: D43, L13

Keywords: Aggregative games; Oligopoly theory; Entry; Strategic substitutes and complements; IIA property; Mergers; Procompetitive effects of trade; Monopolistic competition; Logit/CES; Cournot; Nested Logit

1 Introduction

Many non-cooperative games in economics are aggregative games, where the players' payoff depends on their own action and an aggregate of all players' actions. Examples abound in industrial organization (oligopoly theory, R&D races), public economics (public goods provision games, tragedy of the commons), and political economy (political contests, conflict models), to name a few.[1] In this paper, we consider aggregative oligopoly games with endogenous entry, and compare alternative long-run market structures. Our analysis reveals the key drivers for many existing results, establishes fundamental links, and derives new results.

We compare alternative market structures, such as different objective functions (due to a merger or privatization), different timing of moves (due to leadership), or technological differences. We develop a simple general framework to analyze how the aggregate, producer surplus, and consumer surplus differ across market structures in a free entry equilibrium. Our analysis deploys the inclusive best reply concept introduced by Selten (1970), for which we derive the corresponding maximal profit function as a key tool to characterize the equilibrium.

We show strong neutrality properties across market structures. The aggregate stays the same in the long run. This is despite the fact that the affected firms' equilibrium actions and payoffs, and the number of active firms change, while the unaffected firms' equilibrium actions and payoffs remain unchanged. Thus, free entry completely undoes short-run effects on the aggregate.[2] This neutrality result extends to consumer surplus whenever consumer surplus depends on the aggregate only. We

[1] In oligopoly theory, a prominent example is Cournot oligopoly. Other commonly used models of logit, CES, and linear differentiated demand all fit in the class. Alos-Ferrer and Ania (2005) use aggregative games to provide an evolutionary foundation for the perfect competition paradigm. See Cornes and Hartley (2005, 2007a and 2007b) for examples in contests and public good games.

[2] See Corchón (1994 and 2001) and Acemoglu and Jensen (2013) for comparative statics results for aggregative games in the short run.

show that in Bertrand differentiated products models, consumer surplus is solely a function of the aggregate if and only if demands satisfy the IIA property. Then, the welfare difference is measured simply as the change in payoffs to the directly affected firm(s). Thus, all market structure differences which are privately beneficial are also socially beneficial, calling for a passive policy approach.

These neutrality results show the strong positive and normative implications of using an aggregative game structure, such as oligopoly with CES or logit demand, or Tullock contest game. This is important because these games are widely used in disparate fields. Outside of industrial organization, the CES model is central in theories of international trade (e.g., Helpman and Krugman, 1987; Melitz, 2003), endogenous growth (e.g., Grossman and Helpman, 1993), and new economic geography (e.g., Fujita et al., 2001; Fujita and Thisse, 2002). The logit model forms the basis of the structural revolution in empirical industrial organization. The Tullock contest game has been used in a number of fields, including the economics of advertising, innovation, conflict resolution, lobbying, and electoral competition.

The reason why these models are so popular is uncovered through recognizing them as aggregative games. The oligopoly problem in broad is complex: each firm's actions depend on the actions of all other firms. An aggregative game reduces the degree of complexity drastically to a simple problem in two dimensions. Each firm's action depends only on one variable, the aggregate, yielding a clean characterization of equilibria with asymmetric firms in oligopoly. This feature of aggregative games has been used (sometimes implicitly) by many authors to study existence and uniqueness of equilibria. See, for example, McManus (1962 and 1964), Selten (1970), Novshek (1984 and 1985), and Corchón (1994 and 2001). Our aim is to show the benefits of exploiting the aggregative structure in games with endogenous entry, an aspect that has not been explored in the literature so far.[3]

[3]Indeed, as is also emphasized by Shubik (1982, p. 325), more research needs to be done to

Our framework reveals the underpinning to several results in the literature. We consider mergers, monopolistic competition and international trade in the main text, and cost shocks, leadership, rent-seeking, research joint ventures and privatization in the Online Appendix. Exploiting the aggregative game structure directly yields more general and further results. Importantly, we link together the following results: (i) Merging parties' profits fall but consumer surplus is unchanged in the long run even though the merged parties' prices rise and more varieties enter; (ii) monopolistic competition is the market structure with the highest surplus; (iii) consumer gains from trade are underestimated under monopolistic competition as compared to oligopoly; (iv) Stackelberg leadership raises welfare; and (v) R&D cooperation by some firms has no impact on the long-run total rate of innovation even though cooperation encourages more firms to enter the race.

We show that the toolkit we develop applies more generally to games with a sub-aggregative structure. An important example is nested logit. The neutrality results continue to hold in such cases. Two crucial assumptions behind our neutrality results are that there are no income effects and that the marginal entrant type is the same across the market structures we compare. Both of these assumptions are commonly made in the literature. We show that if they are violated, we no longer get the stark predictions of the main model. With heterogeneous entrants, a beneficial cost shock experienced by a firm causes the aggregate to increase. With income effects, we show that a change that positively affects total profits increases the aggregate. Moreover, when profits are redistributed to consumers, their welfare rises if and only if the change increases total profits.

The rest of the paper proceeds as follows. In Section 2, we present the framework and provide the basic definitions. After defining our equilibrium concept in Section

establish the potential of aggregative games as a useful tool of equilibrium analysis. We take a step in this direction by focusing on games with endogenous entry.

3, we present our core comparative static results in Sections 4 and 5. In Sections 6, 7, and 8, we apply our results to mergers, monopolistic competition, and international trade. We then show how our basic framework can be extended to include sub-aggregative games in Section 9. We consider income effects and heterogeneous entrants in Sections 10 and 11, respectively. Section 12 concludes. Integer constraints and further applications are considered in the Online Appendix.

2 Preliminaries: The IO Aggregative Game Toolkit

We consider two-stage games where firms simultaneously make entry decisions in the first stage. Entry involves a sunk cost K_i for firm i. In the second stage, after observing which firms have entered, active firms simultaneously choose their actions.

2.1 Payoffs

Consider the second (post-entry) stage of the game. Let $\mathcal{S}$ be the set of active entrants. We consider aggregative oligopoly games in which each firm's payoffs depend only on its own action, $a_i \geq 0$, and the sum of the actions of all firms, the aggregate, $A = \sum_{i \in \mathcal{S}} a_i$. We write the (post-entry or gross) profit function as $\pi_i(A, a_i)$.

To illustrate, consider (homogeneous product) Cournot games, where $\pi_i = p(Q) q_i - C_i(q_i)$. The individual action is own output, $q_i = a_i$, and the aggregate is the sum of all firms' outputs, $Q = A$. Consumer surplus depends only on the price, $p(Q)$, so the aggregate is a sufficient statistic for tracking what happens to consumer welfare. In what follows, we shall refer to the case with log-concave (homogeneous products) demand, $p(Q)$, and constant marginal cost, $C_i(q_i) = c_i q_i$, as the Cournot model.

A more subtle example is Bertrand oligopoly with CES demands. The representative consumer's direct utility function in quasi-linear form is $U = \frac{1}{\rho} \ln \left(\sum_{i \in \mathcal{S}} x_i^{\rho} \right) + X_0$, where X_0 denotes numeraire consumption and x_i is consumption of differentiated variant i. Hence, $\pi_i = (p_i - c_i) \frac{p_i^{-\lambda-1}}{\sum_j p_j^{-\lambda}}$ with $\lambda = \frac{\rho}{1-\rho}$. The denominator - the "price index" - constitutes the aggregate. It can be written as the sum of individual firm's choices by defining $a_j = p_j^{-\lambda}$ so that we can think of firms as choosing the values a_j, which vary inversely with prices p_j, without changing the game. Then we write $\pi_i = \left(a_i^{-1/\lambda} - c_i\right) \frac{a_i^{(\lambda+1)/\lambda}}{A}$ and call the function mapping primal price choices to the aggregate value the *aggregator function*.[4] Strategic complementarity of prices implies strategic complementarity of the a's.

Similarly, for Bertrand oligopoly with logit demands, $\pi_i = (p_i - c_i) \frac{\exp[(s_i - p_i)/\mu]}{\sum_{j=0}^{n} \exp[(s_j - p_j)/\mu]}$, where the s_j are "quality" parameters, the p_j are prices, and $\mu > 0$ represents the degree of preference heterogeneity. The "outside" option has price 0. Again, the aggregator function derives from thinking about the firms as choosing $a_j = \exp\left[(s_j - p_j)/\mu\right]$. The denominator in the profit function is the aggregate, so we write $\pi_i = (s_i - \mu \ln a_i - c_i) \frac{a_i}{A}$.[5]

Let $A_{-i} = A - a_i$ be the total choices of all firms in S other than i. Then we can write i's profit function in an aggregative oligopoly game as $\pi_i\left(A_{-i} + a_i, a_i\right)$ and we normalize $\pi_i\left(A_{-i}, 0\right)$ to zero. Assume that each firm's strategy set is compact and convex.[6] Let $r_i\left(A_{-i}\right) = \arg\max_{a_i} \pi_i\left(A_{-i} + a_i, a_i\right)$ denote the standard best reply (or reaction) function. We define $\bar{A}_{-i}$ as the smallest value of A_{-i} such that $r_i\left(\bar{A}_{-i}\right) = 0$.

[4] Cornes and Hartley (2012) show that the aggregative structure may be exploited in any game as long as there exists an additively separable aggregator function which ensures that the interaction between players' choices is summarized by a single aggregate not only in the payoff functions, but also in the marginal payoff functions. See also Jensen (2010). More general classes of aggregative games have been proposed in Jensen (2010) and Martimort and Stole (2012).

[5] In these examples, even though payoffs are a function of the aggregate, consumer welfare does not have to be. Where it is, the aggregative structure of the game can be exploited to dramatically simplify the consumer welfare analysis. We show in Section 5 that this is the case in Bertrand differentiated product games where the demand functions satisfy the IIA property (such as CES and logit).

[6] We can bound actions by ruling out outcomes with negative payoffs. In the Cournot model, we rule out outputs where price must be below marginal cost by setting the maximum value of q_i as the solution to $p\left(q_i\right) = c_i$.

Assumption A1 (Competitiveness) $\pi_i(A_{-i} + a_i, a_i)$ strictly decreases in A_{-i} for $a_i > 0$.

This competitiveness assumption means that firms are hurt when rivals choose larger actions. It also means that $\pi_i(A, a_i)$ is decreasing in A (for given a_i). The aggregator functions we use for Bertrand games vary inversely with price, so competitiveness applies there too.

A1 implies that firms impose negative externalities upon each other. Hence, it rules out games with positive externalities, such as the public goods contribution game (see, e.g., Cornes and Hartley, 2007a and 2007b). However, in such games, it is often not relevant to have a free-entry condition closing the model.

Assumption A2 (Payoffs)

a) $\pi_i(A_{-i} + a_i, a_i)$ is twice differentiable, and strictly quasi-concave in a_i, with a strictly negative second derivative with respect to a_i at any interior maximum.

b) $\pi_i(A, a_i)$ is twice differentiable, and strictly quasi-concave in a_i, with a strictly negative second derivative with respect to a_i at any interior maximum.

A2a is standard, and takes as given the actions of all other players while A2b takes as given the aggregate.[7] A2a implies a continuous best response function $r_i(A_{-i})$ which is differentiable and solves

$$\frac{d\pi_i(A_{-i} + a_i, a_i)}{da_i} = \pi_{i,1}(A_{-i} + a_i, a_i) + \pi_{i,2}(A_{-i} + a_i, a_i) = 0, \qquad i \in \mathcal{S}, \tag{1}$$

for interior solutions, where $\pi_{i,j}(.)$, $j = 1, 2$, refers to the partial derivative with respect to the jth argument.

[7] To see that there is a difference between A2a and A2b, consider Cournot competition with $\pi_i = p(Q)q_i - C(q_i)$, and consider the stronger assumption of profit concavity in q_i. A2a implies that $p''(Q)q_i + 2p'(Q) - C''(q_i) \leq 0$, while A2b implies simply that $C''(q_i) \geq 0$. Neither condition implies the other.

Actions are strategic substitutes when $\frac{d^2\pi_i}{da_i dA_{-i}} < 0$. Then, $r_i(A_{-i})$ is a strictly decreasing function for $A_{-i} < \bar{A}_{-i}$, and is equal to zero otherwise. Conversely, actions are strategic complements when $\frac{d^2\pi_i}{da_i dA_{-i}} > 0$. Then, $r_i(A_{-i})$ is strictly increasing because marginal profits rise with rivals' strategic choices.

The next assumption is readily verified in the Cournot, CES and logit models.[8]

Assumption A3 (Reaction function slope) $\frac{d^2\pi_i}{da_i^2} < \frac{d^2\pi_i}{da_i dA_{-i}}$.

We next show A3 implies that there will be no over-reaction: if all other players collectively increase their actions, the reaction of i should not cause the aggregate to fall (see also McManus, 1962, p. 16, Selten, 1970, Corchón, 1994, and Vives, 1999, p. 42).

Lemma 1 *Under A3, $r_i'(A_{-i}) > -1$ and $A_{-i} + r_i(A_{-i})$ is strictly increasing in A_{-i}.*

Proof. From (1), $r_i'(A_{-i}) = \frac{-d^2\pi_i}{da_i dA_{-i}} / \frac{d^2\pi_i}{da_i^2}$. Because the denominator on the RHS is negative by the second-order condition (see A2a), A3 implies that $r_i'(A_{-i}) > -1$. Then $A_{-i} + r_i(A_{-i})$ strictly increases in A_{-i}. ■

Given the monotonicity established in Lemma 1, we can invert the relation $A = A_{-i} + r_i(A_{-i})$ to write $A_{-i} = f_i(A)$. We can therefore write pertinent relations as functions of A instead of A_{-i}. The construction of A from A_{-i} is illustrated in $d1$ for strategic substitutes. A hat over a variable denotes a specific value. $d1$ shows how knowing $\hat{a}_i = r_i\left(\hat{A}_{-i}\right)$ determines $\hat{A}$, which is the aggregate value consistent with firm i choosing $\hat{a}_i$. $A_{-i} = f_i(A)$ is then given by flipping the axes (inverting the relation).

[8] The Cournot model gives first derivative $p'(Q)q_i + p(Q) - C_i'(q_i)$. A3 implies $p''(Q)q_i + 2p'(Q) - C_i''(q_i) < p''(Q)q_i + p'(Q)$ or $p'(Q) < C_i''(q_i)$, which readily holds for $C_i''(q_i) \geq 0$.

2.2 Inclusive best reply (ibr) function

Selten (1970) first introduced the ibr as an alternative way to formulate the solution to the firm's problem. The ibr is the optimal action of firm i consistent with a given value of the aggregate, A.[9] It is natural to describe the maximization of $\pi_i(A, a_i)$ by writing the action choice as a function of the aggregate. Since Cournot (1838), however, economists have become accustomed to writing the action as a function of the sum of all others' actions. Our intuitions are based on that approach, so the alternative takes some getting used to. Nonetheless, we show that key properties such as strategic substitutability/complementarity are preserved under a mild assumption (A3), so the alternative construction is not too dissimilar. Its advantages are seen in the simple and clean characterizations it affords.

Let $\tilde{r}_i(A)$ stand for this ibr, i.e., the portion of A optimally produced by firm i (hence, $A - A_{-i} = r_i(A_{-i}) = \tilde{r}_i(A)$).[10] A differentiable $r_i(A_{-i})$ gives us a differentiable $\tilde{r}_i(A)$ function by construction.

Geometrically, $\tilde{r}_i(A)$ can be constructed as follows. For strategic substitutes, $a_i = r_i(A_{-i})$ decreases with A_{-i}, with slope above -1 (Lemma 1). At any point on the reaction function, draw down an isoquant (slope -1) to reach the A_{-i} axis, which it attains before the reaction function reaches the axis. The $x-$intercept is the A corresponding to A_{-i} augmented by i's contribution. This gives $a_i = \tilde{r}_i(A)$. Clearly, A and a_i are negatively related. This construction is shown in $d2$, where starting with $r_i\left(\hat{A}_{-i}\right)$ determines $\hat{A}$ and hence $\tilde{r}_i\left(\hat{A}\right)$.

[9]Selten (1970, p.154) calls it the *Einpassungsfunktion*, which Phlips (1995) translates as the "fitting-in function". An alternative translation is the ibr (see, e.g., Wolfstetter, 1999). Novshek (1985) refers to it as the "backwards reaction mapping" while Acemoglu and Jensen (2013) call it the "cumulative best reply" and Cornes and Hartley (2007a and 2007b) call it the "replacement function." McManus (1962 and 1964) graphs the aggregate as a function of the sum of the actions of all other players for the Cournot model, from which one can recover the ibr although he does not directly graph the ibr.

[10]Hence, in Figure 1, $\hat{a}_i = r_i\left(\hat{A}_{-i}\right) = \tilde{r}_i\left(\hat{A}\right)$.

Lemma 2 *If A3 holds, the ibr slope is* $\frac{d\tilde{r}_i}{dA} = \frac{r_i'}{1+r_i'} < 1$. *For strict strategic substitutes* $\tilde{r}_i(A)$ *is strictly decreasing for* $A < \bar{A}_{-i}$. *For strict strategic complements,* $\tilde{r}_i(A)$ *is strictly increasing.*

Proof. By definition, $\tilde{r}_i(A) = r_i(f_i(A))$. Differentiating yields $\frac{d\tilde{r}_i(A)}{dA} = \frac{dr_i(A_{-i})}{dA_{-i}} \frac{df_i(A)}{dA}$. Because $A_{-i} = f_i(A)$ from the relation $A = A_{-i} + r_i(A_{-i})$, applying the implicit function theorem gives us $\frac{\partial f_i}{\partial A} = \frac{1}{1+r_i'}$ and hence $\frac{d\tilde{r}_i}{dA} = \frac{r_i'}{1+r_i'}$. For strategic substitutes, because $-1 < r_i' < 0$ by Lemma 1, $\tilde{r}_i' < 0$. For strategic complements, $0 < \tilde{r}_i' < 1$. ■

Hence, strategic substitutability or complementarity is preserved in the ibr.[11] Note that $\tilde{r}_i' \to 0$ as $r_i' \to 0$ and $\tilde{r}_i' \to -\infty$ as $r_i' \to -1$.

The ibr was constructed by Selten (1970) to establish the existence of an equilibrium. An equilibrium exists if and only if $\sum_{i\in\mathcal{S}} \tilde{r}_i(A)$ has a fixed point. Because $\tilde{r}_i(A)$ is continuous, so too is the sum. Because the individual strategy spaces are compact intervals, then A must lie in a compact interval (its bounds are simply the sum of the individual bounds) and $\sum_{i\in\mathcal{S}} \tilde{r}_i(A)$ maps to the same compact interval. Therefore, there is a fixed point by the intermediate value theorem.

To guarantee uniqueness for a fixed number of firms, it suffices that at any fixed point

$$\sum_{i\in\mathcal{S}} \tilde{r}_i'(A) < 1. \tag{2}$$

We refer to this as the "sum-slope condition" and assume it holds. It automatically holds for strategic substitutes since Lemma 2 implies that $\sum_{i\in\mathcal{S}} \tilde{r}_i(A)$ is decreasing (see Vives, 1999, p. 43). For strategic complements, the condition may be violated, so papers on super-modular games (e.g., Milgrom and Shannon, 1994) often consider

[11] Importantly, although results in the short run critically depend on the slope of the reaction functions, as we will see, we do not need to distinguish between the cases of strategic complements and substitutes for our long-run analysis. Moreover, none of our results rides on the assumption that the ibr is monotone. The only key property for the long-run analysis is that the ibr has slope less than 1, as shown in Lemma 2. Otherwise, the ibr can be non-monotone.

extremal equilibria, at which it holds. We only invoke (2) on those rare occasions when we describe short-run equilibria.

We next present three results which will play a critical role in the development of our core results in Section 4, and their applications in Sections 6, 7, and 8. Let

$$\pi_i^*(A) \equiv \pi_i(A, \tilde{r}_i(A)). \tag{3}$$

It is the value of i's profit when firm i maximizes its profit given the actions of the others and doing so results in A as the total.

Lemma 3 *Under A1-A3, $\pi_i^*(A)$ is strictly decreasing for $A < \bar{A}_{-i}$ and is zero otherwise.*

Proof. For $A \geq \bar{A}_{-i}$, we have $\tilde{r}_i(A) = 0$ by definition, and $\pi_i^*(A) = 0$ for $A \geq \bar{A}_{-i}$. For $A < \bar{A}_{-i}$, from (3), $\frac{d\pi_i^*(A)}{dA} = \frac{d\pi_i(A,\tilde{r}_i(A))}{dA} = \pi_{i,1} + \pi_{i,2}\frac{\partial \tilde{r}_i(A)}{\partial A} = \pi_{i,1}\left(1 - \frac{\partial \tilde{r}_i(A)}{\partial A}\right)$, where the last equality follows from (1). This is negative by A1 and Lemma 2. ■

Lemma 3 helps us establish uniqueness in the long run given the equilibrium concept we introduce in Section 3.

The next result establishes the conditions under which the ibr shifts up. For this, we introduce a shift variable θ_i explicitly into the profit function, $\pi_i(A, a_i; \theta_i)$. We say a difference that raises $\tilde{r}_i(A)$ renders firm i *more aggressive.*

Lemma 4 *(Aggression) $\frac{d\tilde{r}_i(A;\theta_i)}{d\theta_i} > 0$ if and only if $\frac{d^2\pi_i(A,a_i;\theta_i)}{d\theta_i da_i} > 0$.*

Proof. Applying the implicit function theorem to the reaction function shows that $\partial r_i/\partial\theta_i > 0$ if and only if $\frac{\partial^2\pi_i(A,a_i;\theta_i)}{\partial\theta_i\partial a_i} > 0$. Now, by definition, $\tilde{r}_i(A;\theta_i) = r_i(f_i(A,\theta_i);\theta_i)$, where we recall that $f_i(.)$ denotes the A_{-i} locally defined by the relation $A - A_{-i} - r_i(A_{-i};\theta_i) = 0$. Hence, $\frac{d\tilde{r}_i(A;\theta_i)}{d\theta_i} = \frac{\partial r_i(A_{-i};\theta_i)}{\partial A_{-i}}\frac{df_i(A)}{d\theta_i} + \frac{\partial r_i(A_{-i};\theta_i)}{\partial\theta_i}$. Using the implicit function theorem again, we get $\frac{df_i(A)}{d\theta_i} = \frac{-\partial r_i/\partial\theta_i}{1+\partial r_i/\partial A_{-i}}$. Hence,

$$\frac{d\tilde{r}_i(A;\theta_i)}{d\theta_i} = \frac{\partial r_i/\partial\theta_i}{1 + \partial r_i/\partial A_{-i}}, \tag{4}$$

which is positive since the denominator is positive by Lemma 1. ■

The final result will be useful in the analysis of monopolistic competition (Section 7) and leadership (Section D in the Online Appendix). Let $\widehat{r}_i(A)$ stand for the value of a_i that maximizes $\pi_i(A, a_i)$ for any given A.

Lemma 5 *Under A1 and A2b, $\widehat{r}_i(A) > \tilde{r}_i(A)$. Furthermore, $\pi_i(A, \widehat{r}_i(A))$ is the greatest possible profit i can earn for a given A.*

Proof. $\tilde{r}_i(A)$ is defined by $\pi_{i,1}(A, \tilde{r}_i(A)) + \pi_{i,2}(A, \tilde{r}_i(A)) = 0$. The first term is always negative (implied by A1), so the second term must be positive at $a_i = \tilde{r}_i(A)$. Then, for a given A, $\pi_i(A, a_i)$ is increasing in a_i at $a_i = \tilde{r}_i(A)$, and attains its highest value at $a_i = \widehat{r}_i(A)$. Hence, by A2b, the value of a_i that maximizes $\pi_i(A, a_i)$ for given A is larger than $\tilde{r}_i(A)$. ■

3 Free Entry Equilibrium (FEE)

Given the cost of entry K_i for firm i, a Free Entry Equilibrium (FEE) is defined as:

Definition 1 $\left\{(a_i^*)_{i \in \mathcal{S}}\right\}$ *is a FEE with a set $\mathcal{S}$ of active firms if:*

$$\pi_i \left(\sum_{j \in \mathcal{S}} a_j^*, a_i^* \right) \geq K_i \quad \textit{for all } i \in \mathcal{S}$$

and

$$a_i > 0 \Rightarrow \pi_i \left(a_i + \sum_{j \in \mathcal{S}} a_j^*, a_i \right) < K_i \quad \textit{for all } i \notin \mathcal{S}.$$

The first condition means that the firms which are in the market earn more than their entry costs, and therefore do not regret their entry decisions. The second condition means that any firm that is not in the market has no incentive to enter the market. Generally, an equilibrium set of firms will not be unique.

Let $A_{\mathcal{S}}$ be the equilibrium value of the aggregate corresponding to the set $\mathcal{S}$ of active firms. Using this notation, the two conditions in Definition 1 can be restated as $\pi_i^*(A_{\mathcal{S}}) \geq K_i$ for all $i \in \mathcal{S}$ and $\pi_i^*\left(A_{\{\mathcal{S}+i\}}\right) < K_i$ for all $i \notin \mathcal{S}$.

It is common to assume in the literature that in a free entry equilibrium, the marginal firm earns exactly zero profits. To that end, we assume that there is a set $\mathcal{E}$ of firms that we describe as marginal entrants, each of which has the same profit function, $\pi_{\mathcal{E}}(A, a_i)$, and the same entry cost, $K_{\mathcal{E}}$. The set $\mathcal{E}_{\mathcal{A}} = \mathcal{E} \cap \mathcal{S}$ denotes the set of active marginal entrants (those which have sunk the entry cost). Using this notation, we can define the equilibria on which we focus in this paper as follows:

Definition 2 *A Zero Profit Symmetric Entrants Equilibrium (ZPSEE) is a FEE with a set $\mathcal{S}$ of active firms such that $\mathcal{E}_{\mathcal{A}} = \mathcal{E} \cap \mathcal{S} \neq \emptyset$ and $\pi_{\mathcal{E}}\left(\sum_{j \in \mathcal{S}} a_j^*, a_i^*\right) = K_{\mathcal{E}}$ for all $i \in \mathcal{E}_{\mathcal{A}}$.*

Although the ZPSEE is used widely, it does not account for integer constraints. We account for integers using a bounds approach in the Appendix.

Our goal is to present comparative static analyses of how the ZPSEE differs across market structures. We interpret market structure broadly to encompass market institutions (e.g., privatization or nationalization), technological conditions (e.g., cost shocks), etc. We consider market structure differences that directly impact active firms other than the marginal entrants. We refer to the non-marginal firms as insiders, $\mathcal{I}$. We assume that they are in $\mathcal{S}$ in the base market structure and the comparison one. The structural difference can affect some or all of the insiders. We refer to those that are affected as changed insiders, $\mathcal{I}_{\mathcal{C}}$, and those that are not affected as unchanged insiders, $\mathcal{I}_{\mathcal{U}}$.[12]

[12] It is straightforward (though cumbersome) to allow some of the changed firms in $\mathcal{I}_{\mathcal{C}}$ to be inactive under one equilibrium market structure, in which case they earn zero rents.

To illustrate, consider the long-run impact of a cost difference (due to a selective tax or subsidy perhaps). In the base ZPSEE, the set of active firms might comprise five insiders and eight marginal entrants. The cost difference might mean that two of the insiders have lower marginal costs. These two would be the changed insiders, while the other three insiders would be unchanged insiders. We would then compare the ZPSEE in which the two changed insiders have lower costs with the ZPSEE in which they do not. For instance, compare the number of marginal entrants active in the market in each ZPSEE, the price and output levels that they and the insiders choose, the profitability of both the changed and the unchanged insiders, and the welfare implications of the cost difference.

It is important to note that our ZPSEE analysis makes no assumptions regarding the characteristics of the insiders. What is critical is the symmetry of the marginal entrants. In the sections that follow, we present our core comparative static results for ZPSEE, followed by several applications of those results. We then consider FEE other than ZPSEE by modelling heterogeneous marginal entrants in Section 11.

4 Core propositions

We now present our core results. We wish to compare the positive and normative equilibrium characteristics of two different market structures (with the firms $i \in \mathcal{I}_{\mathcal{C}}$ being altered). Let $\mathcal{S}'$ and $\mathcal{S}''$, both of which contain $\mathcal{I}$, stand for the ZPSEE set of firms in the two market structures. Let $A' = A_{\mathcal{S}'}$ and $A'' = A_{\mathcal{S}''}$ be the equilibrium values of the aggregate at the two different equilibrium sets of active firms, and likewise let a_i' and a_i'' be the actions of individual active firms.

4.1 Aggregate and individual actions

Proposition 1 *(Aggregate and individual actions) Suppose that some change to the firms* $i \in \mathcal{I}_{\mathcal{C}}$ *shifts the ZPSEE set of firms from* $\mathcal{S}'$ *to* $\mathcal{S}''$*, both of which contain* $\mathcal{I} = \mathcal{I}_{\mathcal{C}} \cup \mathcal{I}_{\mathcal{U}}$ *and at least one firm from* $\mathcal{E}$*. Then, under A1-A3,* $A' = A''$*,* $a_i' = a_i''$ *for all* $i \in \mathcal{E}_{\mathcal{A}}$*, and* $a_i' = a_i''$ *for all unaffected firms* $i \in \mathcal{I}_{\mathcal{U}}$*.*

Proof. By Lemma 3, $\pi^*_{\mathcal{E}}(A)$ is strictly decreasing in A for $A < \bar{A}_{-i}$, and $\pi^*_{\mathcal{E}}\left(\bar{A}_{-i}\right) = 0$, which implies that there is a unique solution, $A < \bar{A}_{-i}$, for the aggregate at any ZPSEE. In order for there to be at least one active marginal entrant but not all, it must be true that $\pi^*_{\mathcal{E}}(A_{\mathcal{I}}) > K_{\mathcal{E}} > \pi^*_{\mathcal{E}}(A_{\mathcal{I}\cup\mathcal{E}})$, where $A_{\mathcal{I}}$ is the aggregate value with all firms in $\mathcal{I}$ active and $A_{\mathcal{I}\cup\mathcal{E}}$ is the value with all firms in $\mathcal{I}$ and $\mathcal{E}$ active. Hence, we must have $A' = A'' = \pi^{*-1}_{\mathcal{E}}(K_{\mathcal{E}})$.

Since $A' = A''$ and the ibr $\tilde{r}_i(A)$ is the same for all $i \in \mathcal{E}_{\mathcal{A}}$, we have $\tilde{r}_i(A'') = \tilde{r}_i(A')$. Similarly, for each unaffected firm $i \in \mathcal{I}_{\mathcal{U}}$ (that is, insider firms whose profit functions remain unchanged), we have $\tilde{r}_i(A'') = \tilde{r}_i(A')$. ∎

Proposition 1, while simple, is a powerful result that provides a strong benchmark. The composition of A' and A'' may be quite different due to the differences between the infra-marginal firms. There can be more or fewer firms present in the market. The result applies irrespective of whether firms' actions are strategic substitutes or complements. In contrast, in short-run models (without entry), strategic substitutability or complementarity determines equilibrium predictions (which can differ dramatically). Finally, the result applies irrespective of how much heterogeneity there is among the insiders. The aggregative approach significantly reduces the complexity of the problem.

Although the aggregate and the equilibrium action of each active firm from $\mathcal{E}$ stays the same, there may be more or less active marginal entrants in the market as a result of the change. We say that a difference in market structure renders the changed insider firms more (less) aggressive in sum if it raises (decreases) $\sum_{i\in\mathcal{I}_\mathcal{C}} \tilde{r}_i(A)$. Then, an implication of Proposition 1 is that any change making the affected insiders more (less) aggressive in sum will decrease (increase) the number of firms in $\mathcal{E}_\mathcal{A}$. This is because if $A' = A''$ and the affected insiders become more (less) aggressive in sum, then there must be fewer (more) firms from $\mathcal{E}$ because $a_i' = a_i''$ for all $i \notin \mathcal{I}_\mathcal{C}$.[13]

4.2 Total welfare

We next consider how welfare differs across equilibria.

Proposition 2 *(Welfare) Suppose that some change to the firms $i \in \mathcal{I}_\mathcal{C}$ shifts the ZPSEE set of firms from $\mathcal{S}'$ to $\mathcal{S}''$, both of which contain $\mathcal{I} = \mathcal{I}_\mathcal{C} \cup \mathcal{I}_\mathcal{U}$ and at least one firm from $\mathcal{E}$. Suppose also that consumer surplus depends solely on A. Then, under A1-A3:*

(i) Consumer surplus remains unchanged.

(ii) Rents of firms $i \notin \mathcal{I}_\mathcal{C}$ remain unchanged at a ZPSEE, so the change in producer surplus equals the change in rents to the changed insiders, $i \in \mathcal{I}_\mathcal{C}$.

(iii) The change in total surplus is measured solely by the change in the rents of the changed insiders, $i \in \mathcal{I}_\mathcal{C}$.

Proof. (i) By Proposition 1, $A' = A'' = \pi_\mathcal{E}^{*-1}(K)$ at any ZPSEE. The result follows.

(ii) This follows directly from Proposition 1. The aggregate remains the same, the best replies remain the same, and, since the profit functions of the unaffected firms are the same, their rents remain the same. Hence, the total change to producer surplus is measured as the change in the affected firms' rents.

(iii) This is immediate from (i) and (ii). ■

[13] Insiders do not all have to be affected in the same way by a change in the market structure. Some could become more aggressive and others less so, for example. What matters is what happens to the sum of their ibrs. We thank a referee for pointing this out.

As an example, consider an industry where some public firms are privatized. The results above imply that in the long run, consumers neither benefit nor suffer. Total welfare changes by the change in the profits of the privatized firms.[14]

In the oligopoly context, consumers are affected by differences in market structures. Their welfare is an important or even decisive criterion (under a consumer welfare standard) for evaluating the desirability of different market structures.[15] An increase in the aggregate is a sufficient statistic for consumer welfare to rise whenever consumer welfare depends just on the aggregate. For example, in case of Cournot competition with homogeneous goods, the aggregate A is total output, Q, and consumer welfare depends directly on the aggregate via the market price, $p(Q)$. There are a number of other important cases where consumer surplus depends solely on the value of the aggregate (and not its composition). These are discussed in Section 5 below.

Although Proposition 2 follows immediately from Proposition 1, it is not at all obvious a priori that a change in market structure would have no impact on long-run consumer surplus. The result does not hold if the composition of A matters to consumers. This may be so when there is an externality, like pollution, which varies across firms. Then a shift in output composition towards less polluting firms raises consumer welfare.

[14] This generalizes Anderson et al. (1997) who consider the case of a single public firm.

[15] Following standard practice, consumer surplus does not include the transfer of profits back to the consumer. Of course, consumers are better off once they receive profit revenue (which they spend on the numeraire when preferences are quasi-linear). Our discussion follows the standard division of rents to a consumer side and a producer side. We return to this issue in Section 10, where we consider income effects, and we specifically evaluate the benefits to consumers from receiving profits.

5 Consumer welfare and Bertrand differentiated product games

The normative properties of Proposition 2 hold if consumer surplus depends solely on the aggregate. In this section, we show that this is the case in Bertrand (pricing) games with differentiated products where demand satisfies the IIA property.

Suppose the profit function takes the form $\pi_i = (p_i - c_i) D_i(\vec{p})$ where $\vec{p}$ is the vector of prices set by firms and $D_i(\vec{p})$ is firm i's demand function. We are interested in the conditions under which $D_i(\vec{p})$ implies an aggregative game for which the consumer welfare depends only on the aggregate.

So consider a quasi-linear consumer welfare (indirect utility) function $V(\vec{p}, Y) = \phi(\vec{p}) + Y$, where Y is income. Suppose first that we can write $\phi(\vec{p})$ as an increasing function of the sum of decreasing functions of p_i, so $\phi(\vec{p}) = \tilde{\phi}\left(\sum_i g_i(p_i)\right)$ where $\tilde{\phi}' > 0$ and $g_i'(p_i) < 0$. Then, by Roy's Identity, $D_i(\vec{p}) = -\tilde{\phi}'\left(\sum_i g_i(p_i)\right) g_i'(p_i) > 0$, which therefore depends only on the summation and the derivative of $g_i(.)$. Assume further that $D_i(\vec{p})$ is decreasing in own price $\left(\frac{dD_i(\vec{p})}{dp_i} = -\tilde{\phi}''(.)\left[g_i'(p_i)\right]^2 - \tilde{\phi}'(.)\, g_i''(p_i) < 0\right)$.[16] Since $g_i(p_i)$ is decreasing, its value uniquely determines p_i and hence the term $g_i'(p_i)$ in the demand expression. Therefore, demand can be written as a function solely of the summation and $g_i(p_i)$. This means that the game is aggregative, by choosing $a_i = g_i(p_i)$ and $A = \sum_i a_i$.[17] Hence, consumer welfare ($V = \phi(A) + Y$) depends only on A (and not its composition). This structure has another important property, namely that the demand functions satisfy the IIA property: the ratio of any two demands depends only on their own prices (and is independent of the prices of other options in the choice set). That is, $\frac{D_i(\vec{p})}{D_j(\vec{p})} = \frac{g_i'(p_i)}{g_j'(p_j)}$. In summary:

[16] For the logsum formula which generates the logit model, we have $g_i(p_i) = \exp\left[(s_i - p_i)/\mu\right]$ and so $g_i''(p_i) > 0$. However, $\tilde{\phi}$ is concave in its argument, the sum.

[17] Hence, $\pi_i = \left(g_i^{-1}(a_i) - c_i\right)\left(-\phi'(A)\, g_i'\left(g_i^{-1}(a_i)\right)\right)$ as per the earlier logit example.

Proposition 3 *Let* $\pi_i = (p_i - c_i) D_i(\vec{p})$ *and* $D_i(\vec{p})$ *be generated by an indirect utility function* $V(\vec{p}, Y) = \tilde{\phi}\left(\sum_i g_i(p_i)\right) + Y$ *where* $\tilde{\phi}$ *is increasing and twice differentiable, strictly convex in* p_i*, and* $g_i(p_i)$ *is twice differentiable and decreasing. Then demands exhibit the IIA property, the Bertrand pricing game is aggregative, and consumer welfare depends only on the aggregate,* $A = \sum_i a_i$*, where* $a_i = g_i(p_i)$*.*

Important examples include the CES and logit demand models. For the CES model, we have $V = \frac{1}{\lambda} \ln A + Y - 1$, where the action variables are $a_i = p_i^{-\lambda}$ and $Y > 1$ is income. For the logit model, we have the "log-sum" formula $V = \mu \ln A + Y$, and the action variables are $a_i = \exp[(s_i - p_i)/\mu]$.[18]

We also prove a converse to Proposition 3. Suppose that demands exhibit the IIA property, and assume quasi-linearity. Following Theorem 1 in Goldman and Uzawa (1964, p. 389), V must have the form $\tilde{\phi}\left(\sum_i g_i(p_i)\right) + Y$ where $\tilde{\phi}(.)$ is increasing and $g_i(p_i)$ is any function of p_i. If we further stipulate that demands must be differentiable, then the differentiability assumptions made in Proposition 3 must hold. Then, assuming that demands are strictly downward sloping implies that $\tilde{\phi}\left(\sum_i g_i(p_i)\right)$ must be strictly convex in p_i. In summary:

Proposition 4 *Let* $\pi_i = (p_i - c_i) D_i(\vec{p})$ *and* $D_i(\vec{p})$ *be twice continuously differentiable and strictly decreasing in own price. Suppose that the demand functions satisfy the IIA property. Then the demands* $D_i(\vec{p})$ *can be generated by an indirect utility function* $V(\vec{p}, Y) = \tilde{\phi}\left(\sum_i g_i(p_i)\right) + Y$ *where* $\tilde{\phi}$ *is twice differentiable, strictly convex in* p_i*, and* $g_i(p_i)$ *is twice differentiable and decreasing. Then the Bertrand game is aggregative, and consumer welfare depends only on the aggregate,* $A = \sum_i a_i$*, where* $a_i = g_i(p_i)$*.*

[18] See Anderson et al. (1992) for a discussion of the two demand systems. They show that both demand systems can be derived as representative consumer, random utility, and spatial models. The Lucian demand system developed in Anderson and de Palma (2012) provides another example.

However, the fact that a game is aggregative does not imply that the IIA property holds. For example, the linear differentiated products demand system of Ottaviano and Thisse (2004) gives rise to an aggregative game for a fixed number of firms (i.e., in the short-run) with Bertrand competition since demand can be written as a sum of all prices and own price. However, it does not satisfy the IIA property, so the welfare implications do not follow for this specification. The composition of A matters for consumer welfare.

6 Mergers and cooperation

In this section and the next two, we consider applications and show how our toolkit can be used in the context of some commonly considered questions in the literature to gain more insightful and general results.[19] We start with considering the short-run and long-run impact of mergers.

Suppose that two firms cooperate by maximizing the sum of their payoffs (the results easily extend to larger pacts). The merger can be a rationalization of production across plants, or a multi-product firm pricing different variants. Merger synergies can result in both marginal cost and fixed cost savings. We assume that there are no marginal cost savings - these can be incorporated in the analysis by using the effects of cost changes described in the Online Appendix (Section C). We derive existing results in the literature concisely from our framework, and we deliver new results on the long-run impact of mergers in differentiated product markets.

Merged firms jointly solve $\max_{a_j,a_k} \pi_j(A, a_j) + \pi_k(A, a_k)$. The first order conditions take the form

$$\pi_{j,1}(A, a_j) + \pi_{j,2}(A, a_j) + \pi_{k,1}(A, a_k) = 0, \tag{5}$$

which differs from (1) by the last term, which internalizes the aggregate effect on

[19] More applications (on cost changes, contests, leadership, R&D, privatization, etc.) are available in the Online Appendix.

sibling payoff. The two first order conditions can be solved simultaneously to find a_j and a_k as functions of the aggregate, giving $\tilde{r}_j^m(A)$ and $\tilde{r}_k^m(A)$ as the individual ibr functions under merger. Summing these gives the pact's ibr, $\tilde{R}^m(A)$.

Lemma 6 *Consider a merger between firms j and k. Then, for any A, $\tilde{r}_j^m(A) \leq \tilde{r}_j(A)$, $\tilde{r}_k^m(A) \leq \tilde{r}_k(A)$, and $\tilde{R}^m(A) < \tilde{r}_j(A) + \tilde{r}_k(A)$.*

Proof. First suppose both j and k are active under the merger. By A1, $\pi_k(A, a_k)$ is decreasing in A, so the third term in (5) is negative. Thus, for *any* $a_k > 0$, the choice of a_j must be lower at any given A, so $\tilde{r}_j^m(A) < \tilde{r}_j(A)$, and likewise for a_k. Second, if only firm k is active under the merger (e.g., only the lower-cost firm operates when Cournot firms produce homogeneous goods at constant but different marginal costs), then $0 = \tilde{r}_j^m(A) < \tilde{r}_j(A)$ and $\tilde{r}_k^m(A) = \tilde{r}_k(A)$. In both cases, $\tilde{R}^m(A) < \tilde{r}_j(A) + \tilde{r}_k(A)$. ■

For given A, merged firms choose lower actions (lower quantity in Cournot, higher price in Bertrand).[20] That the combined entity has lower total production was stressed by Salant et al. (1983) for Cournot competition with linear demand. Lemma 6 gives this result for any aggregative game using the new concept of the pact ibr.

Consider first mergers in the short run. The equilibrium aggregate, for a given set $\mathcal{S}$ of firms, solves $\sum_{i\in\mathcal{S}} \tilde{r}_i(A) = A$. A merger only affects the ibr functions of the firms involved. Hence, by Lemma 6, $\sum\limits_{i\in\mathcal{S}} \tilde{r}_i(A) > \sum\limits_{i\neq j,k} \tilde{r}_i(A) + \tilde{R}^m(A)$. Since the sum will intersect the 45° line at a lower A, the aggregate falls for strategic substitutes and other firms' actions rise (because $\tilde{r}_i'(A) < 0$ by Lemma 2). In the Cournot model, other firms expand output, so the merged firm's total output must contract by more to render the lower total A. Under the sum-slope condition (2), A also falls

[20] To illustrate, consider a merger in a Cournot market with linear demand. The cost function of firm j is $C_j(q_j) = q_j^2$ and of firm k is $C_k(q_k) = q_k^2/2$. The merged firm maximizes $(1-Q)q_j - q_j^2 + (1-Q)q_k - q_k^2/2$. Solving the FOCs for q_j and q_k yields $\widetilde{r}_j^m(Q) = \frac{1-Q}{5} < \widetilde{r}_j(Q) = \frac{1-Q}{3}$ and $\widetilde{r}_k^m(Q) = \frac{2(1-Q)}{5} < \widetilde{r}_k(Q) = \frac{1-Q}{2}$.

for strategic complements, and others' actions fall (which implies higher prices under Bertrand competition). The merged firm's actions fall for the twin reasons of the direct lowering of the reaction functions and their positive slope.

The next result follows because $\pi_i^*(A)$ is decreasing by Lemma 3.

Proposition 5 *Suppose two firms merge. The aggregate decreases in the short run. Hence, the non-merged firms' profits go up, and consumer welfare goes down when it decreases with A.*

For strategic substitutes, the "Cournot merger paradox" result of Salant et al. (1983) shows that mergers are not profitable unless they include a sufficiently large percentage of the firms in the market. Other firms benefit while merging firms can lose. For strategic complements, the other firms' response reinforces the merged firms' actions and mergers are always profitable (Deneckere and Davidson, 1985). However, non-merged firms still benefit "more" from a merger. This is because each merged firm cannot choose the action that maximizes its individual profits while each non-merged firm does.

In the long run, entry undoes the short-run impact of the merger:

Proposition 6 *Suppose two firms merge and a ZPSEE prevails. Then:*

(i) The aggregate, non-merging firms' actions and profits, and consumer welfare (when it depends solely on A) remain the same.

(ii) There are more entrants, and profits to merging firms are all weakly lower.

Proof. (i) By Propositions 1 and 2.

(ii) By Lemmas 5 and 6, $\tilde{r}_j^m(A) \leq \tilde{r}_j(A) < \widehat{r}_j(A)$. Since $\pi_j(A, a_j)$ is quasi-concave in a_j (A2b), $\pi_j(A, \tilde{r}_j(A)) = \pi_j^*(A) \geq \pi_j\left(A, \tilde{r}_j^m(A)\right)$. There are more entrants in equilibrium because A does not change and merging firms' actions decrease. ■

Proposition 6 applies with asymmetric insiders as long as the marginal entrant's type does not change. If the firms are symmetric and making zero profits to start with, then, with a merger and subsequent entry, the pact firms make negative profits. Hence, cost savings are required in order to give firms a long-run incentive to merge. In this sense, the Cournot merger paradox is now even stronger: absent synergies, pact firms are *always* worse off. Likewise, the profitability of mergers under Bertrand competition no longer holds in the long run.

Proposition 6(i) implies that entry counteracts the short-run negative impact of mergers on consumer welfare. In the long-run, more firms enter and consumers benefit from extra variety. In ZPSEE, the merging firms have higher prices (while all non-merging firms are where they started in terms of price and profit), but the effect of higher prices is *exactly* offset by more variety in consumer welfare.

Davidson and Mukherjee (2007) analyze the long-run impact of a merger in the special case of homogeneous goods Cournot competition with linear demand. Using the aggregative game structure, we are able to make a much broader statement covering multi-product firms and differentiated goods markets with Bertrand competition under IIA (CES and logit). Our positive results also cover Cournot competition with linear differentiated products (Ottaviano and Thisse, 2004), but the normative results do not apply because consumer welfare does not solely depend on the aggregate.

The policy implications of Proposition 6 are very strong. *Under free entry, mergers are socially desirable from a total welfare standpoint if and only if they are profitable.* Laissez-faire is the right policy, and there is no role for antitrust authorities. This conclusion holds even under a consumer-welfare standard for mergers (since consumers remain indifferent by Proposition 2), and even if the merger involves synergies (by Proposition 2). Put another way, our core propositions show that IIA demand systems build in that result.[21] As we discuss later and in the Online Appendix, the result is

[21] Erkal and Piccinin (2010) analyze the long-run impact of mergers under Cournot competition

tempered by income effects, heterogeneous marginal entrants, and integer issues.

7 Monopolistic competition (MC)

Many papers assume firms are monopolistically competitive (MC). This assumption is cleanly interpreted in the aggregative game setting: firms do not internalize the effects of their actions on the aggregate (e.g., in CES models, the "price index" is taken as given). In this sense, their behavior is like the Stackelberg leader's action considered in the Online Appendix (Section D). Hence, for any given value of the aggregate, actions are larger (lower prices/higher quantities) than oligopolistic ones (see Lemma 5). Since the marginal entrants' behavior changes, the equilibrium value of the aggregate also changes.

Under MC, the marginal entrant's zero-profit condition is $\pi_{\mathcal{E}}\left(A, \widehat{r}_{\mathcal{E}}\left(A\right)\right) = K$, while in the oligopoly model thus far, it is $\pi_{\mathcal{E}}\left(A, \tilde{r}_{\mathcal{E}}\left(A\right)\right) = \pi_{\mathcal{E}}^{*}\left(A\right) = K$. The comparison is straightforward. Because the MC marginal entrants maximize profit for any given A, $\pi_{\mathcal{E}}\left(A, \widehat{r}_{\mathcal{E}}\left(A\right)\right) > \pi_{\mathcal{E}}\left(A, \tilde{r}_{\mathcal{E}}\left(A\right)\right)$. In fact, as shown in Lemma 5, MC yields maximal profit for given A. Together with Lemma 3, this implies that the aggregate is the largest one possible. Any other market structure gives a lower value.

Proposition 7 *The aggregate attains its maximum possible value under monopolistically competitive behavior of marginal entrants. If consumer surplus depends only on A, then consumer surplus is maximal under the zero profit condition.*

This result explains (through a new lens) why it is that MC delivers the second best optimum allocation under the zero profit condition (see, e.g., Spence, 1976). By delivering the greatest profit per firm for a given aggregate, it generates the largest possible aggregate value.

with linear differentiated product demand. The game is aggregative both in the short run and the long run in this case, and the merger has no impact on the aggregate, but since the demand system does not satisfy IIA, the consumer welfare conclusions are different.

For example, for CES/Logit, all firms have *lower* prices under MC compared to oligopoly with free entry. The higher aggregate gives higher actions (lower prices) by strategic complementarity, and higher actions still by the Stackelberg-like property (see Lemma 5). This leads to higher consumer surplus and total welfare under MC.

Finally, it is insightful to apply a MC assumption to the Cournot context. Under symmetry, firms solve $\max_q \pi(Q, q) = p(Q)q - C(q)$ taking Q as given. The solution is perfect competition with free entry. As we know, this is the optimal outcome, as Proposition 7 attests. The aggregative game lens brings out this common structure.

8 Gains from trade

We now apply the analysis of aggregate games with the ZPSEE to trade liberalization. A standard approach in trade is to use models of monopolistic competition with CES preferences (see, e.g., Melitz, 2003). Because the CES yields an aggregative game for Bertrand oligopoly, we can apply our framework to determine how oligopoly outcomes compare to monopolistically competitive ones.

Our setting is to take an autarchic economy and replicate it k-fold. We investigate the impact of such market expansion on mark-ups, product variety, and firm selection (i.e., whether more production is done by the more productive firms). Under monopolistic competition with CES preferences, expanding the economy has no effect on equilibrium mark-ups and firm selection, and simply increases k-fold the diversity of products, thus increasing consumer welfare by increasing choice. This raises the question of whether monopolistic competition overestimates or underestimates the gain in consumer welfare if the true situation were properly characterized by oligopoly. We show that with Bertrand oligopoly, mark-ups decrease and consequently product variety increases less than k-fold. Moreover, there is selection of more productive firms into the market. When we compare to a situation where mark-ups and firm selection are constant, and variety rises k-fold, there are two conflicting impacts on consumer welfare; a detrimental variety effect and a beneficial mark-up plus selection effect. A priori, this could cause consumer surplus to rise more or less under oligopoly than under monopolistic competition.

Note that in a replication environment, monopolistic competition both starts and finishes at a higher aggregate value (from Proposition 7), so it gives higher consumer welfare than a ZPSEE. However, we show that the aggregate, A, rises k-fold with monopolistic competition and proportionately more with oligopoly. Hence, if consumer welfare is concave in A, it is unclear a priori whether the increase in welfare is higher under monopolistic competition or oligopoly. We show that in the central cases of logit and CES, the gain is higher under oligopoly, and in that sense the gains from trade are underestimated under monopolistic competition (even though monopolistic competition yields higher welfare per se). The interpretation is that the benefits from tougher competition and better firm selection more than outweigh the loss of variety (relative to monopolistic competition).

Although pro-competitive effects have been central to discussions of the gains from trade, it has been challenging to capture them in a tractable trade model. This is largely because of the shortcomings of the standard approach in trade, the CES monopolistic competition model, which does not allow for pro-competitive effects. In response, two alternative paths have been pursued. The first approach maintains the assumption of monopolistic competition but assumes other, non-CES, type of preferences (see, e.g., Krugman, 1979; Melitz and Ottaviano, 2008; Arkolakis et al., 2012; Behrens and Murata, 2012). The second approach considers models of oligopoly, with or without CES preferences (see, e.g., Devereux and Lee, 2001; Bernard et al., 2003; Atkeson and Burstein, 2008; de Blas and Russ, 2010; Epifani and Gancia, 2011; Holmes et al., 2014). The results on the pro-competitive effects of trade are mixed and depend on considerations such as the asymmetries between countries, intersectoral differences, and free entry.

In what follows, we demonstrate how the toolkit of aggregative oligopoly games with the ZPSEE can be used to address the pro-competitive effects of trade in a very tractable way, taking into account the channels of endogenous mark-ups, product variety, and firm selection. None of these channels are new in the literature. Our contribution is to show how the toolkit of aggregative oligopoly games with ZPSEE can be used (i) to analyze all these channels in a unified framework, and (ii) to compare the welfare gains under monopolistic competition and oligopoly. Our analysis has CES preferences as a special case, but it is more general than that because it encompasses any type of preferences that yield an aggregative game structure.

As in Melitz (2003), we consider a framework where firms are differentiated based on their productivity, represented by θ_i. Let $G(\theta)$ stand for the distribution function of the productivity levels and $g(\theta)$ stand for the corresponding density function. We assume that firms know their productivity levels and make production decisions based on that knowledge. Production involves a fixed cost, K. As in our main framework, we use ZPSEE as our equilibrium concept.

We interpret trade as an increase in market size and suppress trade costs. When the market size scales up, the mass of firms (and their types) scales up by the same proportion. However, the type of the marginal entrant does not change and there are always some marginal entrants that are active in equilibrium. We let N represent the mass of potential firms and x represent the endogenous fraction of marginal entrants that are active in equilibrium. We assume that all the marginal entrants have the lowest possible productivity level, $\underline{\theta}$.

We first consider how x and A change as the size of the economy changes, before turning our attention to the consumer welfare gains. An change in x corresponds to both a selection (of the more efficient firms) and a variety effect.

Letting Z denote market size, we assume that the profit function has the form

$$\pi_i = \frac{Zh(a_i, \theta_i)}{A}. \tag{6}$$

This functional form covers CES and logit demand functions. For example, $h(a_i, \theta_i) = \left(a_i^{-1/\lambda} - c(\theta_i)\right) a_i^{(\lambda+1)/\lambda}$ in the case of CES and $h(a_i, \theta_i) = (s_i - \mu \ln a_i - c(\theta_i))\, a_i$ in the case of logit, where $c(\theta_i)$ is a decreasing function denoting higher marginal production costs for lower productivity levels.

The inclusive best response function, $\widetilde{r}_i(A, \theta_i)$, is the implicit solution a_i to the first order condition:

$$Ah_a(a_i, \theta_i) - h(a_i, \theta_i) = 0. \tag{7}$$

We assume $h_{aa}(a_i, \theta_i) < 0$, so the second order condition holds. This implies strategic complementarity (i.e., the slope of the ibr is positive).

The ZPSEE is described by the following two equations. The first one is the zero-profit condition for the marginal entrants:

$$\pi_i^*(A, \underline{\theta}) = \frac{Zh^*(A, \underline{\theta})}{A} = K, \tag{8}$$

where $h^*(A, \underline{\theta}) = h(\widetilde{r}_i(A, \underline{\theta}), \underline{\theta})$. It is readily shown that $\pi_i^*(A, \theta_i)$, the maximized profit function, is decreasing in A (as per Lemma 3), so the equilibrium value of A is tied down by (8).

The second equation describes the composition of A:

$$N\left[xG(\underline{\theta})\,\widetilde{r}_i(A, \underline{\theta}) + \int_{\theta > \underline{\theta}} \widetilde{r}_i(A, \theta)\, g(\theta)\, d\theta\right] = A, \tag{9}$$

where x represents the (endogenous) fraction of active firms which have the worst productivity draws ($\underline{\theta}$). After substituting for A (defined by (8)) and $\widetilde{r}_i(A, \theta_i)$ (defined by (7)) in (9), we can solve for x.

Consider first monopolistic competition. As per Section 7, $a_i^* = \arg\max \frac{Zh(a_i, \theta_i)}{A}$ is independent of A. Suppose the economy is scaled up k-fold, which means the market size becomes kZ and the mass of potential firms becomes kN. Then, consistent with Melitz (2003, pp. 1705-6), a remains the same, A increases by the same proportion, and x remains the same.[22]

The following proposition states how a ZPSEE oligopoly outcome differs from the monopolistic competition, where we define $\varepsilon_A^a(\theta_i)$ as the elasticity of i's action with respect to A.

Proposition 8 *Suppose the economy scales up k-fold. Under monopolistic competition, the aggregate scales up k-fold and the equilibrium variety scales up k-fold. In a ZPSEE, the aggregate increases more than k-fold, while the equilibrium variety increases less than k-fold if $\varepsilon_A^a(\theta_i) \geq \varepsilon_A^a(\underline{\theta})$ for all θ.*

Proof. See the Appendix. ■

Proposition 8 states that in a ZPSEE, if Z and N scale up k-fold, A increases more than proportionately, but variety increases less than proportionately. This implies that the increase in A is mainly driven by the increase in competition (mark-up effect) rather than variety. In the proof of Proposition 8 we show that x, the number of active marginal entrants in equilibrium, decreases. This is the reason for the decrease in variety. The decrease in x also implies that there is selection of better (more productive) firms into the market. Importantly, the elasticity condition given in Proposition 8 holds for the central cases of CES and logit demand systems.

In case of oligopoly, because actions are strategic complements, when A increases, a increases. For CES and logit demand systems, this implies that as A increases, prices fall and competition intensifies. This is the pro-competitive effect of trade that ensues under oligopoly.

[22] For this reason, Melitz (2003) notes that in the absence of trade costs, "trade allows the individual countries to replicate the outcome of the integrated world economy" (p. 1706). He then considers trade costs in order to bring out the impact of trade in the context of firm heterogeneity, specifically to show how trade results in reallocations between firms and increases the average productivity.

The pro-competitive effect of trade can also be analyzed by tracing what happens to the mark-ups, given by $\frac{h(a_i,\theta_i)}{a_i}$. In a logit model, this is equivalent to $(p_i - c_i)$. In the case of CES, it is the Lerner index.

Under monopolistic competition, since the mark-up is independent of A, it does not change when the market size increases. In a ZPSEE, the mark-up is $\frac{h^*(A,\theta_i)}{a_i^*(A,\theta_i)}$, and we want to see how this changes as the market size increases. It is straightforward to show that mark-ups decrease across the board. One question is whether the decrease in the mark-ups increases with θ_i. That is, do the more productive firms decrease their mark-ups by more? It is straightforward to verify that in the case of the logit and CES demand systems, although the mark-up is increasing in θ_i, the reduction in the mark-up is also increasing in θ_i. That is, the more productive firms have higher mark-ups to start with and they decrease their mark-ups by more when the economy is scaled up.[23]

Finally, we compare consumer welfare gains from trade under monopolistic competition and oligopoly. We are interested in seeing whether monopolistic competition underestimates the gains from trade because it does not take into account the pro-competitive impact of trade.

Proposition 9 *Suppose the economy scales up k-fold. For CES and logit demand systems, the increase in consumer welfare is higher under oligopoly than under monopolistic competition.*

Proof. We determine whether $\frac{dW}{dA}\frac{dA}{dk}$ is larger under oligopoly or monopolistic competition. From the zero-profit condition under monopolistic competition, we have

$$\frac{dA}{dk} = \frac{A}{k}$$

[23] The details are available from the authors on request.

Under oligopoly we have

$$\frac{dA}{dk} = \frac{-\left(h_a^2 - hh_{aa}\right)}{kh_a h_{aa}} = \left(\frac{A}{k} - \frac{h_a}{kh_{aa}}\right) > \frac{A}{k}$$

since $h_{aa} < 0$.

Under logit, $W = \mu \ln A$ and $\frac{dW}{dA} = \frac{\mu}{A}$. Hence, under monopolistic competition with a logit demand system, the welfare gain from a marginal increase in k is

$$\frac{dW}{dA}\frac{dA}{dk} = \frac{\mu}{A}\frac{A}{k} = \frac{\mu}{k}.$$

This is clearly lower than the marginal welfare gain under oligopoly since we established that $\frac{dA}{dk}$ is larger under oligopoly. A similar comparison holds for CES because $W = \frac{1}{\lambda}\ln A$. ■

We show in Section 7 that the value of the aggregate is the highest under monopolistic competition. This implies that consumer welfare will be higher under monopolistic competition than under oligopoly both before and after opening to trade, but Proposition 9 implies that the *gains* from trade will be higher under oligopoly.

9 Sub-aggregative Games and the Nested Logit

In this section, we show that our analysis of aggregative oligopoly games extends to a class of what we term sub-aggregative games. A leading example is the nested logit model, in which each nest yields a sub-aggregate, and these nest sub-aggregates can be aggregated to furnish an overall aggregate. We provide a toolkit for both the short-run and long-run (ZPSEE) analysis, and show that our main results still apply.

Suppose that the profit function of firm i can be written as a function of own action, an aggregator of firms' actions in the firm's immediate class, J, and an overall aggregator: $\pi_i\left(A, A_J, a_{iJ}\right)$. We are interested in games where, in the spirit of aggregative games, neither the composition of others' actions in the sub-aggregator nor the composition of others' actions outside the immediate class (or "nest" in the nested logit context) matters to profit.

Consider the nested logit structure.[24] The choice probability for option i in nest J is given by

$$\mathbb{P}_{iJ} = \mathbb{P}_{i|J}\mathbb{P}_J,\ J = 1, ..., N; i = 1, ..., n_J,$$

where n_J is the number of options in nest $J = 1, .., N$. Here, both of the choice probabilities on the RHS take a logit structure. Specifically, the probability of conditional choice of i from nest J is

$$\mathbb{P}_{i|J} = \frac{\exp \frac{(\alpha_i - p_i)}{\mu_J}}{\sum\limits_{j \in J} \exp \frac{(\alpha_j - p_j)}{\mu_J}}.$$

Taking again the action variable as $a_{iJ} = \exp \frac{(\alpha_i - p_i)}{\mu_J}$, we can write this as

$$\mathbb{P}_{i|J} = \frac{a_{iJ}}{A_J},$$

where we refer to the value A_J as the sub-aggregator for nest J.

The choice probability of nest J is

$$\mathbb{P}_J = \frac{\exp \frac{V_J}{\mu}}{\sum\limits_{I=1,...,N} \exp \frac{V_I}{\mu_1}} = \frac{a_J}{A},$$

where V_J is the attractiveness of a nest, given by the standard log-sum formula applied at the nest level

$$V_J = \mu_J \ln \left(\sum_{j \in J} \exp \frac{(\alpha_j - p_j)}{\mu_J} \right) = \mu_J \ln A_J,$$

$a_J = A_J^{\mu_J/\mu}$ denotes the transformation of the sub-aggregates, and A (the sum of the a_J) is the overall aggregate. Note that $\mu_J \leq \mu$ is McFadden's (1978) consistency condition for intra-nest substitution patterns to be more elastic than cross-nest ones.[25]

[24] See Ben-Akiva (1973) for the original development and Anderson et al. (1992) for a more detailed exposition.

[25] Note that if all the μ_J's are equal to μ, we have the standard logit structure with no nests, i.e., all variants are equally substitutable.

We can write firm i's profit in terms of the two levels of aggregate:

$$\pi_{iJ}\left(A, A_J, a_{iJ}\right) = \left(p_i - c_i\right)\mathbb{P}_{i|J}\mathbb{P}_J = \left(\alpha_i - \mu_J \ln a_{iJ} - c_i\right)\frac{a_{iJ}}{A_J}\frac{A_J^{\mu_J/\mu}}{A}.$$

Our assumptions to deal with the sub-aggregative game set-up are as follows. For A1 (competitiveness), we assume that own profit strictly decreases in both A and in A_J, so that higher aggregator values are harmful in both dimensions. It can be readily verified that this assumption holds in the nested logit example above. For A2a, we assume that profits are strictly quasi-concave so that the first order conditions deliver a unique maximum for reaction functions.

We now write the ibr as $\tilde{r}_i\left(A, A_J\right)$, which depends on both constituent aggregates. In the case of nested logit, $\tilde{r}$ is increasing in both arguments so there is strategic complementarity in both dimensions.[26]

9.1 Short-run analysis

Suppose first that the set of active agents is given, as is their membership into the various classes. To determine the short-run equilibrium, we proceed as follows. First, for a given value of the aggregate, $\overline{A}$, determine the ibr $\tilde{r}_i\left(\overline{A}, A_J\right)$ for each $i \in J$. Summing over all $i \in J$ and setting equal to A_J delivers the sub-class equilibrium value of A_J as a fixed point, namely $A_J^*\left(\overline{A}\right) = \sum_{i\in J}\tilde{r}_i\left(\overline{A}, A_J^*\right)$. This is shown in Panel A of Figure 3 (under the assumption that actions are strategic complements). Notice that the less aggressive are the firms (in terms of our earlier terminology of weaker ibrs), the smaller is the subsequent $A_J^*\left(\overline{A}\right)$.

We now proceed analogously for finding A. That is, the equilibrium value for A is determined as

$$A^* = \sum_{J\in N} A_J^*\left(A^*\right)$$

where N is the set of nests. This is shown in Panel B of Figure 3.

[26]The details can be found in Section B of the Online Appendix.

We can now perform comparative static analysis with the model. Take the merger example. When firms in the same nest merge, the combined reaction function for the nest falls. This implies that A falls in equilibrium, and so too do the other equilibrium A_J^*'s if the sub-aggregates are strategic complements. So all fall. For nested logit, prices rise everywhere, extending our previous results. One interesting difference is that if the merging parties are in different nests, the merged entity only internalizes the A effect, not the A_J effect. Hence, the impact of a within-nest merger on actions will be higher than the impact of an across-nest merger. In the case of nested logit, this means a within-nest merger would result in higher prices than an across-nest merger.

9.2 Long-run analysis

We now apply the ZPSEE to the sub-aggregative structure. Suppose that each nest comprises of a set of insider firms, and a set of symmetric marginal potential entrants, as before. Denote by $\pi_i^*(A, A_J) = \pi_i(A, A_J, \tilde{r}_i(A, A_J))$ the maximized value function, and assume that the various partial derivatives are such that (analogous to the arguments substantiating L3) $\pi_i^*(A, A_J)$ is decreasing in each argument. This implies that the direct effect coming through the competitiveness assumption is not overturned by the indirect effect coming through the ibr. We substantiate in Section B of the Online Appendix that the nested logit does indeed satisfy these properties.

We can then define the ZPSEE from the corresponding level curve (where the maximized profit is equal to the corresponding marginal entrant entry cost). The slope of this locus is given by

$$\frac{dA_J}{dA} = -\frac{\partial \pi_i^*(A, A_J) / \partial A}{\partial \pi_i^*(A, A_J) / \partial A_J}$$

and is negative under the partial derivative property.

There is such a ZPSEE curve for each sector. Summing them up yields the equilibrium value of A consistent with a ZPSEE in each sector. It also yields the sub-aggregator values in each sector, the A_J's. This construction is given in the top panel of Figure 4. From this, the individual actions in class J are determined, as per the lower panel of Figure 4. The number of marginal entrants is determined residually: since each chooses the same action, their number must ensure total actions sum to the purported equilibrium, A_J.

The following results now follow from the toolkit analysis. First, parallel to Proposition 1, any change to the insider firms does not change the ZPSEE equilibrium locus, and so does not change the overall aggregate, the constituent aggregates, and the actions of unaffected firms. The total welfare proposition is likewise unchanged: as long as consumer surplus depends only on the values of the sub-aggregates (note that nested logit has the stronger property that it depends only on the aggregate), then consumer surplus is unchanged, and the change in total surplus is just the change in insiders' rents. For a merger, whether it is across or within nests, the neutrality property of Proposition 6 continues to hold.

This toolkit allows us to analyze the welfare consequences of other changes. Suppose, for example, that K_J, the entry cost, falls for a nest. Then, A_J goes up, the other A_I's fall, and A rises because the increase in A_J dominates. As long as consumer surplus is increasing in the value of the sub-aggregates (and the composition does not matter, as is the case for nested logit), consumers are better off.

10 Income effects

The benchmark results in Section 4 rely on the assumption that consumer preferences are quasi-linear, i.e., there are no income effects. Although this assumption is commonly made in the literature focusing on partial equilibrium analysis, income effects are important in many contexts. For example, much of the trade literature assumes unit income elasticity (so, a richer country is just a larger poor country).

Results are more nuanced with income effects, but policy implications are stronger. We no longer get the neutrality result that the aggregate is the same, and we no longer have consumer surplus neutrality. With income effects, differences in profits under different market structures (redistributed to consumers) cause demand effects that affect the outcome. Ultimately, consumer welfare rises if and only if total profits rise.[27]

Suppose then that demands increase with income. We wish to analyze the case where consumer surplus depends only on the aggregate, which restricts attention to the IIA forms. We explicitly include profits in consumer income, Y, so we evaluate changes in consumer welfare incorporating extra income from profits (or losses).

Using Proposition 4, we can write the profit function for firm i in a multiplicatively separable form as

$$\pi_i = \omega_i(a_i)\,\sigma(A)\,\psi(Y),$$

where $\omega_i(a_i)$ denotes the i-specific profit component and $\psi(.)$ is an increasing function. Moreover, A3 implies that $\sigma(.)$ is a decreasing function. For example, consider the CES model with income share α devoted to the differentiated product sector. The demand for product i is $D_i = \frac{p_i^{-\lambda-1}}{\sum\limits_{j=1,\ldots,n} p_j^{-\lambda}}\alpha Y$, so $a_j = p_j^{-\lambda}$ (and $\psi(Y) = \alpha Y$).[28] Then, $\pi_i = (p_i - c_i) D_i = \frac{\omega_i(a_i)\alpha Y}{A}$, where $\omega_i(a_i) = \alpha a_i \left(1 - c_i a_i^{\frac{1}{\lambda}}\right)$, and $V = Y A^{\frac{\alpha}{\lambda}}$.

Proposition 10 *Assume that (i) demand satisfies the IIA property, and is increasing in Y; (ii) Y includes the sum of firms' profits; and (iii) consumer welfare V is*

[27] Consumer welfare here is total welfare because the profits are passed back to consumers.

[28] This is the classic demand generated from a representative consumer utility of the form $U = \left(\sum\limits_{j=1,\ldots,n} x_j^{\rho}\right)^{\frac{\alpha}{\rho}} x_0^{1-\alpha}$ where x_0 is consumption of the numeraire, x_j is consumption of variant j, and $\lambda = \frac{\rho}{1-\rho} > 0$, where the elasticity of substitution, $\rho \in (0,1)$ for (imperfect) substitute products. See, for example, Dixit and Stiglitz (1977).

increasing in both A and Y. Let $\mathcal{S}'$ and $\mathcal{S}''$ stand for the sets of firms in two ZPSEE, and suppose that total profits are higher in the second one. Then, $Y' < Y''$, $A' < A''$, and $V' < V''$.

Proof. Because the total profits are higher (and the marginal entrants make zero at both ZPSEE), $Y' < Y''$. The zero-profit condition for marginal entrants at the two ZPSEE are $\omega(a')\psi(Y')\sigma(A') = K$ and $\omega(a'')\psi(Y'')\sigma(A'') = K$. Since $Y' < Y''$ and $\psi(.)$ is an increasing function, it follows that $\omega(a')\sigma(A') > \omega(a'')\sigma(A'')$. Lemma 3 implies that $\omega(a^*)\sigma(A)$ is a decreasing function of A, so $A'' > A'$. Since V is increasing in both A and Y, $V' < V''$. ■

An important implication of Proposition 10 is that circumstances which are beneficial for firms (and hence cause Y to increase) are also a fortiori beneficial for consumers because the aggregate increases through the income effect. This reinforces the total welfare result we had in Section 4, without income effects. With income effects, when Y increases via extra profits (due to, e.g., a merger with synergies), total welfare increases because both the firms and the consumers are better off, through the twin channels of a higher income reinforced by a higher aggregate.

To illustrate Proposition 10, consider a reduction in marginal cost for some insider firm (a lower c_i) such that total profits rise. The increased consumer income increases the demand for each variant, ceteris paribus. Proposition 10 shows that the higher profits benefit consumers through both the extra income and also the higher value of A (as expressed through lower equilibrium prices and/or more variety). By contrast, when there are no income effects, there is no change in the aggregate because extra profits are spent solely on the numeraire.

Next consider a merger. If there are no synergies, profits of the merged entity are below those of the other non-merged firms (Proposition 6). In the long run, the merger makes a loss, which reduces consumer income. Profits go down, as does the aggregate and consumer surplus. If, however, there are sufficient synergies (expressed, e.g., through lower marginal production costs), then total profits after merger may be higher. In this case, welfare must be higher because the consumers are better off whenever the firms are better off.[29]

11 Heterogeneous entrants

We have assumed until now that the firms in $\mathcal{E}$ all have the same profit function. The simplest generalization is when firms differ by entry costs (differences in production costs and qualities are treated in the Online Appendix).

Suppose that firms from $\mathcal{E}$ have the same profit functions up to idiosyncratic K. Note that Lemmas 1, 2 and 3 still hold since they apply to the post-entry sub-games. Similar to a supply curve, rank firms by entry costs. Let $K(n)$ denote the entry cost of the nth lowest cost entrant. Assume the marginal firm earns zero profit. Then the equilibrium solution for any set of active firms, $\mathcal{S}$, is given by the fixed point condition $\sum_{i \in \mathcal{S}} \tilde{r}_i(A) = A$. By the sum-slope condition (2), the LHS has slope less than 1.

Suppose now that one insider j becomes more aggressive (in the sense of Lemma 4), and the equilibrium set of firms moves from $\mathcal{S}'$ to $\mathcal{S}''$. Proposition 1 implies that if all firms in E have the same entry cost, such a change increases a_j while leaving A and the actions of all other firms unchanged. These results now change:

Proposition 11 *Let entry costs differ across firms in* $\mathcal{E}$. *Let* $\mathcal{S}'$ *and* $\mathcal{S}''$ *be the sets of firms in two zero profit, free entry equilibria, and suppose that insider firm* j *is more aggressive in the second one. Then: (i)* $A' < A''$*; (ii) fewer firms are active; (iii) each*

[29] Shimomura and Thisse (2012) consider a model with CES demand and income effects to analyze mixed markets. They assume a given (small) number of large incumbents, which behave strategically, and a symmetric monopolistically competitive fringe. They show that an extra large incumbent raises profits for the other large firms, lowers the price index, and raises consumer welfare. Our results in Section 4 indicate that positive income effects are crucial for their results.

firm in $\mathcal{E}_{\mathcal{A}}$ and $\mathcal{I}_{\mathcal{U}}$ chooses a lower (higher) action if and only if actions are strategic substitutes (complements); and (iv) insider firm j chooses a higher action.

Proof. (i) Suppose instead that $A' \geq A''$. By Lemma 3, $\pi_i^*(A)$ is strictly decreasing. Hence, since $A' \geq A''$, $\pi_{\mathcal{E}}^*(A') \leq \pi_{\mathcal{E}}^*(A'')$. The equilibrium condition for a marginal active firm to make zero profit, $\pi_{\mathcal{E}}(A) = K(n)$, implies that $n' \leq n''$ since the marginal firm has a higher gross profit and hence a higher entry cost. If actions are strategic substitutes, this is a contradiction because at A'', there are purportedly more entrants, and the action of each is (weakly) greater. Moreover, firm j produces strictly more because firm j's ibr is higher (Lemma 4). Hence, we cannot have $A' \geq A''$.

The proof for strategic complementarity uses the sum-slope condition (2). For firms $i \in \mathcal{S}'$, $\sum_{i\in\mathcal{S}'} \tilde{r}_i(A') - \sum_{i\in\mathcal{S}'} \tilde{r}_i(A'') < A' - A''$. But $\sum_{i\in\mathcal{S}''} \tilde{r}_i(A'') > \sum_{i\in\mathcal{S}'} \tilde{r}_i(A'')$ because there are extra firms in $\mathcal{S}''$ and j is more aggressive (with a higher ibr by Lemma 4). Hence, $\sum_{i\in\mathcal{S}'} \tilde{r}_i(A') - \sum_{i\in\mathcal{S}''} \tilde{r}_i(A'') < A' - A''$, but equality must attain at any pair of equilibria, so there is a contradiction.

(ii) From Lemma 3, $\pi_i^*(A)$ is strictly decreasing. Since $A'' > A'$, then $\pi_{\mathcal{E}}^*(A'') < \pi_{\mathcal{E}}^*(A')$. The zero-profit condition for the marginal entrant, $\pi_{\mathcal{E}}(A) = K(n)$, implies that $n'' < n'$.

(iii) Lemma 2 implies that since $A' < A''$, firms choose a lower (higher) action iff actions are strategic substitutes (complements).

(iv) By definition, when firm j is more aggressive, it has a higher ibr (see Lemma 4). Since $A' < A''$, j chooses a higher action still if actions are strategic complements. Under strategic substitutes, suppose instead that j chose a lower action. But then the aggregate would have to be larger to overturn the impact of the shift in the ibr. From (ii) and (iii), there would be fewer firms in $\mathcal{E}_{\mathcal{A}}$ and each such firm would choose a lower action under strategic substitutes. Then, every action level would be smaller, which is inconsistent with the purported higher aggregate. Hence, firm j's action must be larger in both cases. ■

In contrast to the neutrality results of Section 4, a more aggressive firm raises the aggregate. For the Cournot model, this means a higher total output, and for the Bertrand model with logit or CES demand, a lower price (implying a higher total output). When consumer surplus increases in A, consumers must be better off.

Although the firm which experiences the change reacts positively to it by increasing its own action, whether the actions of all other firms increase or decrease depends on the sign of the slope of their ibr functions. By Lemma 3, because A rises, the firms which remain active must earn lower rents.

A merger without synergies works in the opposite direction: the aggregate falls, and despite further variety through entry, consumer surplus is lower. Hence, laissez-faire is no longer the optimal policy and an active merger policy is desirable because mergers, absent synergies, now reduce consumer surplus.

12 Discussion

This paper introduces a free entry condition into aggregative oligopoly games to yield strong benchmark conditions for long-run equilibria across market structures. We show how the benchmark neutrality results are modified when we consider income effects and entrants that are heterogeneous in costs and qualities.[30] Allowing income effects extends our strong result that higher profit entails higher welfare, but entrant heterogeneity means a necessary condition for welfare improvement is that producer surplus should rise.

[30] With heterogeneous entrants, the benchmark neutrality results change because the type of the marginal entrant differs between alternative market structures. This would also be the case if the difference between alternative market structures affected all the firms in $\mathcal{E}$. Consider, for example, two market structures with cost differences. In one of them, the marginal entrants are more aggressive. It is possible to show, by extending the analysis in Section C of the Online Appendix, that the aggregate will be higher under the market structure where the cost difference renders the marginal entrant more aggressive.

Our analysis shows the benefits of exploiting the aggregative structure in games with endogenous entry. It is well-acknowledged in the literature that aggregative games offer an attractive way to analyze games involving many heterogeneous players by reducing the dimensionality of the problem. However, their potential for analyzing long-run equilibria has not been explored so far.

We make several other contributions. First, we develop the toolkit for analyzing aggregative oligopoly games, which are ubiquitous in a range of fields from industrial organization to international trade to public economics. We relate the inclusive best reply to the standard best reply function, and show how the former simply delivers clean results. Strategic substitutability and complementarity of the best reply are preserved in the inclusive version. We derive a maximum value result to show that maximized profits decrease in the aggregate. This is a key device for analyzing long-run equilibrium. The simplicity of our analysis provides a basis on which models which assume monopolistic competition for reasons of tractability (e.g., in international trade) can deliver results with strategic interaction instead.

Second, we prove that consumer surplus depends only on the aggregate in Bertrand oligopoly games if and only if the demand function satisfies the IIA property. The central examples are Logit and CES models. This is important because it allows us to obtain welfare results in a range of applications where things would otherwise be intractable. Moreover, our results also show the extent to which some of the existing welfare results in the literature are "baked in" by the choice of the demand function.

Third, we explain how the toolkit can be extended in a straightforward way to apply to sub-aggregative games, and show that the benchmark neutrality results continue to hold in this case.

Fourth, we posit the combined inclusive best reply function as a simple tool for merger analysis. Using it, we show that mergers are socially desirable in the long run from a total welfare standpoint if and only if they are profitable. The analysis generalizes and explains results from the mergers literature that had been derived only for specific demand systems or forms of competition (Cournot or Bertrand).

Fifth, we derive a general proof of the proposition that monopolistic competition is “second best” from a welfare perspective under a zero profit constraint.

Sixth, we compare consumer gains from trade under monopolistic competition and oligopoly, and show that they are higher under oligopoly.

The aggregative game approach builds in global competition between firms. A key caveat is that it therefore builds in the neutrality results from the outset. Models of localized competition are quite intractable beyond simple symmetric cases (e.g. the circle model) or for small numbers of firms.[31] Yet they can suggest quite different results, with a wide divergence between optimal and equilibrium actions. Further work will evaluate these differences.

References

[1] Acemoglu, D. and M. K. Jensen. 2013. Aggregate Comparative Statics, *Games and Economic Behavior*, 81, 27-49.

[2] Alos-Ferrer, C. and A. B. Ania. 2005. The Evolutionary Stability of Perfectly Competitive Behavior, *Economic Theory*, 26, 497-516.

[3] Anderson, S. and A. de Palma. 2012. Oligopoly and Luce's Choice Axiom, *Regional Science and Urban Economics*, 42, 1053-1060.

[4] Anderson, S., A. de Palma, and J.-F. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. MIT Press: Cambridge, MA, USA.

[31] Special cases of localized competition which are aggregative games include the Hotelling model with two firms and the circular city model with three firms. Our short run results then apply.

[5] Anderson, S., A. de Palma, and J.-F. Thisse. 1997. Privatization and Efficiency in a Differentiated Industry, *European Economic Review*, 41, 1635-1654.

[6] Arkolakis, C., A. Costinot, and A. Rodriguez-Clare. 2012. New Trade Models, Same Old Gains?, *American Economic Review*, 102(1), 94-130.

[7] Atkeson, A. and A. Burstein. 2008. Pricing-to-Market, Trade Costs, and International Relative Prices, *American Economic Review*, 98(5), 1998-2031.

[8] Behrens, K. and Y. Murata. 2012. Trade, Competition, and Efficiency, *Journal of International Economics*, 87, 1-17.

[9] Ben-Akiva, M. 1973. Structure of Passenger Travel Demand Models, *Unpublished Ph.D. Dissertation*, M.I.T. Department of Civil Engineering, Cambridge, MA, USA.

[10] Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. Plants and Productivity in International Trade, *American Economic Review*, 93(4), 1268-1290.

[11] Corchón, L. 1994. Comparative Statics for Aggregative Games: The Strong Concavity Case, *Mathematical Social Sciences*, 28, 151-165.

[12] Corchón, L. 2001. Differentiable Comparative Statics with Payoff Functions Not Differentiable Everywhere, *Economics Letters*, 72(3), 397-401.

[13] Cornes, R. and R. Hartley. 2005. Asymmetric Contests with General Technologies, *Economic Theory*, 26, 923-946.

[14] Cornes, R. and R. Hartley. 2007a. Aggregative Public Good Games, *Journal of Public Economic Theory*, 9, 201-219.

[15] Cornes, R. and R. Hartley. 2007b. Weak Links, Good Shots and Other Public Good Games: Building on BBV, *Journal of Public Economics*, 91, 1684-1707.

[16] Cornes, R., and R. Hartley. 2012. Fully Aggregative Games, *Economics Letters*, 116(3), 631-633.

[17] Cournot, A. A. 1838. *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Paris: L. Hachette.

[18] Davidson, C. and A. Mukherjee. 2007. Horizontal Mergers with Free Entry, *International Journal of Industrial Organization*, 25(1), 157-172.

[19] de Blas, B. and K. Russ. 2010. Understanding Markups in the Open Economy under Bertrand Competition, *NBER Working Papers*, 16587.

[20] Deneckere, R. and C. Davidson. 1985. Incentives to Form Coalitions with Bertrand Competition, *RAND Journal of Economics*, 16, 473-486.

[21] Devereux, M. B. and K. M. Lee. 2001. Dynamic Gains from International Trade with Imperfect Competition and Market Power, *Review Of Development Economics*, 5(2), 239-255.

[22] Dixit, A. K. and J. E. Stiglitz. 1977. Monopolistic Competition and Optimum Product Diversity, *American Economic Review*, 67(3), 297-308.

[23] Epifani, P. and G. Gancia. 2011. Trade, Markup Heterogeneity and Misallocations, *Journal of International Economics*, 83(1), 1-13.

[24] Erkal, N. and D. Piccinin. 2010. Welfare-Reducing Mergers in Differentiated Oligopolies with Free Entry, *Economic Record*, 86(273), 178-184.

[25] Fujita, M., Krugman, P., and A. J. Venables. 2001. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press: Cambridge, MA, USA.

[26] Fujita, M. and J.-F. Thisse. 2002. *Economics of Agglomeration; Cities, Industrial Location and Regional Growth*. Cambridge University Press: Cambridge, UK.

[27] Goldman, S. M. and H. Uzawa. 1964. A Note on Separability in Demand Analysis, *Econometrica*, 32, 387-398.

[28] Grossman, G. M. and E. Helpman. 1993. *Innovation and Growth in the Global Economy*. MIT Press: Cambridge, MA, USA.

[29] Helpman, E. and P. Krugman. 1987. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. MIT Press: Cambridge, MA, USA.

[30] Holmes, T. J., W.-T. Hsu, and S. Lee. 2014. Allocative Efficiency, Mark-Ups, and the Welfare Gains from Trade, *Journal of International Economics*, 94(2), 195-206.

[31] Jensen, M. K. 2010. Aggregative Games and Best-Reply Potentials, *Economic Theory*, 43(1), 45-66.

[32] Krugman, P. R. 1979. Increasing Returns, Monopolistic Competition, and International Trade, *Journal of International Economics*, 9(4), 469-479.

[33] Martimort, D. and L. Stole. 2012. Representing Equilibrium Aggregates in Aggregate Games with Applications to Common Agency, *Games And Economic Behavior*, 76(2), 753-772.

[34] McFadden, Daniel L. 1978. Modelling the Choice of Residential Location, in A. Karlquist, L. Lundquist, F. Snickers, and J. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, 75-96. North Holland: Amsterdam.

[35] McManus, M. 1962. Numbers and Size in Cournot Oligopoly, *Bulletin of Economic Research*, 14, 14-22.

[36] McManus, M. 1964. Equilibrium, Numbers and Size in Cournot Oligopoly, *Bulletin of Economic Research*, 16(2), 68-75.

[37] Melitz, M. J. 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica*, 71(6), 1695-1725.

[38] Melitz, M. J. and G. I. P. Ottaviano. 2008. Market Size, Trade, and Productivity, *Review of Economic Studies*, 75(1), 295-316.

[39] Milgrom, P. and C. Shannon. 1994. Monotone Comparative Statics, *Econometrica*, 62(1), 157-180.

[40] Novshek, W. 1984. Finding All n-Firm Cournot Equilibria, *International Economic Review*, 25(1), 61-70.

[41] Novshek, W. 1985. On the Existence of Cournot Equilibrium, *Review Of Economic Studies*, 52(1), 85-98.

[42] Ottaviano, G. and J.-F. Thisse. 2004. Agglomeration and Economic Geography, in V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*. Vol. IV. Elsevier: Amsterdam, Netherlands.

[43] Phlips, L. 1995. *Competition Policy: A Game-Theoretic Perspective*. Cambridge University Press: Cambridge, UK.

[44] Salant, S. W., S. Switzer, and R. J. Reynolds. 1983. Losses from Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium, *Quarterly Journal of Economics*, 98, 185-199.

[45] Selten, R. 1970. *Preispolitik der Mehrproduktenunternehmung in der Statischen Theorie*. Springer Verlag: Berlin, Germany.

[46] Shimomura, K.-I. and J.-F. Thisse. 2012. Competition Among the Big and the Small, *RAND Journal of Economics*, 43(2), 329-347.

[47] Shubik, M. 1982. *Game Theory in the Social Sciences: Concepts and Solutions.* Cambridge: MIT Press.

[48] Spence, M. 1976. Product Selection, Fixed Costs, and Monopolistic Competition, *Review of Economic Studies*, 43(134), 217-235.

[49] Vives, X. 1999. *Oligopoly Pricing: Old Ideas and New Tools.* MIT Press: Cambridge, MA, USA.

[50] Wolfstetter, E. 1999. *Topics in Microeconomics: Industrial Organization, Auctions, and Incentives.* Cambridge University Press: Cambridge, UK.

APPENDIX

A Proof of Proposition 8

When the economy is scaled up k-fold, the equilibrium conditions (8) and (9) become

$$\frac{kZh^*(A,\underline{\theta})}{A} = K \tag{10}$$

$$kN\left(xG(\underline{\theta})\,\widetilde{r}_i(A,\underline{\theta}) + \int_{\theta>\underline{\theta}} \widetilde{r}_i(A,\theta)\,g(\theta)\,d\theta\right) = A \tag{11}$$

We show the impact on A by considering the elasticity of A w.r.t. k. Using the implicit function theorem on (10), we get the following expression:

$$\frac{k}{A}\frac{dA}{dk} = -\frac{k}{A}\frac{Zh^*(A,\underline{\theta})/A}{kZ\frac{dh^*(A,\underline{\theta})/A}{dA}} = -\frac{k}{A}\frac{1}{k}\left(\frac{A\left(h_a^2 - hh_{aa}\right)}{hh_{aa}}\right) = \left(1 - \frac{1}{A}\frac{h_a}{h_{aa}}\right) \tag{12}$$

which is larger than 1 as desired because $h_{aa} < 0$ from the second order condition.

We next consider how the equilibrium variety is affected when the economy scales up k-fold.[32] This is given by the change in the number of marginal entrants that are active in equilibrium, given by x. Letting $\Omega = \left(xG(\underline{\theta})\,\widetilde{r}_i(A,\underline{\theta}) + \int_{\theta>\underline{\theta}} \widetilde{r}_i(A,\theta)\,g(\theta)\,d\theta\right)$ in (11), we have

$$\begin{aligned} kN\Omega &= A \\ \ln kN + \ln\Omega &= \ln A \end{aligned}$$

Then,

$$\frac{d\ln\Omega}{d\ln k} = \frac{d\ln A}{d\ln k} - \frac{d\ln kN}{d\ln k}.$$

The first term on the RHS is > 1 (see (12)) and the second term is $= 1$. Hence,

$$\frac{d\ln\Omega}{d\ln k} = \varepsilon_k^A(\underline{\theta}) - 1 > 0.$$

[32] The result also follows from (10) since the zero-profit curve is above the rectangular hyperbola.

A change in k changes Ω through the twin channels of x and $\widetilde{r}_i(A,\theta_i)$. Suppose k increases. Since $\widetilde{r}_i(A,\theta_i)$ also increases, if $\varepsilon_k^a(\theta_i) > \varepsilon_k^A(\underline{\theta}) - 1$, x must decrease for (11) to continue to hold. This is what we aim to show. Note that

$$\varepsilon_k^a(\theta_i) = \frac{k}{a}\frac{d\widetilde{r}_i(A,\theta_i)}{dk} = \left(\frac{A}{a}\frac{d\widetilde{r}_i(A,\theta_i)}{dA}\right)\left(\frac{k}{A}\frac{dA}{dk}\right) = \varepsilon_A^a(\theta_i)\cdot\varepsilon_k^A(\underline{\theta}),$$

so we seek

$$\begin{aligned} \varepsilon_A^a(\theta_i)\,\varepsilon_k^A(\underline{\theta}) - \varepsilon_k^A(\underline{\theta}) &> -1 \\ \varepsilon_k^A(\underline{\theta})\left(1-\varepsilon_A^a(\theta_i)\right) &< 1. \end{aligned} \tag{13}$$

The slope of the ibr is

$$\frac{d\widetilde{r}_i(A,\theta_i)}{dA} = \frac{h_a^2}{h_a^2 - hh_{aa}},$$

so we have

$$\varepsilon_A^a(\theta_i) = \frac{A}{a}\frac{h_a^2}{h_a^2 - hh_{aa}}$$

which is equal to

$$\frac{1}{a}\frac{h}{h_a}\frac{h_a^2}{h_a^2 - hh_{aa}} = \frac{h_a}{h_a^2 - hh_{aa}}\frac{h}{a}$$

once we substitute for $A = \frac{h}{h_a}$ from the first order condition.

From here, it is straightforward to show that (13) holds if all θ are the same. Equation (13) becomes

$$-\left(\frac{h_a^2 - hh_{aa}}{hh_{aa}}\right)\left(1 - \frac{A}{a}\frac{h_a^2}{h_a^2 - hh_{aa}}\right) < 1$$

which simplifies to

$$-\frac{(h_a^2 - hh_{aa})}{hh_{aa}} + \frac{A}{a}\frac{h_a^2}{hh_{aa}} < 1$$

or

$$\frac{h_a^2}{hh_{aa}}\left(\frac{A}{a} - 1\right) < 0$$

which holds for $h_{aa} < 0$, as assumed.

For heterogeneous firms, (13) still holds if $\varepsilon_A^a(\theta_i) \geq \varepsilon_A^a(\underline{\theta})$ for all θ, which is what the proposition states.

Figure 1: Derivation of A from A_{-i}

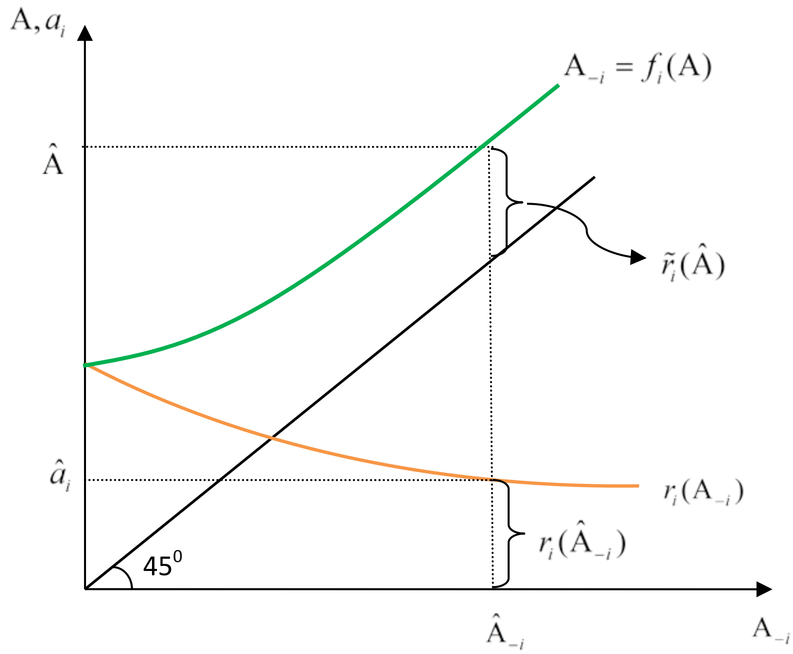


Figure 2: Construction of $\tilde{r}_i(\mathrm{A})$, Strategic Substitutes Case

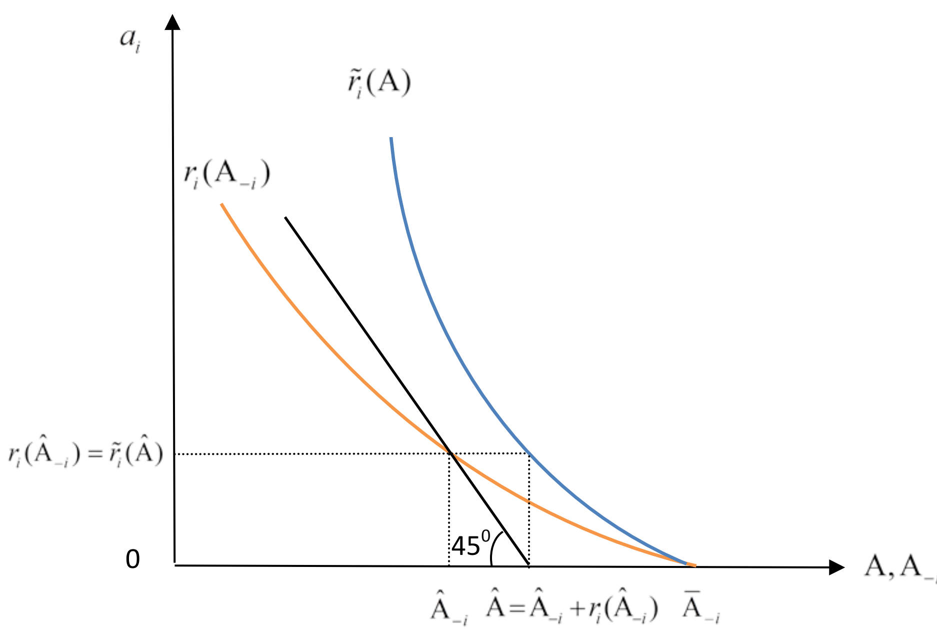
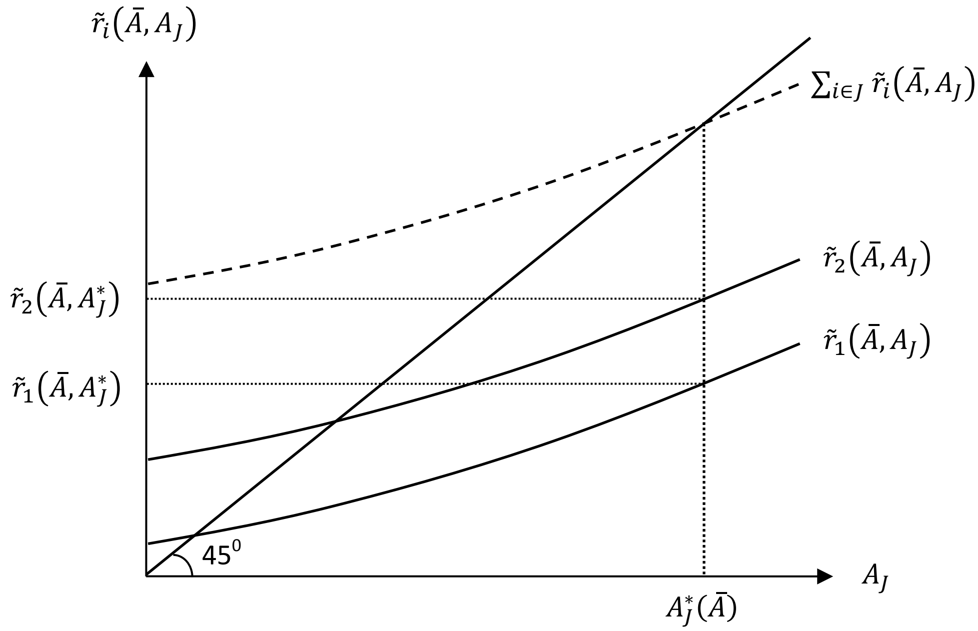
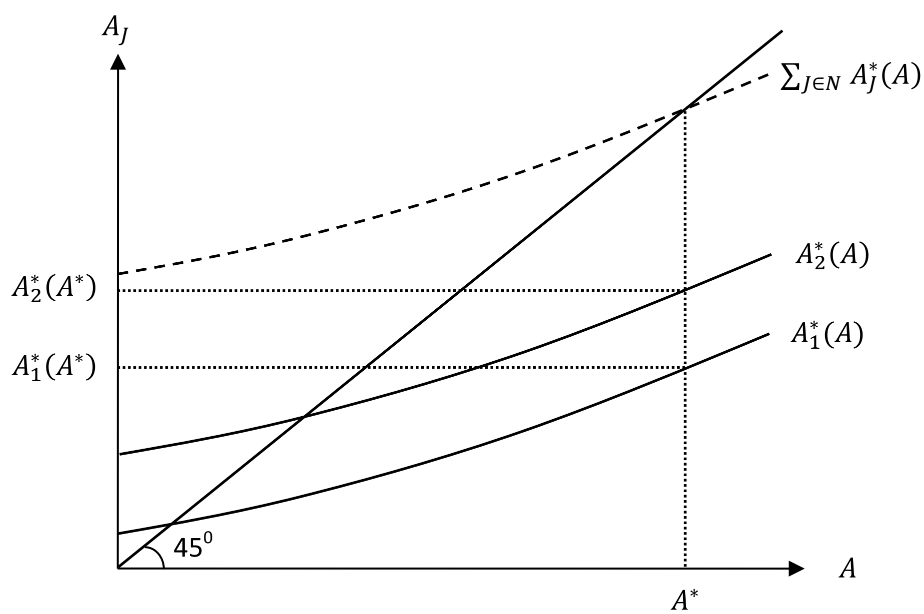


Figure 3: Sub-aggregative Games - Short Run

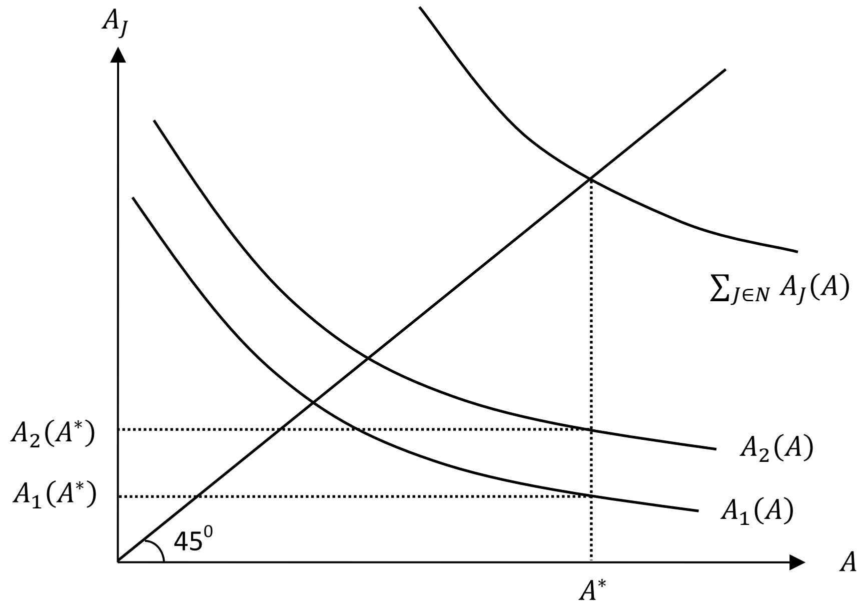


Panel A: The ibr and A_J^* for given $\bar{A}$

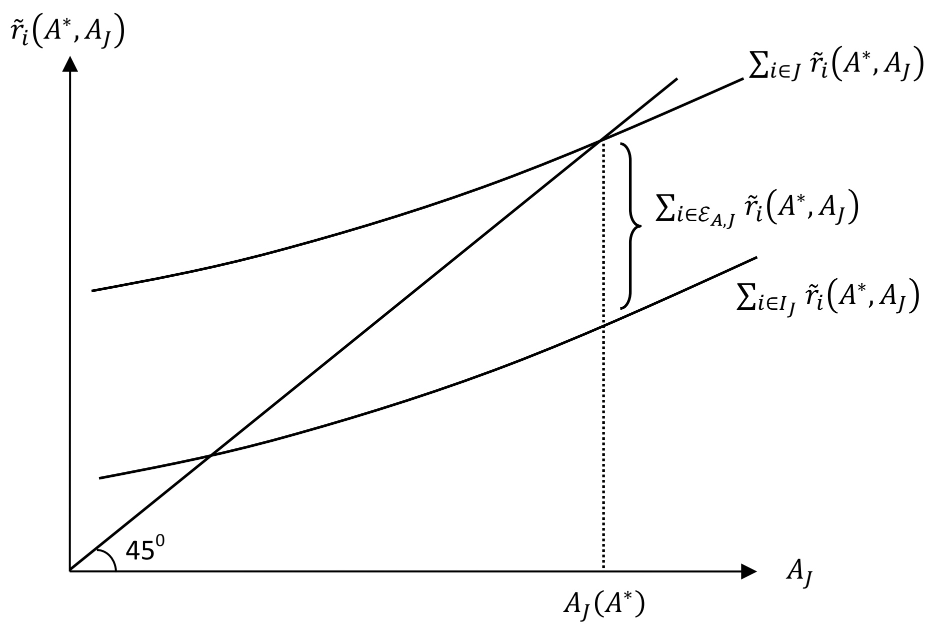


Panel B: Determination of the equilibrium values of the aggregate and sub-aggregates

Figure 4: Sub-aggregative Games - Long Run



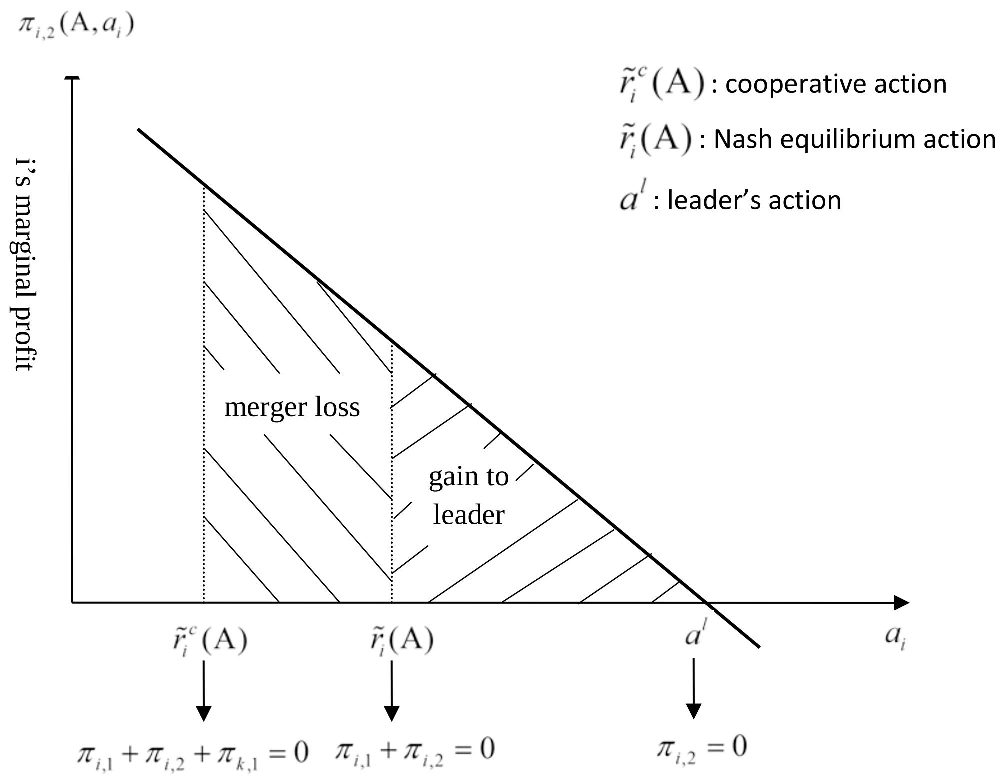
Panel A: ZPSEE level curves for each nest and the equilibrium A^*



Panel B: For given A^*, decomposition of equilibrium sub-aggregate, $A_J(A^*)$
(where I_J stand for the insiders in nest J and $\mathcal{E}_{A,J}$ stand for the active marginal entrants in nest J)

Figure 5: Comparison of Solutions by Profitability

(FOR THE ONLINE APPENDIX)



ONLINE APPENDIX

A Integer constraints

In this section, we take the integer constraint into account, and thus determine bounds on the equilibrium aggregate. We can then bound the welfare differences.

Let A_L and A_U stand for the lower and the upper bound on the equilibrium aggregate with a discrete number of firms and at least one firm in $\mathcal{E}_{\mathcal{A}}$. We seek the range of values of A such that there is no entry. If $A < A_L$, there will be entry. If $A > A_U$, there will be exit.

First note that the aggregate cannot exceed the equilibrium level when the number of firms from $\mathcal{E}$ is treated as a continuous variable. Hence, A_U is defined by $\pi^*_{\mathcal{E}}(A_U) = K$.

We determine the lower bound, A_L, by considering the incentives to enter. If a potential entrant from $\mathcal{E}$ expects the total reaction of the rivals to be $A_{-i} < A_U - \widetilde{r}_{\mathcal{E}}(A_U)$, it will enter the market. This is because, by definition, $\pi^*_{\mathcal{E}}(A_U) = K$, and so for all $A_{-i} < A_U - \widetilde{r}_{\mathcal{E}}(A_U)$, the potential entrant will expect to make positive profits upon entry. If it expects $A_{-i} \geqslant A_U - \widetilde{r}_{\mathcal{E}}(A_U)$, it will not enter.

Hence, $A_L = A_U - \widetilde{r}_{\mathcal{E}}(A_U)$ is a lower bound on A under strategic substitutability. The critical value of A will be higher than A_L because the incumbents tend to accommodate entry by reducing their equilibrium actions under strategic substitutability.

We can now show the following result.

Proposition 12 *Assume actions are strategic substitutes. Any Symmetric Entrants Equilibrium with at least one firm in $\mathcal{E}_{\mathcal{A}}$ must have $A \in [A_L, A_U]$, where A_U is defined by $\pi^*_{\mathcal{E}}(A_U) = K$ and $A_L = A_U - \widetilde{r}_{\mathcal{E}}(A_U)$.*

Proof. First, note that $A_U = \pi^{*-1}_{\mathcal{E}}(K)$ is the upper bound on the aggregate with at least one firm in $\mathcal{E}_{\mathcal{A}}$, since $\pi_i^*(A)$ is decreasing in A by Lemma 3 and A_U is the largest aggregate value at which a firm in $\mathcal{E}_{\mathcal{A}}$ can make a non-negative profit.

Second, we wish to show that any $A < A_L$ cannot be an equilibrium because it must attract profitable entry by a firm from $\mathcal{E}$. This is equivalent to showing that if an entrant joins the set $\mathcal{S}$ of active firms, then the ensuing equilibrium aggregate value will not be above $A_U = \pi_{\mathcal{E}}^{*-1}(K)$. Suppose this is not the case and suppose $A_{\{\mathcal{S}+i\}} > A_U$ with $i \in \mathcal{E}$. Since, by Lemma 2, $\tilde{r}_{\mathcal{E}}(A)$ is decreasing under strategic substitutes, each firm in $\mathcal{S}$ must be choosing a lower action than before (i.e., at the equilibrium with the extra firm from $\mathcal{E}$ excluded). This means their actions sum to less than A. Likewise, for the incremental firm in $\mathcal{E}_{\mathcal{A}}$, $\tilde{r}_{\mathcal{E}}\left(A_{\{\mathcal{S}+i\}}\right) < \tilde{r}_{\mathcal{E}}(A_U) = A_U - A_L$ under strategic substitutability (also by Lemma 2). Hence, its action must be less than its action at A_U. But then the sum of the actions cannot exceed A_U, a contradiction.

■

Hence, the equilibrium value of the aggregate lies in $[A_L, A_U]$ for all market structures. The maximum consumer welfare difference across equilibria $\mathcal{S}'$ and $\mathcal{S}''$ is $|CW(A_U) - CW(A_L)|$.

For example, consider Cournot competition with linear demand $P = 1 - Q$ and zero variable costs. Then, $Q_U = 1 - \sqrt{K}$, because each firm would just make zero profit by producing its equilibrium output of $q_i = \sqrt{K}$. The best response function is given by $r_i(Q_{-i}) = \frac{1-Q}{2}$ and the ibr is given by $\tilde{r}_i(Q) = 1 - Q$. Hence, the lower bound is given as $Q_L = Q_U - \tilde{r}_i(Q_U) = 1 - 2\sqrt{K}$. Note that there are $n = \frac{1-\sqrt{K}}{\sqrt{K}}$ firms at Q_U if this is an integer. Suppose then we took out one (indifferent) firm. The new equilibrium total quantity is $\frac{n-1}{n}$. Substituting the value of n given above yields the actual lower bound as $\frac{1-2\sqrt{K}}{1-\sqrt{K}} > Q_L$. The ratio of Q_L to the actual bound gets small as K gets small, as indeed does the ratio $\frac{Q_L}{Q_U}$.

Finding the bounds for the strategic complements case is more tricky because $\sum \tilde{r}_i'(A)$ could be very close to 1. The equilibrium action with one less firm may be very far away from A_U. At the other extreme, if $\sum \tilde{r}_i'(A)$ is close to zero, then A_L is close to the bound found above for the case of strategic substitutes.

B Nested Logit

In this section, we show that in the case of nested logit, (i) $\frac{d\widetilde{r}(A,A_k)}{dA} > 0$, (ii) $\frac{d\widetilde{r}(A,A_k)}{dA_k} > 0$, and (iii) $\frac{dA_k}{dA} < 0$ along an iso-profit line.

Consider

$$\pi(A, A_k, a_i) = (\alpha - \ln a_i)\frac{a_i}{A_k} \cdot \frac{A_k^\mu}{A},$$

where $\mu = \frac{\mu_2}{\mu_1} \in (0,1)$, $c = 0$, $A_k = \sum a_i$, and $A = \sum A_k^\mu$. Let

$$\tilde{\alpha}(a_i) = \alpha - \ln a_i \tag{14}$$

$$T = \frac{1}{A_k} \cdot \frac{A_k^\mu}{A} = \frac{A_k^{\mu-1}}{A}.$$

Then,

$$\pi(A, A_k, a_i) = \tilde{\alpha}(a_i)a_i \cdot T. \tag{15}$$

The first order condition w.r.t. a_i is given by

$$\pi_1 \frac{dA}{da_i} + \pi_2 \frac{dA_k}{da_i} + \pi_3 = 0$$

where π_1, π_2, and π_3 stand for the partial derivative of π w.r.t. its first, second, and third argument, respectively. Since $\frac{dA_k}{da_i} = 1$, we have

$$\pi_1 \frac{dA}{da_i} + \pi_2 + \pi_3 = 0 \tag{16}$$

The partial derivatives of π are given by

$$\pi_1 = \tilde{\alpha}(a_i)a_i \frac{\partial T}{\partial A} = -\tilde{\alpha}(a_i)a_i \frac{A_k^{\mu-1}}{A^2} = -\frac{\tilde{\alpha}(a_i)a_i}{A}T$$

$$\pi_2 = \tilde{\alpha}(a_i)a_i \frac{\partial T}{\partial A_k} = \tilde{\alpha}(a_i)a_i \frac{(\mu-1)A_k^{\mu-2}}{A} = \frac{\tilde{\alpha}(a_i)a_i(\mu-1)T}{A_k}$$

$$\pi_3 = \frac{\partial \tilde{\alpha}(a_i)}{\partial a_i} a_i T + \tilde{\alpha}(a_i)T = -\frac{1}{a_i} a_i T + \tilde{\alpha}(a_i)T = -T + \tilde{\alpha}(a_i)T = T(\tilde{\alpha}(a_i) - 1)$$

Note that $\pi_3 > 0$ since π_1 and π_2 are < 0.

Substituting for π_1, π_2, π_3, and $\frac{dA}{da_i} = \mu A_k^{\mu-1} = \mu TA$ in (16) gives us

$$-\tilde{\alpha}(a_i)a_iT^2\mu + \frac{\tilde{\alpha}(a_i)a_i(\mu-1)T}{A_k} + T(\tilde{\alpha}(a_i)-1) = 0 \tag{17}$$

Dividing through by $-\tilde{\alpha}(a_i)a_iT$ gives

$$-T\mu + \frac{\mu-1}{A_k} + \frac{1}{a_i} - \frac{1}{\tilde{\alpha}(a_i)a_i} = 0 \tag{18}$$

Let $f(a_i)$ stand for the last two terms of this equation:

$$f(a_i) = \frac{1}{a_i}\left(1 - \frac{1}{\tilde{\alpha}(a_i)}\right) = \frac{\tilde{\alpha}(a_i)-1}{a_i\tilde{\alpha}(a_i)}.$$

Then, $f(a_i) > 0$ because $\pi_3 > 0$. (18) becomes

$$-T\mu + \frac{\mu-1}{A_k} + f(a_i) = 0 \tag{19}$$

Note that

$$f'(a_i) = -\frac{f(a_i)}{a_i} - \frac{1}{a_i^2\tilde{\alpha}(a_i)^2} < 0$$

Since $f(a_i) = \frac{\tilde{\alpha}(a_i)-1}{a_i\tilde{\alpha}(a_i)}$, we can also write $f'(a_i)$ as

$$f'(a_i) = -\frac{1}{a_i^2\tilde{\alpha}(a_i)} - f(a_i)^2.$$

To show that $\frac{dA_k}{dA} < 0$ along an iso-profit curve, we use the zero-profit condition of the marginal entrants:

$$\pi^*(A, A_k, \widetilde{r}(A, A_k)) = f$$

Hence,

$$\frac{dA_k}{dA} = \frac{-\partial\pi^*/\partial A}{\partial\pi^*/\partial A_k} = \frac{-\left(\pi_1 + \pi_3\frac{d\widetilde{r}(A,A_k)}{dA}\right)}{\pi_2 + \pi_3\frac{d\widetilde{r}(A,A_k)}{dA_k}} \tag{20}$$

We totally differentiate the first order condition to get $\frac{d\widetilde{r}(A,A_k)}{dA}$ and $\frac{d\widetilde{r}(A,A_k)}{dA_k}$. From (19) we have

$$f(a_i) = \mu T - \frac{\mu-1}{A_k} = \frac{\mu A_k^{\mu-1}}{A} - \frac{\mu-1}{A_k} \tag{21}$$

Totally differentiating gives us

$$f'(a_i)da_i = -\frac{\mu T}{A}dA + (\mu - 1)\left(\frac{\mu T}{A_k} + \frac{1}{A_k^2}\right)dA_k.$$

Hence,

$$\frac{d\widetilde{r}\,(A, A_k)}{dA} = \frac{-\mu(T/A)}{f'(a_i)} > 0$$

since $f'(a_i) < 0$, and

$$\frac{d\widetilde{r}\,(A, A_k)}{dA_k} = \frac{(\mu - 1)\left(\frac{\mu T}{A_k} + \frac{1}{A_k^2}\right)}{f'(a)} > 0$$

since $\mu - 1 < 0$ and $f'(a_i) < 0$.

Now, consider the numerator of (20). Substituting for π_1, π_3, $\frac{d\widetilde{r}(A,A_k)}{dA}$ and $f'(a_i)$, and simplifying gives us

$$\begin{aligned} &-\left(\pi_1 + \pi_3\frac{d\widetilde{r}\,(A, A_k)}{dA}\right) \\ &= \frac{T}{A}\left[\frac{-a_i\tilde{\alpha}(a_i)\left[\tilde{\alpha}(a_i) + (\tilde{\alpha}(a_i) - 1)^2\right] + \mu T(\tilde{\alpha}(a_i) - 1)a_i^2\tilde{\alpha}(a_i)^2}{-\tilde{\alpha}(a_i) - (\tilde{\alpha}(a_i) - 1)^2}\right]. \end{aligned}$$

First note that the denominator of the term inside the brackets is negative:

$$-\tilde{\alpha}(a_i) - (\tilde{\alpha}(a_i) - 1)^2 < 0$$

Now consider the numerator of the term inside the brackets:

$$\begin{aligned} &- a_i\tilde{\alpha}(a_i)\left[\tilde{\alpha}(a_i) + (\tilde{\alpha}(a_i) - 1)^2\right] + \mu T(\tilde{\alpha}(a_i) - 1)a_i^2\tilde{\alpha}(a_i)^2 \\ &\overset{sign}{=} \left[-\tilde{\alpha}(a_i) - (\tilde{\alpha}(a_i) - 1)^2 + \mu T a_i\tilde{\alpha}(a_i)(\tilde{\alpha}(a_i) - 1)\right] \end{aligned} \tag{22}$$

We can re-write (18) in the following way.

$$\frac{(\mu - 1)\,\tilde{\alpha}(a_i)a_i}{A_k} + (\tilde{\alpha}(a_i) - 1) = \mu T a_i\tilde{\alpha}(a_i)$$

Substituting for $\mu T a_i\tilde{\alpha}(a_i)$ in (22) and simplifying yields

$$-\tilde{\alpha}(a_i) + \frac{(\tilde{\alpha}(a_i) - 1)(\mu - 1)a_i\tilde{\alpha}(a_i)}{A_k} < 0$$

Hence, the numerator of (20) is > 0.

Next consider the denominator of (20). Substituting for π_2, π_3, $\frac{d\widetilde{r}(A,A_k)}{dA_k}$ and $f'(a_i)$, and simplifying yields

$$\pi_2 + \pi_3 \frac{d\widetilde{r}\,(A, A_k)}{dA_k} \overset{sign}{=} \tilde{\alpha}(a_i) \left[-1 + \frac{\mu(\tilde{\alpha}(a_i) - 1)a_i}{A_k} \right] \tag{23}$$

For (20) to be negative, (23) must be negative. From (21), after substituting for $f\,(a_i)$, we have

$$\frac{\tilde{\alpha}(a_i) - 1}{a_i \tilde{\alpha}(a_i)} = \mu \frac{A_k^{\mu-1}}{A} + \frac{1-\mu}{A_k}.$$

Solving for $\tilde{\alpha}(a_i) - 1$ yields

$$\tilde{\alpha}(a_i) - 1 = \frac{a_i \left(\mu \frac{A_k^{\mu-1}}{A} + \frac{1-\mu}{A_k} \right)}{1 - a_i \left(\mu \frac{A_k^{\mu-1}}{A} + \frac{1-\mu}{A_k} \right)}.$$

Substituting for $\tilde{\alpha}(a_i) - 1$ in (23) and simplifying yields

$$\frac{a_i}{A_k} \left(1 + \mu \frac{a_i}{A_k} \right) \left(\frac{A_k^{\mu}}{A} \mu + 1 - \mu \right) < 1.$$

Notice that $\left(\frac{A_k^{\mu}}{A} \mu + 1 - \mu \right) < 1$, so it suffices to show that $\frac{a_i}{A_k} \left(1 + \mu \frac{a_i}{A_k} \right) < 1$. Let $s_i = \frac{a_i}{A_k}$ denote the share of firm i in the sub-aggregate. Then, we want to show that

$$1 - s_i - \mu s_i^2 > 0.$$

If $\mu = 0$, the inequality holds. Consider the worst case, $\mu = 1$, in which case we want to show that $1 - s_i - s_i^2 > 0$. This is guaranteed for $s_i < 0.618$.

Now, note that in a given nest, in equilibrium, s_i is increasing in α. This implies that if there are two firms, the weaker one has at most 50% of the nest. In a ZPSEE, the marginal entrants have lower α. Since there are at least two firms, the restriction on s_i must be satisfied.

C Cost changes and producer surplus (rents)

In the rest of the Online Appendix, we consider further applications of the toolkit we developed. Consider two equilibria with cost or quality differences. For example, a selectively-applied exogenous tax or subsidy affects the marginal costs of firms (see, e.g., Besley, 1989; Anderson et al. 2001). Or, a government subsidizes production costs (Brander and Spencer, 1985) or R&D activities (Spencer and Brander, 1983) of domestic firms engaged in international rivalry and the number of foreign firms is determined by a free-entry condition.

Even if several firms are impacted, the total effect is the cumulative effect, so we can consider changes as if they happened one firm at a time. Thus, we analyze what happens if a single insider is affected. We distinguish between the total profit and the marginal profit effects on the changed firm's rents. Denote the changed firm i's type parameter by θ_i, and assume that $\partial\pi_i(A, a_i; \theta_i)/\partial\theta_i > 0$ so that a higher θ_i makes the firm better off *if* it does not change its action.

Proposition 1 implies that at a ZPSEE, A is unchanged if θ_i rises. From Lemma 4, a firm's equilibrium action rises at a ZPSEE if a change makes the firm more aggressive. Because A is the same, the number of entrants must be lower.

Proposition 13 *A higher θ_i raises firm i's rents at a ZPSEE if $\frac{\partial^2\pi_i(A,a_i;\theta_i)}{\partial\theta_i\partial a_i} \geq 0$.*

Proof. Since A is unchanged, we show that $\frac{d\pi_i^*(A;\theta_i)}{d\theta_i} > 0$ with A fixed. Indeed,

$$\frac{d\pi_i^*(A;\theta_i)}{d\theta_i} = \frac{d\pi_i(A, \tilde{r}_i(A); \theta_i)}{d\theta_i} = \pi_{i,2}\frac{\partial\tilde{r}_i(A;\theta_i)}{\partial\theta_i} + \pi_{i,\theta_i}. \tag{24}$$

The last term is positive by assumption; $\pi_{i,2} > 0$ by A1 and (1); $\partial\tilde{r}_i(A;\theta_i)/\partial\theta_i > 0$ by Lemma 4, so the whole expression is positive, as claimed. ■

The qualification $\frac{\partial^2\pi_i(A_{-i}+a_i,a_i;\theta_i)}{\partial\theta_i\partial a_i} \geq 0$ in Proposition 13 represents an increasing *marginal* profitability. If, however, marginal profits decrease with θ_i, there is a tension between the direct effect of the improvement to i's situation and the induced effect through a lower action.[33] There are examples in the literature where the response of rivals can overwhelm the direct effect (although we know of no examples using the free entry mechanism). Bulow et al. (1985) analyze multi-market contact where a purported benefit turns into a liability once reactions are factored in. The Cournot merger paradox of Salant et al. (1983) shows merging firms can be worse off.

D Leaders and followers

Etro (2006, 2007, and 2008) first introduced a Stackelberg leader into the free-entry model. His main results can be derived succinctly and his welfare conclusions can be extended using our framework. The game structure is amended to 3 stages. The leader incurs its sunk cost and chooses a_l, rationally anticipating the subsequent entry and follower action levels. Then the other potential entrants (i.e., the other firms in $\mathcal{I}$ and $\mathcal{E}$) choose whether or not to incur their sunk costs and enter. Finally, those that have entered choose their actions.

A first result on welfare is quite immediate:

Proposition 14 *Assume a Stackelberg leader, and that the subsequent equilibrium is a ZPSEE. Assume also that consumer surplus depends only on A. Then welfare is higher than at the Nash equilibrium, but consumer surplus is the same.*

[33] This tension is illustrated in an example where a cost improvement with a "direct" effect of raising profits may nonetheless end up decreasing them after the free entry equilibrium reaction. Consider a Cournot model with linear demand. Costs are $C_1(q) = (c+\theta)q_1 - \beta\theta$ for firm 1 and $C(q) = cq$ for all other firms. Output for each other firm is determined by $1 - Q - c = q$, and the zero profit condition is $q = \sqrt{K}$. Firm 1's cumulative best reply is $1 - Q - c - \theta = q_1$, so a higher marginal cost reduces its output. Hence, $q_1 = q - \theta = \sqrt{K} - \theta$. Since firm 1's equilibrium profit is $\pi_1^* = q_1^2 + \beta\theta - K$, then $\pi_1^* = \left(\sqrt{K} - \theta\right)^2 + \beta\theta - K$ at the ZPSEE. Hence, $\frac{d\pi_1^*}{d\theta} = -2\left(\sqrt{K} - \theta\right) + \beta = -2q_1 + \beta$. Notice that the "direct" effect of a marginal change in θ is $-q_1 + \beta$, which is the change in profit if all outputs were held constant (except for firm 1's, by the envelope theorem). Clearly, depending on the size of β, a positive direct effect can nonetheless mean a negative final effect, once we factor in the entry response and the output contraction of the affected firm.

Proof. The consumer surplus result follows because A is the same, given the outcome is a ZPSEE. Welfare is higher because the leader's rents must rise. It can always choose the Nash action level, and can generally do strictly better. ■

From Section 5, this welfare result covers all demand systems with the IIA property (including CES and logit) as well as the Cournot model.

The ibr $\tilde{r}_i(A)$ is implicitly defined by $\pi_{i,1}(A, \tilde{r}_i(A)) + \pi_{i,2}(A, \tilde{r}_i(A)) = 0$. A1 implies $\pi_{i,1}(A, \tilde{r}_i(A)) < 0$, so the second term must be positive at the solution. A Stackelberg leader rationally anticipates that A is unchanged by its own actions (Proposition 1), so its optimal choice of action is determined by

$$\pi_{i,2}(A, a_l) = 0. \tag{25}$$

Hence, by A2b, the leader's long-run action must be larger than that in a simultaneous-move game (see Lemma 5).

Proposition 15 *(Replacement Effect) Assume a Stackelberg leader, and that the subsequent equilibrium is a ZPSEE. Then its action level is higher, and there are fewer active marginal entrants although they retain the same action level.*

We term this the Replacement Effect because, with a fixed A, the leader would rather do more of it itself, knowing that it crowds out one-for-one the follower firms from $\mathcal{E}$. In some cases, the leader wants to fully crowd them out. For example, in the Cournot model with $\pi_i(Q, q_i) = p(Q) q_i - cq_i$, we have $\frac{\partial \pi_i(Q,q_i)}{\partial q_i} = p(Q) - c$, so the leader will always fully crowd out the firms from $\mathcal{E}$ since $p(Q) > c$ at a ZPSEE.

Finally, we compare with the short run, when the number of firms is fixed. A leader takes into account the impact of its action on the behavior of the followers. In contrast to (25), the leader's action is determined by

$$\pi_{i,1}(A, a_l) \frac{dA}{da_l} + \pi_{i,2}(A, a_l) = 0. \tag{26}$$

If actions are strategic complements, $dA/da_i > 1$. Since $dA/da_i = 1$ in a simultaneous-move Nash equilibrium, the leader acts less aggressively than it would in a simultaneous-move game. If actions are strategic substitutes (i.e., $dA/da_i < 1$), the leader acts more aggressively than it would in a simultaneous-move game.

The comparison of short-run and long-run equilibria is most striking for strategic complements. Consider Bertrand differentiated products. The leader sets a higher price to induce a higher price from the followers (so reducing A, as desired).[34] At the ZPSEE, by contrast, the leader sets a *lower* price (higher a_l) and all firms in $\mathcal{E}_{\mathcal{A}}$ have the *same* price, regardless of the leader's presence.

The merger and leadership results can be tied together with a simple graph. A2b (quasi-concavity) implies that firm i's marginal profit, $\pi_{i,2}(A, a_i)$, is decreasing. In Figure 5, firm i's profit is represented as the area under this derivative because A is determined at a ZPSEE independently of i's actions. The leadership point is the value of a_l where $\pi_{i,2}(A, a_l) = 0$. Clearly, it gives the highest profit of any solution. In comparison, the solution where i plays simultaneously with the other firms after entry involves $\pi_{i,1}(A, a_i) + \pi_{i,2}(A, a_i) = 0$. Hence, the action level is lower, and the corresponding profit level is lower (see Lemma 5). The smaller profit is the triangle in Figure 5.

Now consider merger. From Lemma 6, each merger partner chooses an even lower action level, so each now nets an even lower payoff. The trapezoid in Figure 5 shows the loss compared to simultaneous Nash equilibrium actions.

E Contests

Aggregative games are common in contests (starting with Tullock, 1967), where players exert effort to win a prize. We consider applications in R&D and lobbying.

[34] These results can be quite readily derived within our framework.

E.1 Cooperation in R&D

Starting with Loury (1979) and Lee and Wilde (1980), the standard approach to R&D competition assumes that the size of the innovation is exogenously given, but its timing depends stochastically on the R&D investments chosen by the firms through a Poisson process. Time is continuous, and firms share a common discount rate r. Firms choose an investment level x at the beginning of the race which provides a stochastic time of success that is exponentially distributed with hazard rate $h(x)$. A higher value of $h(x)$ corresponds to a shorter expected time to discovery. Suppose that $h'(x) > 0$, $h''(x) < 0$, $h(0) = 0$, $\lim_{x \to 0} h'(x)$ is sufficiently large to guarantee an interior equilibrium, and $\lim_{x \to \infty} h'(x) = 0$.

Following Lee and Wilde (1980), assume that each firm i pays a fixed cost K_i at $t = 0$ and a flow cost x_i as long as it stays active. Then firm i's payoff is

$$\frac{h_i(x_i) V_i - x_i}{r + \sum_{j \in \mathcal{S}} h_j(x_j)} - K_i,$$

where V_i is the private value of the innovation and $\sum_{j \in \mathcal{S}} h_j(x_j)$ is the combined hazard rate. Equivalently, each firm chooses $a_i = h_i(x_i)$. Hence, $A = \sum_{j \in \mathcal{S}} h_j(x_j)$ and we can write the firm's payoff function as $\pi_i(A, a_i) = \frac{a_i V_i - h_i^{-1}(a_i)}{r+A} - K_i$. This aggregative game satisfies assumptions A1-A3.

Using this set-up, Erkal and Piccinin (2010) compare free entry equilibria with R&D competition to free entry equilibria with R&D cooperation. Under R&D co-operation, partner firms choose effort levels to maximize their joint profits, and may or may not share research outcomes (Kamien et al. 1992). Proposition 1 implies that the total rate of innovation, $A = \sum_i h_i(x_i)$, is the same regardless of the type of cooperation. This is despite the fact that the number of participants in the R&D race is different. This surprising neutrality result implies that any welfare gain from R&D cooperation cannot be driven by its impact on total innovation.

E.2 Lobbying

Following Tullock's (1967) model of contestants lobbying for a political prize, write the probability of success for firm i exerting effort x_i as $\frac{h_i(x_i)}{\Omega+\sum_{j\in\mathcal{S}} h_j(x_j)}$, where $\Omega \geq 0$ represents the probability that the prize is not awarded to any lobbyist (see Skaperdas, 1996, for an axiomatic approach to contest success functions). Typically, the lobbying model is analyzed with fixed protagonists, but now introduce a free-entry condition for the marginal lobbyists. Results are direct from our core propositions and their extensions. Namely, comparing two equilibria, the aggregate is the same (as are marginal lobbyists' actions) and, hence, there is no difference in the total chance of success. If one scenario involves a "dominant" or leader lobbyist, that lobbyist will exert more effort in order to crowd out marginal entrants. The overall chance of success remains the same, so there is an efficiency gain because the same result is attained with less sunk cost, and the surplus gain is measured by the increase in surplus to the dominant lobbyist. A similar result attains if a lobbyist is more efficient (i.e., if its marginal effort is more aggressive in the sense of Lemma 4).[35]

F R&D subsidies

R&D subsidies are used in many countries throughout the world. This section uses some of the results derived in Section C to derive new results on the long-run impact of R&D subsidies.

Consider a subsidy program that affects only a subset of the firms in an industry (the firms in $\mathcal{I}_\mathcal{C}$). Suppose that, as in Lee and Wilde (1980), investment in R&D entails the payment of a fixed cost K_i at $t = 0$ and a flow cost, and that the subsidy decreases the recipient's marginal cost of R&D. The R&D subsidy causes the recipients' ibr

[35] See, e.g., Konrad (2009), pp. 72-76, for a discussion of rent-seeking contests with voluntary participation. See Gradstein (1995) on entry deterrence by a leading rent-seeker.

functions to shift up. Since actions are strategic complements in Lee and Wilde (1980), this causes the rate of innovation in the short run, A, to increase. Proposition 1 implies that the long-run rate of innovation is unchanged with the subsidy. Lemma 4 implies that the individual efforts of the firms in $\mathcal{I}_{\mathcal{C}}$ increase while Proposition 1 states that those of the firms in $\mathcal{I}_U$ and $\mathcal{E}_{\mathcal{A}}$ do not change, so the number of participants in the R&D race decreases. Finally, Propositions 2 and 13 imply that the expected profits of the subsidized firms in $\mathcal{I}_{\mathcal{C}}$ go up, and the expected profits of the firms in $\mathcal{I}_U$ remain unchanged.

These results imply that although the government can increase the rate of innovation in the short run by adopting a selective R&D subsidy policy, it cannot affect the rate of innovation in the long run.

G The Logit model with differentiated quality-costs

The analysis in Section 11 readily adapts to the case of firms with different quality-costs and the same entry cost, K. Anderson and de Palma (2001) consider this model, showing that higher quality-cost firms have higher mark-ups and sell more, while entry is excessive. We extend their results by determining the comparative static properties of the equilibrium.

Suppose that $\pi_i = (p_i - c_i) \frac{\exp(s_i - p_i)/\mu}{\sum\limits_{j=0,..,n} \exp(s_j - p_j)/\mu}$, where the s_j represent vertical "quality" parameters and $\mu > 0$ represents the degree of preference heterogeneity across products. The "outside" option has price 0 and "quality" s_0. Since we can think of firms as choosing the values $a_j = \exp(s_j - p_j)/\mu$, we can write $\pi_i = (s_i - \mu \ln a_i - c_i) \frac{a_i}{A}$.

Label firms by decreasing quality-cost so that $s_1 - c_1 \geq s_2 - c_2 \geq ... \geq s_n - c_n$. Let $\mathcal{S}$ be the set of active firms, i.e., the first n firms. The marginal firm, firm n, makes zero in a free-entry equilibrium.

Now suppose that an insider firm $j < n$ is more aggressive (it has a lower marginal cost, for example). Then the aggregate must rise (the argument follows the lines of the proof of Proposition 11). Fewer firms are active at the equilibrium where j is more aggressive, and each one except j has a higher action, meaning a lower mark-up. Intuitively, if j is more aggressive, conditions become more competitive and marginal firms are forced out. Consumers are better off because the aggregate has risen.

H Privatization of public firms

Anderson et al. (1997) use a CES model to compare free entry equilibria with and without privatization. Since the CES model has the IIA property, Proposition 4 applies: the game is aggregative, and consumer surplus depends only on the aggregate value.

When some firms are run as public companies, they maximize their contribution to social surplus. The public firms may make a profit at a ZPSEE, even though the private firms do not. Public firms price lower, but produce more. Following privatization, although consumers suffer from a price rise, this is exactly offset by the increase in product variety as new entrants are attracted by relaxed price competition (Proposition 2). This means privatization changes total welfare by the decrease in the rents of the public firms only. Profitable public firms ought not be privatized if entry is free, and if demands are well characterized by IIA.

References (Online Appendix)

[1] Anderson, S. and A. de Palma. 2001. Product Diversity in Asymmetric Oligopoly: Is the Quality of Consumer Goods Too Low?, *Journal of Industrial Economics*, 49, 113-135.

[2] Anderson, S., A. de Palma, and B. Kreider. 2001. Tax Incidence in Differentiated Product Oligopoly, *Journal of Public Economics*, 81, 173-192.

[3] Anderson, S., A. de Palma, and J.-F. Thisse. 1997. Privatization and Efficiency in a Differentiated Industry, *European Economic Review*, 41, 1635-1654.

[4] Besley, T. 1989. Commodity Taxation and Imperfect Competition: A Note on the Effects of Entry, *Journal of Public Economics*, 40(3), 359-367.

[5] Brander, J. A. and B. J. Spencer. 1985. Export Subsidies and International Market Share Rivalry, *Journal of International Economics*, 18, 83-100.

[6] Bulow, J., J. Geanakoplos, and P. Klemperer. 1985. Multimarket Oligopoly: Strategic Substitutes and Complements, *Journal of Political Economy*, 93, 488-511.

[7] Erkal, N. and D. Piccinin. 2010. Cooperative R&D under Uncertainty with Free Entry, *International Journal of Industrial Organization*, 28, 74-85.

[8] Etro, F. 2006. Aggressive Leaders, *RAND Journal of Economics*, 37, 146-154.

[9] Etro, F. 2007. *Competition, Innovation, and Antitrust*. Springer Verlag: Berlin, Germany.

[10] Etro, F. 2008. Stackelberg Competition with Endogenous Entry, *Economic Journal*, 118, 1670-1697.

[11] Gradstein, M. 1995. Intensity of Competition, Entry and Entry Deterrence in Rent Seeking Contests, *Economics and Politics*, 7(1), 79-91.

[12] Kamien, M., E. Muller, and I. Zang. 1992. Research Joint Ventures and R&D Cartels, *American Economic Review*, 82, 1293-1306.

[13] Konrad, K. A. 2009. *Strategy and Dynamics in Contests.* Oxford University Press: Oxford, UK.

[14] Lee, T. and L. L. Wilde. 1980. Market Structure and Innovation: A Reformulation, *Quarterly Journal of Economics*, 94, 429-436.

[15] Loury, G. 1979. Market Structure and Innovation, *Quarterly Journal of Economics*, 93, 395-410.

[16] Salant, S. W., S. Switzer, and R. J. Reynolds. 1983. Losses from Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium, *Quarterly Journal of Economics*, 98, 185-199.

[17] Skaperdas, S. 1996. Contest Success Functions, *Economic Theory*, 7, 283-290.

[18] Spencer, B. J. and J. A. Brander. 1983. International R&D Rivalry and Industrial Strategy, *Review of Economic Studies*, 50, 707-722.

[19] Tullock, G. 1967. The Welfare Costs of Tariffs, Monopolies, and Theft, *Western Economic Journal*, 5(3), 224-232.