Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 154 (2014) 112–125

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Free entry versus socially optimal entry ☆

Rabah Amir [a,*], Luciano De Castro [a], Leonidas Koutsougeras [b]

[a] *Department of Economics, University of Iowa, Iowa City, IA 52242, USA*
[b] *School of Economic Studies, University of Manchester, M13 9PL, UK*

Received 7 October 2013; final version received 30 July 2014; accepted 10 September 2014

Available online 16 September 2014

Abstract

This paper reconsiders the well-known comparison of equilibrium entry levels into a Cournot industry under free entry, second best (control of entry but not production) and first best (control of entry and production). Allowing for the possibility of limited increasing returns to scale in production, this paper generalizes the conclusion of Mankiw and Whinston (1986) [10], that under business-stealing competition, free entry yields more firms than the second-best solution. We also show that under-entry always holds under business-enhancing competition. This confirms the general intuition given by Mankiw and Whinston, which does not rely on the convexity of the cost function. The same result is shown to extend (at a similar level of generality) to the comparison between free entry and the first best socially optimal solution, irrespective of business-stealing. Three illustrative examples are provided, one showing that the second-best and free entry solutions may actually coincide.

JEL classification: C72; D43; L13

Keywords: Free entry; Entry regulation; Socially optimal entry; Cournot oligopoly; Supermodularity

☆ All three authors gratefully acknowledge the wonderful hospitality and stimulating atmosphere at the Hausdorff Institute at the University of Bonn during the summer of 2013, when this work was completed. The revised version of the paper has benefited from constructive suggestions from the referees and Xavier Vives as the Editor in charge.

* Corresponding author.
E-mail address: rabah-amir@iowa.edu (R. Amir).

http://dx.doi.org/10.1016/j.jet.2014.09.003

1. Introduction

In a drastic reversal of a long-standing belief, von Weizsacker [23] proved that free entry into a Cournot industry with identical firms, linear demand, quadratic production costs and fixed entry cost leads to excessive entry relative to a first-best socially optimal solution. He postulated an omniscient social planner who can dictate both the number of firms to enter the industry and their market conduct or production levels. Perry [15] derived the same conclusion but with a second-best social planner, one that decides the number of entering firms but not their market conduct. Suzumura and Kiyono [19] address both comparisons–free entry versus first and second-best entry–with competition based on conjectural variations (including Cournot as a special case), but no fixed (entry) costs. With the latter proviso, they extend both von Weizsacker's first-best and Perry's second-best results. Mankiw and Whinston [10], or MW, reconsider Perry's question in a general setting encompassing Cournot competition (among others) with convex costs and a fixed entry cost, and provide a rigorous analysis accounting for the integer nature of the number of firms. Nachbar et al. [13] consider the effect of sunk costs on the welfare comparison at hand.[1] Finally, Amir and Lambson [4] offer an extension to a form of dynamic competition.

The present paper compares free entry into a symmetric Cournot industry with both first-best and second-best socially optimal entry. The main result is that free entry leads to an excessive (resp., insufficient) number of firms relative to second-best planning *if and only if* a "business stealing" (resp., "business enhancing") effect is present, i.e. each firm's Cournot equilibrium output contracts (resp., expands) as more firms enter the market. When holding globally, these two effects correspond respectively to the properties of strategic substitutes and strategic complements of Cournot outputs. This result is obtained under the most general assumption that guarantees the existence of symmetric pure-strategy Cournot equilibrium (based on supermodularity conditions), namely that price or inverse demand falls faster than marginal cost in a global sense.[2]

The second result is that, relative to first-best planning, excessive entry prevails, irrespective of strategic substitutes/complements. For this case, the relevant separation is also into two cases, but depending on whether the cost function is convex or concave. Indeed, the first best number of firms is always one in the latter case, making it a special case for the issue of entry.

With respect to MW's elegant analysis of the second-best comparison, given Cournot competition in the second stage of the game, the present paper adds by establishing that business stealing is essentially necessary as well as sufficient, by not requiring that production costs be convex, or that industry output be monotone in the number of firms. By specifying a Cournot framework at the outset, on the one hand, the present paper forsakes relevant generality in modeling the structure of competition, but on the other hand gains in clarity in allowing assumptions to be placed directly on the primitives of the oligopoly model. The latter step is guided by the requirement that the same assumptions must in the first place guarantee existence of a Cournot equilibrium. With that in mind, the basic strategy underlying the present paper is that MW's result ought to be investigated with as much generality as possible, subject to the conditions for existence. One advantage is that attention need not be restricted a priori to symmetric market equilibria. Rather, asymmetric equilibria are essentially precluded by our basic assumptions. Furthermore, this approach also links together all the important underlying issues that are relevant to

[1] In contrast to the present literature review, past literature on this topic has not always distinguished whether a first best or a second best criterion is used, despite the obvious relevance of the issue.

[2] This translates into a reaction correspondence will all slopes above -1: A firm responds in such a way that total output rises whenever rivals' total output rises, a property consistent with both strategic substitutes and complements.

the main question at hand, namely existence, multiplicity of equilibrium and comparative statics as the number of firms changes. Aside from these differences, the present work has been directly inspired by MW's rigorous analysis, in particular in dealing with the integer constraint. Relative to Suzumura and Kiyono [19], the present set-up is more general for both comparisons as far as the Cournot model is concerned (conjectural variations are not considered here).

Since a key part of the added generality in the present analysis consists of allowing for increasing returns in the form of decreasing marginal cost (as opposed to fixed costs), it is important to note that the prevalence of this feature has been documented for manufacturing industries in several empirical studies: See Friedlaender et al. [9], Ramey [16], and Diewert and Wales [5].

The intuition behind the second-best comparison, as laid out in MW, is elementary. A potential entrant does not take into account the business-stealing externality it confers on rivals in making its entry decision. A social planner internalizes this externality and allows one extra firm only if its profit exceeds the social value of the output contraction that this would cause for other firms. Thus, entry is always more desirable privately than socially. Given the present paper's reliance on necessary and sufficient conditions, an analogous interpretation applies directly to the converse result that under a business-enhancing effect, free entry yields too few firms relative to the second-best solution. Thus, our analysis fully confirms MW's claim that the business-stealing effect is the key assumption behind the excess-entry result. In a Cournot oligopoly, the business-stealing effect has a much broader scope than the business-enhancing effect. In particular, the latter can hold globally only in the absence of variable costs (for details, see Amir [1]).

As to the first-best comparison, the intuition and proof are quite elementary, even though the second stage is not the same as for the free entry scenario. Under first-best, for a given number of firms, per-firm output is chosen to equate price and marginal cost, and is therefore always larger than under Cournot. Per-firm profit is thus zero, and hence lower than under Cournot. It follows that if the first best number of firms were to engage in Cournot competition, per firm profit would be positive, thus making further entry (beyond the first-best level) profitable.

Sections 2 and 3 deal with the second-best and first-best comparisons respectively. Section 4 provides an illustration of interest and Section 5 briefly concludes. Appendix A collects the proofs.

2. Free versus second-best entry

We begin by laying out the basic set-up and notation for the Cournot model and the two entry scenarios. A symmetric Cournot oligopoly is described by a triplet (P, C, n) where $P : R^+ \to R^+$ is the inverse demand function, $C : R^+ \to R^+$ is the (common) cost function, and n is the number of firms. The average cost curve is defined by $A(x) = C(x)/x$ if $x > 0$ and $A(0) = C'(0)$.

Let x denote the output of the firm under consideration, y the total output of its $(n-1)$ rivals, and z industry output ($z = x + y$). Let x_n, y_n, z_n denote the symmetric n-firm equilibrium sets for those variables, and π_n and W_n the corresponding per-firm profit and social welfare.

The profit function is $\Pi(x, y) = xP(x + y) - C(x)$, or with total output $z = x + y$ as decision,

$$\widetilde{\Pi}(z, y) = \Pi(z - y, y) = (z - y)P(z) - C(z - y).$$

Let $\triangle(z, y)$ denote the cross-partial derivative of $\widetilde{\Pi}$ with respect to z and y, i.e.,

$$\triangle(z, y) \triangleq -P'(z) + C''(z - y)$$

Clearly, both $\widetilde{\Pi}$ and $\triangle$ are defined on the lattice $\varphi \triangleq \{(z, y) : y \geq 0, z \geq y\}$.

We assume throughout that (i) $P(\cdot)$ and $C(\cdot)$ are smooth, $P'(\cdot) < 0$ and $C'(\cdot) \geq 0$, (ii) $P(x) < A(x)$ for all $x > \widehat{x}$, for some $\widehat{x} > 0$, and (iii) $\pi_1 > K$, where K is the market entry cost per firm.[3]

Here, (ii) implies that a firm's reaction curve $r(y) \triangleq \arg\max_x \Pi(x, y)$ is eventually zero, so that a firm's effective output set is bounded. (iii) says that at least one firm can survive in the market.

The free entry number of firms, N, is defined via (the subgame-perfect equilibria of) the standard two-stage entry game (see e.g., MW), as the integer solution to

$$\pi_N \geq 0 \quad \text{and} \quad \pi_{N+1} < 0.$$

The second-best number of firms, N^*, is defined as usual as the integer solution to

$$N^* \triangleq \arg\max_n \{W_n - nK\}.$$

For a rigorous analysis of the issue at hand, it is critical to include provisions for existence of (pure-strategy) Cournot equilibrium. Our guiding principle is that the relevant comparisons will be established under each set of minimal sufficient conditions that is known to ensure existence of Cournot equilibrium. Accordingly, the presentation is divided into two separate cases defined by $\Delta > 0$ and $\Delta < 0$. For each of these cases, we first provide a summary of known useful properties of Cournot equilibrium (for details and proofs, see Amir [1] and Amir and Lambson [3]).[4]

2.1. The case $\Delta > 0$

When $\Delta > 0$ globally, the case of convex costs is covered, but so is the case of mildly increasing returns to scale, as well as hybrids of these two cases. This condition is thus very general.

Below, we shall invoke (sometimes tacitly) the following known facts:

(i) The slopes of $r(\cdot)$ are all > -1 (so r has no downward jumps).[5] From this slope property alone (no quasi-concavity needed here), symmetric Cournot equilibria (and no asymmetric ones) exist.

(ii) The minimal (resp., maximal) equilibrium is Pareto-dominant for firms (resp., consumers). Thus these equilibria satisfy one of the most commonly used selection criteria. In addition, they possess well-defined and intuitive comparative statics properties, namely p_n and π_n always decrease in n.

(iii) For a non-extremal, locally stable equilibrium (for Cournot dynamics) for all $n \in \{n_1, ..., n_t\}$, $t > 1$, p_n and π_n decrease in n on $\{n_1, ..., n_t\}$, if the equilibrium set has the same cardinality for each n in $\{n_1, ..., n_t\}$. The latter condition is needed to identify the equilibrium unambiguously.[6]

---

[3] Our notation is different from MW's in that π_n and W_n do not include the entry cost K here.

[4] Several arguments and background results of this paper rely on insights from the theory of supermodular games (Topkis [21], Vives [24], Milgrom and Roberts [11], Milgrom and Shannon [12], Amir [2], and others). We refer to the book/survey treatments in Topkis [22] or Vives [25,26] for the relevant definitions and results. On Cournot oligopoly, see [1,8,14,17,18].

[5] Throughout, "larger" and "increasing" refer to the weak notions. Else the qualifier "strictly" is added.

[6] Without this one-to-one correspondence between equilibria, one cannot necessarily make meaningful statements for non-extremal equilibria. For instance, the middle of three equilibria for $n = n_1$ need not be comparable to one of the middle three equilibria out of five for $n = n_1 + 1$. When the parameter of interest is a real number, further restrictions will be needed to preclude non-generic features such as equilibria occuring at tangency or discontinuity points of r. The two extremal equilibria are free of all these complications. In contrast, Cournot-unstable equilibria typically satisfy the opposite (thus counter-intuitive) comparative statics properties. If r is continuous, the stable and unstable equilibria will

Below, we call an equilibrium regular as n varies if it is either an extremal equilibrium or it satisfies all the assumptions in (iii), and thus has the given intuitive comparative statics properties.

Proposition 1. *Let $\Delta > 0$ globally. For a common regular Cournot equilibrium for both games,*
(a) If $x_{N^*-1} \geq x_{N^*}$, then $N \geq N^* - 1$.
(b) If $x_{N^*+1} \geq x_{N^*}$, then $N \leq N^*$.

Alternatively, one may state a global version of this result by replacing the condition $x_{N^*-1} \geq x_{N^*}$ (resp., $x_{N^*+1} \geq x_{N^*}$) by a global strategic substitutes (resp., complements) assumption. Using the local conditions as stated has the advantage of preserving this key result even when the game is neither of strategic substitutes nor of strategic complements in a global sense, both possibilities being fully consistent with the key assumption that $\Delta > 0$ (as is easily confirmed in examples).

Going back to primitives, the Cournot game is *globally* of strategic substitutes (resp., complements) if $P(\cdot)$ is log-concave (resp., log-convex and $C(\cdot) = 0$). In the latter case, the local condition $x_{N^*+1} \geq x_{N^*}$ becomes even more attractive, since it may hold in the presence of costs. Indeed, $x_{N^*+1} \geq x_{N^*}$ is implied by strategic complementarity only on $[0, x_{N^*+1}]$, for which a sufficient condition on primitives is, say with linear costs cx, that $[P(\cdot) - c]$ is log-convex on $[0, z_{N^*+1}]$, or

$$\big[P(z) - c\big]P''(z) - \big[P'\big]^2(z) \geq 0 \quad \text{for all } z \in [0, z_{N^*+1}]. \tag{1}$$

For a log-convex $P(\cdot)$, (1) is more likely to hold the smaller z_{N^*+1} is, but (1) cannot hold for all z, unless $c = 0$ (to see this, take $z \geq P^{-1}(c)$ in (1)). In other words, global strategic complementarity is not consistent with the presence of non-trivial costs. For more on this point, see Amir [1].

Parts (a) and (b) together essentially provide a full characterization of the comparison between the number of firms under the two regimes. This characterization can be succinctly phrased as follows: Free entry leads to a socially excessive (resp., insufficient) number of firms whenever a business-stealing (resp., business-enhancing) effect is present in the industry. This confirms the crucial role played by the nature of the externality an entering firm confers on the existing firms, which provided the main economic interpretation that MW associated with the under-entry result. Since the assumptions validating insufficient entry are far more restrictive, one naturally expects over-entry to constitute the norm for Cournot industries. Furthermore, since strategic complementarity cannot hold globally for Cournot competition in the presence of production costs, under-entry can only hold in a local sense, and typically for highly concentrated industries.

The following examples serve as illustrations of the new result of this section, namely Proposition 1(b). Example 1 shows that when r is increasing (so x_n is increasing in n), N can be much smaller than N^*, so the conventional view on entry may actually hold in a robust manner.[7]

Example 1. Let $P(z) = \frac{1}{(z+1)^5}$ and $C(x) \equiv 0$. The profit function is $\Pi(x, y) = \frac{x}{(x+y+1)^5}$.

Clearly, $\Delta > 0$ here, and, for $n \leq 4$, there exists a unique Cournot equilibrium with

---

tend to alternate, along with their respective comparative statics properties. For further discussion of these issues, see Amir and Lambson [3] and Echenique [6].

[7] In all the examples in Perry [15] purported to disprove that $N \geq N^*$, Proposition 1(a) actually holds. In other words, at worst he had $N = N^* - 1$. Interestingly then, this is the first example in the entry literature that actually confirms (the strict version of) the conventional wisdom, that free entry can lead to insufficient entry.

$$x_n = \frac{1}{5-n}, \qquad \pi_n = \frac{(5-n)^4}{5^5}, \quad \text{and} \quad W_n = \frac{1}{4}\left[1 - \frac{(5-n)^4}{5^4}\right], \quad \text{for } n \le 4.$$

Upon computation, x_n and W_n increase, and π_n decreases, in n. As is easily verified,

$$W_1 = 0.1476, \qquad W_2 = 0.2176, \qquad W_3 = 0.2436, \qquad W_4 = 0.2496$$

and

$$\pi_1 = 0.08192, \qquad \pi_2 = 0.02592, \qquad \pi_3 = 0.00512, \qquad \pi_4 = 0.00032.$$

It is easy to see that for all K, the planner lets in at least as many firms as free entry, and for some K, many more firms than free entry would. In particular, for $K = 0.026$, we obtain $N = 1$ and $N^* = 4$. Hence, Proposition 1(b) holds: The conventional wisdom on insufficient entry can hold in a global sense under restrictive conditions (a globally upward-sloping reaction curve).

The qualitative features of this example are robust to the inclusion of costs, say $C(x) = cx$, as long as c is small enough that the game retains strategic complementarity over $[0, x_{N^*+1}]$, and K takes values that are compatible with this interval being the relevant range.[8] However, the associated computations become much more cumbersome. □

The second example is an affirmative reply to the question, can free entry ever coincide with the second-best outcome? This is motivated by the observation that exponential demand is log-linear, thus at the same time log-concave and log-convex, so both parts of Proposition 1 apply.

Example 2. Let $P(z) = ae^{-z}$, $a > 0$, and $C(x) \equiv 0$. The profit function is $\Pi(x, y) = axe^{-x-y}$.

The first order condition is $(1 - x)e^{-x-y} = 0$, so $r(y) \equiv 1$ (i.e., 1 is a dominant strategy). Hence

$$x_n = 1, \qquad z_n = n, \qquad \pi_n = ae^{-n}, \quad \text{and} \quad N = \log a/K, \quad \forall K \in [0, ae^{-1}).$$

The second-best problem is $\max_n\{\int_0^n ae^{-t}dt - nK\}$, or $\max_n\{a(1 - e^{-n}) - nK\}$, and thus

$$N^* = \log a/K = N.$$

Hence, *for every* $K \in [0, ae^{-1})$, free entry yields the socially optimal number of firms.

It can be shown that this example is non-robust to the inclusion of non-zero costs, which would give rise to strict strategic substitutes and thus excess entry (no closed-form solutions are possible).

2.2. *The case* $\Delta < 0$

When $\Delta < 0$, production has strong scale economies, which is not typical of many industries.

Here, we shall make use of the following well known facts (Amir and Lambson [3]):

(i) The slopes of $r(\cdot)$ are all < -1, so the game is strongly of strategic substitutes.

(ii) As r may have downward jumps in this case, quasi-concavity of the profit function in own output is needed to guarantee existence of a unique symmetric Cournot equilibrium for each n.

(iii) This symmetric equilibrium possesses counter-intuitive comparative statics properties with respect to n (e.g., p_n increases in n), though π_n decreases in n.

---

[8] With $c > 0$, the reaction curve is inverse U-shaped, and its argmax is a decreasing function of c (Amir [1]).

When $\Delta < 0$, there are also many asymmetric Cournot equilibria, each involving a subset $m \le n-1$ of inactive firms (producing zero output), and the other $n-m$ firms each producing x_{n-m} (Amir and Lambson [3]). However, for the free entry problem, the only Cournot equilibrium selection consistent with subgame perfection is the unique symmetric equilibrium with *all* entering firms active. Indeed, any inactive firm in the second stage would deviate in the first stage, i.e., not enter so as to save on its entry cost $K > 0$. For the planner's problem, as social welfare is maximized by one firm, the solution is monopoly for all second-stage Cournot equilibrium selections (monopoly is the market structure that maximizes both consumer surplus and industry profits).

Proposition 2. *Assume $\Delta < 0$ and $\Pi(\cdot, y)$ is quasi-concave for each y. Then $N \ge N^* = 1$. Furthermore, if $K = 0$, then $N = +\infty$ and $N^* = 1$.*

This result is intuitive but the extreme over entry with zero (or small) entry costs is noteworthy.

3. Free vs. first-best entry

In the first-best situation, the social planner is assumed to be fully in charge of the industry at hand, setting both the industry concentration level and the conduct of the firms in the industry.[9] Thus, his objective function is, with n denoting the number of firms and x the output per firm,

$$\max\left\{F(n,x) \triangleq \int_0^{nx} P(t)dt - nC(x) - nK : n \ge 1, x \ge 0\right\}. \tag{2}$$

Let $(\overline{N}, \bar{x})$ be the global argmax in (2), under the constraint that n is an integer.

We also need to consider the subproblem of maximizing welfare with respect to per-firm output with n fixed. For each n, let $\bar{x}_n$ denote the corresponding (first-best) per-firm output, or

$$\bar{x}_n \triangleq \arg\max\left\{\int_0^{nx} P(t)dt - nC(x) - nK : x \ge 0\right\}. \tag{3}$$

Let $\bar{x}_n, \bar{y}_n, \bar{z}_n, \bar{p}_n, \bar{\pi}_n$ and $\overline{W}_n$ denote the same equilibrium quantities as before but now corresponding to a symmetric n-firm optimal outcome for the first-best solution.

Our first result provides a characterization of this solution for use below.

Lemma 3. *Assume $\Delta > 0$ globally. Then the first best solution satisfies the following:*
(a) For each n, each firm produces more than at the corresponding Cournot equilibrium, or $\bar{x}_n \ge x_n$.
(b) For each n, each firm's profit is lower than at the Cournot equilibrium, or $\bar{\pi}_n \le \pi_n$.
(c) $\bar{x}_n$ is decreasing in n.

Unlike the comparison between free entry and the second-best solution, the present comparison is a priori complex due to fact that the first best planning problem is a static maximization

[9] While first-best social planning solutions are regularly investigated in various settings in industrial and public economics, they remain useful more as benchmark ideals rather than outcomes of much practical significance.

problem where the planner makes two simultaneous choices. However, since this is an optimization problem (and not a strategic game) without uncertainty, the first-best welfare maximization problem may be equivalently converted into a two-stage sequential optimization problem where the planner chooses n in the first period and then x in the second period. Together with the conclusions of Lemma 3, this observation underlies the approach that we shall follow in making the comparison at hand.

We are now ready for the results on excess entry relative to a first best standard. We organize the results in two propositions, since the conclusions are sharply different. The first deals with the case of convex costs; the second case deals with concave costs and yields an extreme outcome.

Proposition 4. *If C is strictly convex, then $N \geq \overline{N} - 1$.*

The proof formalizes the simple intuition, in line with Lemma 3(b), that if the first-best number of firms were to engage in Cournot competition, they would make positive (as opposed to zero) profits, thereby inviting further entry. Hence free entry would lead to more firms than the first-best solution.

The integer constraint requires extra care (see proof). Finally, existence of Cournot equilibrium here follows from convexity of C (since $\Delta > 0$), so quasi-concavity of profit is not needed.

For the case of concave costs, the result is fully intuitive (here, quasi-concavity is needed only to ensure existence of a symmetric Cournot equilibrium).

Proposition 5. *If C is concave and $\Pi(\cdot, y)$ is strictly quasi-concave for each y, then $N \geq \overline{N} = 1$. Furthermore, if $K = 0$, then $N = +\infty$ and $\overline{N} = 1$*

Thus, the implications of these two intuitive results are clear-cut. In contrast to the second best case, business-enhancing competition is not enough to imply insufficient free entry relative to the first best criterion. In fact, insufficient free entry never happens in this case.

4. Extension: Δ not uni-signed

For the second-best criterion, the overall analysis has restricted attention to cases where either $\Delta > 0$ or $\Delta < 0$, both in a global sense. This is motivated both by the need to include provisions for the existence of Cournot equilibria and by the nature of the comparative statics of equilibria as n varies. Unfortunately, this leaves out potentially important cases where Δ does not have a globally uniform sign (due to the presence of local scale economies). For such cases, the present analysis need not apply a priori. For illustration, we provide one such insightful example.

Example 3. Let $P(z) = \frac{1}{z+1}$ and $C(x) = \frac{1}{2}\log(x+1)$. Then $\Pi(x, y) = \frac{x}{x+y+1} - \frac{1}{2}\log(x+1)$. Here $\Delta(z, y) = \frac{1}{(z+1)^2} - \frac{1}{2(z-y+1)^2}$ changes signs on φ (so our results need not apply).

The reaction curve is given by the positive arc of the unit circle $r(y) = \sqrt{1-y^2}$ for $y \leq 1$.

Simple computations yield that (i) the (unique and symmetric) Cournot equilibrium has each firm producing $x_n = (n^2 - 2n + 2)^{-1/2}$, and (ii) x_n is decreasing in n, for all n. Social welfare is

$$W_n = \log\left(\frac{n}{\sqrt{n^2 - 2n + 2}} + 1\right) - \frac{n}{2}\log\left(\frac{1}{\sqrt{n^2 - 2n + 2}} + 1\right).$$

In particular, $W_1 = \frac{1}{2}\log 2 = 0.34657$, and $W_2 = \log(\sqrt{2} + 1) - \log(\frac{1}{2}\sqrt{2} + 1) = 0.34657$, so that $W_2 = W_1$ (exactly). In fact, W_n is single-peaked in n with a maximum at $n \simeq 1.36$.

Thus, with $K = 0$, $N^* = \{1, 2\}$, i.e., a second-best social planner is indifferent between monopoly and duopoly as the optimal choice! In addition, it is easy to see that the free entry number is then $N = \infty$. Furthermore, simple computations lead to $W_{N^*} = W_1 = 0.34657$, $W_\infty = 0.1931$. Thus with costless entry, the size of the welfare loss due to free entry is $0.34657 - 0.1931 = 0.15347$, or 44%!

However, for any $K > 0$, we have $N^* = 1$ (2 firms being suboptimal due to the double entry cost). Hence, as long as $K \leq \pi_1$, we know that $N \geq N^*$. Excessive free entry thus obtains.

As to the first-best solution, one clearly has $\overline{N} = 1$ (since $A(\cdot)$ is strictly decreasing in x) and $\bar{x} = +\infty$, for any $K \geq 0$. Yet, one can have the free entry number of firms N be as large as desired (by taking K sufficiently small). In particular, with $K = 0$, $N - \overline{N} = \infty$ and the (first-best) welfare loss due to free entry is also infinite (since, as is easily checked, $\overline{W} = \infty$).

Here, excess entry leads to drastic welfare losses (relative to both first and second-best) because more competition leads each firm to produce less and thus at higher average cost (as C is concave).

5. Conclusion

This paper has provided a comprehensive and rigorous treatment of endogenous entry by comparing free entry with both first-best and second-best regulated entry, for Cournot markets. The analysis is conducted under general conditions (allowing for increasing returns to scale) that ensure the existence of Cournot equilibrium for the case of identical firms. The main results are as follows: (i) free entry results in an excessive number of firms relative to a second-best optimum essentially if and only if per-firm Cournot equilibrium output is decreasing in the number of firms (in the relevant range); and (ii) free entry results in an excessive number of firms relative to the first-best optimum in all cases. Along with insightful illustrative examples, these conclusions confirm unambiguously the critical importance of business-stealing versus business-enhancing assumptions; the possibility that the conventional view can hold under robust, though restrictive, conditions, and the irrelevance of the nature of the returns to scale in production in the general intuition given by MW for the second-best result. Due to the overall high level of generality, the two-way results for the second-best criterion, and the fact that both first and second best comparisons are covered, we hope that these results will also serve to unify the existing literature on entry for the case of Cournot markets.

Appendix A

This Appendix contains all the proofs of the results of this paper, given in order of appearance.

Proof of Proposition 1. We give separate proofs for Parts (a)–(b). As optimal monopoly profit $\pi_1 > K$ by assumption, at least one firm can survive in the market and therefore one clearly has $N \geq 1$ and $N^* \geq 1$.

Part (a): Consider the same regular or extremal Cournot equilibrium (to ensure that π_n is decreasing in n) for both the free entry and the second-best scenario. We show that $\pi_{N^*-1} > K$.

First, by the mean value theorem, there exists $\widetilde{x}$ such that $x_{N^*} \leq \widetilde{x} \leq x_{N^*-1}$ and

$$C(x_{N^*-1}) - C(x_{N^*}) = C'(\widetilde{x})(x_{N^*-1} - x_{N^*}). \tag{4}$$

By definition of N^*, $W_{N^*} - N^*K \geq W_{N^*-1} - (N^* - 1)K$, i.e.,

$$\int_0^{z_{N^*}} P(t)dt - N^*C(x_{N^*}) - N^*K$$
$$\geq \int_0^{z_{N^*-1}} P(t)dt - (N^*-1)C(x_{N^*-1}) - (N^*-1)K. \tag{5}$$

With $\pi_{N^*-1} = x_{N^*-1}P(z_{N^*-1}) - C(x_{N^*-1})$, (5) can be rewritten as

$$\pi_{N^*-1} - K \geq x_{N^*-1}P(z_{N^*-1}) - \int_{z_{N^*-1}}^{z_{N^*}} P(t)dt - N^*\big[C(x_{N^*-1}) - C(x_{N^*})\big].$$

Using the fact that P is strictly decreasing,

$$\begin{aligned}
&\pi_{N^*-1} - K \\
&\quad > x_{N^*-1}P(z_{N^*-1}) - (z_{N^*} - z_{N^*-1})P(z_{N^*-1}) - N^*\big[C(x_{N^*-1}) - C(x_{N^*})\big] \\
&\quad = x_{N^*-1}P(z_{N^*-1}) - (z_{N^*} - z_{N^*-1})P(z_{N^*-1}) - N^*C'(\widetilde{x})(x_{N^*-1} - x_{N^*}), \quad \text{by (4)} \\
&\quad = N^*\big[P(z_{N^*-1}) - C'(\widetilde{x})\big](x_{N^*-1} - x_{N^*}).
\end{aligned} \tag{6}$$

Since $\Delta > 0$, the correspondence r can have no slope ≤ -1 (see [2] or [7]), and hence no downward jumps. Since, in addition, $y_{N^*-1} \leq y_{N^*}$ and $x_{N^*} = r(y_{N^*}) < x_{N^*-1} = r(y_{N^*-1})$, we claim that r is onto (or surjective) over the subset $[x_{N^*}, x_{N^*-1}]$ of its range, in the sense that, given $\widetilde{x}$ with $x_{N^*} \leq \widetilde{x} \leq x_{N^*-1}$, there exists (at least one) $\widetilde{y}$ such that $r(\widetilde{y}) = \widetilde{x}$. Moreover, we must have $y_{N^*-1} \leq \widetilde{y} \leq y_{N^*}$. This follows from a direct application of Tarski's intersection point theorem (Tarski [20] 290), the two maps being the horizontal (or constant) map $x = \widetilde{x}$ and (a selection of) r, both taken on the common domain $[y_{N^*-1}, y_{N^*}]$. Indeed, the map $x = \widetilde{x}$ is higher than r at y_{N^*-1} but lower at y_{N^*}. The map $x = \widetilde{x}$ is continuous (thus quasi-decreasing, see Tarski [20]) and r is quasi-increasing. Hence, the two maps must intersect at a point, $\widetilde{y}$, as described above.

Since the slope property of r is equivalent to $r(y) + y$ being increasing in y, it follows from the fact $y_{N^*-1} \leq \widetilde{y}$ that $z_{N^*-1} = y_{N^*-1} + x_{N^*-1} = y_{N^*-1} + r(y_{N^*-1}) \leq \widetilde{y} + r(\widetilde{y})$, so that $P(z_{N^*-1}) \geq P[\widetilde{y} + r(\widetilde{y})]$. Also, we must have $P[\widetilde{y} + r(\widetilde{y})] \geq C'(\widetilde{x})$ since $\widetilde{x} \in r(\widetilde{y})$. Hence, $P(z_{N^*-1}) \geq C'(\widetilde{x})$. Since (by assumption) $x_{N^*-1} - x_{N^*} \geq 0$, we conclude from (6) that $\pi_{N^*-1} > K$. Since π_n is decreasing in n (Amir and Lambson [3]), it follows that $N \geq N^* - 1$. □

Part (b): Assume $x_{N^*} \leq x_{N^*+1}$. Given any regular or extremal Cournot equilibrium, we show that $\pi_{N^*+1} < K$. We begin the proof with two preliminary observations for use below.

First, since P is strictly decreasing, we have

$$x_{N^*+1}P(z_{N^*+1}) - \int_0^{(N^*+1)x_{N^*+1}} P(t)dt < -\int_0^{N^*x_{N^*+1}} P(t)dt. \tag{7}$$

Second, the fact that $\Delta > 0$ here implies that the function $V_n(x) \triangleq \int_0^{nx} P(t)dt - nC(x)$ is concave in x (for fixed n), since $V_n''(x) = n[nP'(nx) - C''(x)] < 0$. Hence, for any $x \geq x'$,

$$\left[\int_0^{nx} P(t)dt - nC(x)\right] - \left[\int_0^{nx'} P(t)dt - nC(x')\right] \geq n\big[P(nx) - C'(x)\big](x - x'), \tag{8}$$

where use is made of the fact that $V_n'(x) = n[P(nx) - C'(x)]$.

By definition of N^*, $W_{N^*} - N^*K \geq W_{N^*+1} - (N^*+1)K$, i.e.,

$$\int_0^{z_{N^*}} P(t)dt - N^*C(x_{N^*}) - N^*K$$
$$\geq \int_0^{z_{N^*+1}} P(t)dt - \big(N^*+1\big)C(x_{N^*+1}) - \big(N^*+1\big)K. \tag{9}$$

Adding $\pi_{N^*+1} = x_{N^*+1}P(z_{N^*+1}) - C(x_{N^*+1})$ to both sides, (9) can be rewritten as

$$\begin{aligned}
&\pi_{N^*+1} - K \\
&\leq x_{N^*+1}P(z_{N^*+1}) - \int_0^{z_{N^*+1}} P(t)dt + N^*C(x_{N^*+1}) + \int_0^{z_{N^*}} P(t)dt - N^*C(x_{N^*}) \\
&< -\left\{\left[\int_0^{N^*x_{N^*+1}} P(t)dt - N^*C(x_{N^*+1})\right] - \left[\int_0^{N^*x_{N^*}} P(t)dt - N^*C(x_{N^*})\right]\right\}, \quad \text{by (7)} \\
&= -\big\{V_{N^*}(x_{N^*+1}) - V_{N^*}(x_{N^*})\big\} \\
&\leq -N^*\big[P\big(N^*x_{N^*+1}\big) - C'(x_{N^*+1})\big](x_{N^*+1} - x_{N^*}), \quad \text{by (8) since } x_{N^*+1} \geq x_{N^*} \\
&\leq -N^*\big[P(z_{N^*+1}) - C'(x_{N^*+1})\big](x_{N^*+1} - x_{N^*}) \\
&\quad \text{as } z_{N^*+1} = \big(N^*+1\big)x_{N^*+1} \geq N^*x_{N^*+1} \\
&\leq 0,
\end{aligned} \tag{10}$$

where the last inequality follows from the facts that $P(z_{N^*+1}) \geq C'(x_{N^*+1})$ and $x_{N^*+1} \geq x_{N^*}$. Hence, $\pi_{N^*+1} < K$. Since π_n is decreasing in n (Amir and Lambson [3]), the result follows. □

Proof of Proposition 2. Since $K \leq \pi_1$, $N \geq 1$. Both consumer and producer surpluses are maximized by $N^* = 1$.

That $N = +\infty$ when $K = 0$ follows from the continuity of the reaction function $r(\cdot)$. □

Proof of Lemma 3. (a) We show that $\bar{x}_n \geq x_n$, for every $n \geq 1$.

Recall that the planner's objective is

$$\max\left\{F(n,x) \triangleq \int_0^{nx} P(t)dt - nC(x) - nK : n \geq 1, x \geq 0\right\}. \tag{11}$$

The first order conditions for this maximization are the familiar relationships

$$\frac{\partial F(n,x)}{\partial x} = n\big[P(nx) - C'(x)\big] = 0 \tag{12}$$

and (this is stated here for later use)

$$\frac{\partial F(n,x)}{\partial n} = xP(nx) - C(x) - K = 0. \tag{13}$$

We first show that F is strictly concave in x. Taking $\frac{d}{dx}$ once more in (12), we see that $\partial^2 F(n,x)/\partial x^2 = n[nP'(nx) - C''(x)] < 0$ for all $x \geq 0$, since $\Delta > 0$. Hence, for fixed n, F is strictly concave in x. Evaluating (12) at the Cournot equilibrium output, we have $\partial F(n,x_n)/\partial x = n[P(nx_n) - C'(x_n)] > 0$, so that we can conclude from (12) that $\bar{x}_n > x_n$.

(b) We show that $\pi_n \geq \overline{\pi}_n$, for every $n \geq 1$, where $\overline{\pi}_n$ denote the per-firm profit resulting from (2), i.e. $\overline{\pi}_n \triangleq \bar{x}_n[P(n\bar{x}_n) - C(\bar{x}_n)]$. Consider

$$
\begin{aligned}
\pi_n &= x_n\big[P(nx_n)\big] - C(x_n) \\
&\geq \bar{x}_n P\big[\bar{x}_n + (n-1)x_n\big] - C(\bar{x}_n), \quad \text{by the Cournot property} \\
&> \bar{x}_n P\big[\bar{x}_n + (n-1)\bar{x}_n\big] - C(\bar{x}_n), \quad \text{since } \bar{x}_n > x_n \\
&= \bar{x}_n P\big(n\bar{x}_n - C(\bar{x}_n)\big) \\
&= \overline{\pi}_n.
\end{aligned}
$$

(c) We show that $\bar{x}_n$ is decreasing in n. To this end, we show that the argmax in (11) satisfies the strict dual single-crossing property in $(x;n)$, i.e. that for any $x' > x$ and $n' > n$,

$$
\left[\int_0^{nx} P(t)dt - nC(x) - nK\right] \geq \left[\int_0^{nx'} P(t)dt - nC(x') - nK\right] \implies
$$

$$
\left[\int_0^{n'x} P(t)dt - n'C(x) - n'K\right] > \left[\int_0^{n'x'} P(t)dt - n'C(x') - n'K\right],
$$

or equivalently that

$$
C(x') - C(x) \geq \frac{1}{n}\int_{nx}^{nx'} P(t)dt \quad \implies \quad C(x') - C(x) > \frac{1}{n'}\int_{n'x}^{n'x'} P(t)dt. \tag{14}
$$

A sufficient condition for (14) is $\frac{1}{n}\int_{nx}^{nx'} P(t)dt > \frac{1}{n'}\int_{n'x}^{n'x'} P(t)dt$, i.e., the function $H(n) \triangleq \frac{1}{n}\int_{nx}^{nx'} P(t)dt$, for $x' > x$ fixed, is strictly decreasing in n. Consider (treating n as a real variable w.l.o.g.)

$$
\begin{aligned}
H'(n) &= -\frac{1}{n^2}\int_{nx}^{nx'} P(t)dt + \frac{1}{n}\big[x'P(nx') - xP(nx)\big] \quad \text{by Leibnitz's integral rule} \\
&< -\frac{1}{n}(x'-x)P(nx') + \frac{1}{n}\big[x'P(nx') - xP(nx)\big], \quad \text{since } P \text{ is decreasing.} \\
&= \frac{x}{n}\big[P(nx') - P(nx)\big] \\
&< 0, \quad \text{since } x' > x.
\end{aligned}
$$

As (14) holds, $F(n,x)$ satisfies the strict dual single-crossing property in $(x;n)$. Hence $\bar{x}_n$, being an argmax of $F(n,x)$ for fixed n, is decreasing in n (Milgrom and Shannon [12]). □

Proof of Proposition 4. Since $\Delta > 0$ when C is convex, a symmetric Cournot equilibrium exists for any n. The aim of the proof is to show that, for any regular or extremal equilibrium, one has $\pi_{\overline{N}-1} \geq K$.

We need to consider the optimization problem (2) both with real choice variables and with the formulation with integer n. Since $C'' > 0$, the objective (11) is easily shown to be strictly quasi-concave in (x, n). Let $(\overline{N}, \bar{x})$ be the argmax in (11) with the integer constraint on the number of firms. Then there exists a unique pair of real numbers $(m, \bar{x}_m)$ satisfying (12)–(13), or

$$P(m\bar{x}_m) = C'(\bar{x}_m) \quad \text{and} \quad \bar{\pi}_m = \bar{x}_m P(m\bar{x}_m) - C(\bar{x}_m) = K$$

as well as $\overline{N} = m^+$ (the smallest integer larger than m) or $\overline{N} = m^-$ (the largest integer smaller than m). We consider the latter two cases separately.

If $\overline{N} = m^+$, then $\overline{N} - 1 \le m$ and, since $F(n, x)$ is strictly concave in n,

$$\begin{aligned}\left.\frac{\partial F(n,x)}{\partial n}\right|_m &= \bar{x}_m P(m\bar{x}_m) - C(\bar{x}_m) - K = \bar{\pi}_m - K = 0\\ &\le \left.\frac{\partial F(n,x)}{\partial n}\right|_{\overline{N}-1} = \bar{x}_{\overline{N}-1} P\big[(\overline{N}-1)\bar{x}_{\overline{N}-1}\big] - C(\bar{x}_{\overline{N}-1}) - K\\ &= \bar{\pi}_{\overline{N}-1} - K.\end{aligned} \tag{15}$$

Likewise, if $\overline{N} = m^-$, then $\overline{N} \le m$ and

$$\begin{aligned}\left.\frac{\partial F(n,x)}{\partial n}\right|_m &= \bar{x}_m P(m\bar{x}_m) - C(\bar{x}_m) - K = \bar{\pi}_m - K = 0\\ &\le \left.\frac{\partial F(n,x)}{\partial n}\right|_{\overline{N}} = \bar{x}_{\overline{N}} P(\overline{N}\bar{x}_{\overline{N}}) - C(\bar{x}_{\overline{N}}) - K = \bar{\pi}_{\overline{N}} - K.\end{aligned} \tag{16}$$

Since by Lemma 3, $\pi_n \ge \bar{\pi}_n$ $\forall n$, (15) and (16) imply that $\pi_{\overline{N}-1} \ge \bar{\pi}_{\overline{N}-1} \ge K$. Hence, $N \ge \overline{N} - 1$. □

Proof of Proposition 5. We claim that $\overline{N} = 1$. Indeed, rewriting (2) with industry output z and n as the decision variables and $A(\cdot)$ as the average cost curve yields the equivalent objective $\int_0^z P(t)dt - zA(z/n) - nK$. Clearly, for fixed z, this objective is decreasing in n if $A(\cdot)$ is decreasing. Since the latter is implied by the concavity of C, it follows that $\overline{N} = 1$. Hence, $N \ge \overline{N}$. □

References

[1] R. Amir, Cournot oligopoly and the theory of supermodular games, Games Econ. Behav. 15 (1996) 132–148.
[2] R. Amir, Sensitivity analysis in multisector optimal economic dynamics, J. Math. Econ. 25 (1996) 123–141.
[3] R. Amir, V.E. Lambson, On the effects of entry in Cournot markets, Rev. Econ. Stud. 67 (2000) 235–254.
[4] R. Amir, V.E. Lambson, Entry, exit, and imperfect competition in the long run, J. Econ. Theory (2003) 191–203.
[5] W. Diewert, T. Wales, Flexible functional forms and global curvature conditions, Econometrica 5 (1987) 43–68.
[6] F. Echenique, Comparative statics by adaptive dynamics and the correspondence principle, Econometrica 70 (2002) 833–844.
[7] A. Edlin, C. Shannon, Strict monotonicity in comparative statics, J. Econ. Theory 81 (1998) 201–219.
[8] C. Ewerhart, Cournot games with biconcave demand, Games Econ. Behav. 85 (2014) 37–47.
[9] A. Friedlaender, C. Winston, K. Wang, Costs, technology and productivity in the US automobile industry, Bell J. Econ. 14 (1983) 1–20.
[10] N.G. Mankiw, M.D. Whinston, Free entry and social inefficiency, RAND J. Econ. 17 (1986) 48–58.
[11] P. Milgrom, J. Roberts, Rationalizability, learning, and equilibrium in games with strategic complementarities, Econometrica 58 (1990) 1255–1278.
[12] P. Milgrom, C. Shannon, Monotone comparative statics, Econometrica 62 (1994) 157–180.

[13] J. Nachbar, B. Petersen, I. Hwang, Costs, accommodation, and the welfare effects of entry, J. Ind. Econ. 46 (1998) 317–332.
[14] W. Novshek, On the existence of Cournot equilibrium, Rev. Econ. Stud. L II (1985) 85–98.
[15] M. Perry, Scale economies, imperfect competition and public policy, J. Ind. Econ. 32 (1984) 313–330.
[16] V. Ramey, Nonconvex costs and the behavior of inventories, J. Polit. Economy 99 (1991) 306–334.
[17] J. Roberts, H. Sonnenschein, On the existence of Cournot equilibrium without concave profit functions, J. Econ. Theory 13 (1976) 112–117.
[18] R.J. Ruffin, Cournot oligopoly and competitive behavior, Rev. Econ. Stud. 38 (1971) 493–502.
[19] K. Suzumura, K. Kiyono, Entry barriers and economic welfare, Rev. Econ. Stud. 54 (1987) 157–167.
[20] A. Tarski, A lattice-theoretic fixed point theorem and its applications, Pacific J. Math. 5 (1955) 285–309.
[21] D. Topkis, Minimizing a submodular function on a lattice, Operations Res. 26 (1978) 305–321.
[22] D. Topkis, Supermodularity and Complementarity, Princeton University Press, 1998.
[23] C. von Weizsacker, A welfare analysis of barriers to entry, RAND J. Econ. 11 (1980) 399–420.
[24] X. Vives, Nash equilibrium with strategic complementarities, J. Math. Econ. 19 (1990) 305–321.
[25] X. Vives, Oligopoly Pricing: Old Ideas and New Tools, The MIT Press, 1999.
[26] X. Vives, Complementarities and games: new developments, J. Econ. Lit. 43 (2005) 437–479.