International Journal of Industrial Organization 70 (2020) 102605

Contents lists available at ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

Market size and competition: A "hump-shaped" result☆

Iris Kesternich[a], Heiner Schumacher[a], Johannes Van Biesebroeck[a,b,*], Iris Grant[c]

[a] KU Leuven, Belgium
[b] CEPR, UK
[c] InterMutualistisch Agentschap, Belgium

ARTICLE INFO

Article history:
Received 10 April 2019
Revised 21 February 2020
Accepted 2 March 2020
Available online 7 March 2020

JEL classification:
D43
L13

Keywords:
Competition
Market entry
Market size
Entry threshold ratio

ABSTRACT

An active empirical literature estimates entry threshold ratios (ETRs), introduced by Bresnahan and Reiss (1991), to learn about the impact of firm entry on competition. We show that in the standard homogeneous goods oligopoly model, there is no monotonic relationship with the price-cost margin, one measure for the strength of competition. Regardless of the shape of demand, the ETR is hump-shaped in the number of active firms. It can also increase with entry in the Salop model of product differentiation or in a game of repeated interactions where collusion is possible. Empirical applications should use caution and only interpret changes in the ratio as indicative of a change in competition when the number of firms is sufficiently large.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The entry threshold ratio (ETR) was introduced in a seminal paper of Bresnahan and Reiss (1991). It is defined as the ratio of two consecutive entry thresholds, which are both normalized by the number of active firms. The entry threshold S_n indicates the minimum market size or number of customers needed for n active firms to break even. The ratio $(S_n/n)/(S_{n-1}/(n-1))$ then measures the increase in market size per firm that is needed for an additional firm to be able to enter a market with $n-1$ incumbents without incurring a loss.

The ETR is intended to measure the rate at which variable profits fall with entry. If competition becomes more intense such that the price-cost margin falls when additional firms compete and fixed entry costs are nondecreasing in the order of entry, the ratio will exceed one. When market competition approaches the monopolistically competitive benchmark, additional entry no longer changes the price-cost margin and the ETR will converge to unity.

The original application studied a few local service professions, namely doctors, dentists, druggists, plumbers, and tire dealers, but it has since been applied in a wide range of circumstances. This includes industries as varied as banking (Feinberg and Reynolds, 2010), hospitals (Abraham et al., 2007), brewing (Manuszak, 2002), broadband

☆ We thank Frank Verboven for discussions on this topic. The authors gratefully acknowledge financial support through the Flemish Research Foundation – FWO (Kesternich, Schumacher, and Van Biesebroeck), KU Leuven research funds (Kesternich, Schumacher, and Grant) and Methusalem project METH/15/004 (Van Biesebroeck).

* Corresponding author.
E-mail addresses: iris.kesternich@kuleuven.be (I. Kesternich), heiner.schumacher@kuleuven.be (H. Schumacher), jo.vanbiesebroeck@kuleuven.be (J. Van Biesebroeck), iris.grant@intermut.be (I. Grant).

https://doi.org/10.1016/j.ijindorg.2020.102605
0167-7187/© 2020 Elsevier B.V. All rights reserved.

(Xiao and Orazem, 2011), newspapers (Pfann and Van Kranenburg, 2003), or TV stations (Nishida and Gil, 2014), among many more applications.[1]

While most applications identify entry thresholds solely from cross-sectional variation in market size, a few studies have relied directly on actual entry or exit events (Varela, 2018) or verified whether entry or exit occurs in the expected direction when the market size changes (Carree and Dejardin, 2007). The ETRs have also been adapted to product differentiation where entry can expand the market (Schaumans and Verboven, 2015) or where two types of competitors have asymmetric competitive effects (Dranove et al., 2003; Cleeren et al., 2010).

Each of the studies cited above reports a sequence of ETRs to gauge how the intensity of competition changes with the number of active firms. Using s_n for the per-firm entry threshold (S_n/n), an estimate of s_2/s_1 that exceeds one is interpreted as evidence that a duopoly is more competitive and leads to lower variable profits than a monopoly. Results are generally discussed relative to a benchmark of a diminished effect on the price-cost margin for each additional entrant, and an expectation that the ratio s_3/s_2 will be lower than s_2/s_1. However, we show that such a monotone decline in ETRs is not predicted by the standard model that assumes Cournot quantity competition and a linear demand curve. With constant marginal costs and no heterogeneity across firms, we show that $s_2/s_1 < s_3/s_2$, i.e., the ratio for the third entrant is higher than for the second, even though the price-cost margin is lower. In the benchmark model, ETRs decline monotonically only once there are more than three entrants, but it takes seven firms before it falls below the value for s_2/s_1.

The intuition for this surprising theoretical finding can be seen from the fact that market size per firm and industry variable profits change at the same rate, but in opposite directions, with the number of firms. The first rises and the second declines monotonically with aggregate quantity, and thus with the number of active firms, but they do not change at a constant rate. Entry has two opposing effects on aggregate profits, a negative effect through a reduced price-cost margin, but also a positive effect due to higher total output. The rate of decline is determined by the net effect. It turns out that profits decline only slowly initially, for low values of n, and the rate of decline increases with n over an initial range.

This hump-shaped pattern in the dependency of the ETR on the number of active firms is not limited to the simple benchmark model. We show that it also holds for more general forms of demand. However, when one considers only integer numbers of firms, the hump will not appear in the data if demand is highly inelastic. I.e. the peak of the hump might be that close to s_2/s_1 that $s_2/s_1 > s_3/s_2$. We further show that alternative models of competition can also generate a hump, i.e. a rising ETR with entry when the number of incumbents is low, but for different underlying reasons. In the product differentiation model of Salop (1979), the expansion of the relevant market size with entry lowers the necessary market size per firm and depresses the ratio for low n.[2] In a model of repeated interaction, it is well known that collusion is more likely to prevail when the number of competitors is low (e.g., Tirole, 1988). When collusion becomes unsustainable, the ETR jumps.

These insights are not only interesting from a theoretical perspective, they are important to keep in mind from an applied perspective. They imply that the evolution of estimated ETRs with the number of active firms does not map directly into a change in intensity of market competition or a change in the effect of entry on the price-cost margin.[3] The change in ETRs only becomes a reliable predictor for the evolution of competition once the number of firms is sufficiently large.[4]

A solution could be to add additional assumptions, for example, a functional form for demand and cost homogeneity, in which case more features of the economy could be identified. However, the ETR's appeal primarily comes from the light modeling and data requirements. The main message then is to avoid attaching any interpretation to the change from s_2/s_1 to s_3/s_2. It is particularly important to keep this in mind given that the approach is most frequently used for small markets with only a few active firms.

The remainder of the paper is organized as follows. In Section 2, we derive the hump-shaped pattern for the ETRs in a model of Cournot competition between homogeneous goods' producers. We use the case of linear demand for illustration and to discuss intuition, but then establish that it holds as well for more general demand functions. In Section 3, we show that a hump-shape pattern can occur for other modes of competition as well. In Section 4, we review implications for empirical work and Section 5 concludes.

2. Entry threshold ratios in Cournot oligopolies

2.1. An example with linear demand

We analyze the entry threshold ratio in the canonical oligopoly model of Cournot competition in quantities. We first consider the linear demand setting that was also used in Bresnahan and Reiss (1991) and provide some intuition for the

[1] The literature contains several more specialized applications, for example there are studies on health insurers entering the marketplaces created by the Affordable Care Act (Abraham et al., 2017), funeral homes (Chevalier et al., 2009), fertilizer plants (Itin-Shwartz, 2017), notaries (Lee, 2007), charitable nonprofits (Gayle et al., 2017), driving schools (Asplund and Sandin, 1999), African-American movie theaters (Gil and Marion, 2018), etc.

[2] The alternative model of product differentiation in Schaumans and Verboven (2015) would also generate a hump for most demand systems.

[3] Our finding is not about the *level* of ETRs for successive entrants or whether the ratios are above or below one. It pertains to what the *change* in ETRs implies for the change in competitive effect of the marginal entrant, or the *relative* competitive effect compared to previous entrants.

[4] In the one-shot Cournot oligopoly, the interpretation of a change in ETRs is unambiguous once there are three active firms, but other models might require more firms.

pattern that we uncover. In the following subsection we derive a general statement on how the ETR varies with the number of firms for an arbitrary demand function.

Consider a situation with n identical firms and constant marginal costs. Market demand takes the form of $Q = S(a/b - p/b)$, where S represents the total market size. Let q_i be the quantity produced by firm i such that $Q = q_1 + \ldots + q_n$ is the total market quantity. The inverse demand function is then

$$p = a - \frac{b}{S}Q. \tag{1}$$

The firms' constant marginal costs are c and firm i chooses its quantity q_i to maximize profit

$$\pi_i = \left(a - \frac{b}{S}Q - c\right)q_i - F.$$

Standard arguments show that there is a unique symmetric equilibrium and that the allocation in this equilibrium is as follows:

$$\begin{aligned}
\text{quantity per firm:}\quad & q_n^* = \frac{1}{n+1}\frac{S}{b}(a-c)\\
\text{total quantity:}\quad & Q_n^* = \frac{n}{n+1}\frac{S}{b}(a-c)\\
\text{equilibrium price:}\quad & p_n^* = \frac{a+cn}{n+1}\\
\text{markup per unit:}\quad & p_n^* - c = \frac{a-c}{n+1}\\
\text{industry variable profit:}\quad & n \times q_n^* \times (p_n^* - c) = \frac{n}{(n+1)^2}\frac{S}{b}(a-c)^2\\
\text{firm profit:}\quad & q_n^* \times (p_n^* - c) - F = \frac{1}{(n+1)^2}\frac{S}{b}(a-c)^2 - F
\end{aligned}$$

Denote by $\Pi(n)$ the industry variable profits in the symmetric equilibrium when there are n firms. The entry threshold per firm for a market with n firms is the number of customers that each firm needs to serve such that the n'th firm can enter the market without making a loss. Since firm profits decrease in n, the market size S_n that is needed to support n firms will increase in n. Setting firm profit equal to zero and solving for S_n, we find

$$S_n = (n+1)^2 \frac{Fb}{(a-c)^2}, \tag{2}$$

that is, S_n equals fixed costs over variable firm profits. The entry threshold per firm is therefore given by $s_n = S_n/n$ and the entry threshold ratio is defined as $g_n = s_n/s_{n-1}$ for $n \geq 2$. One can interpret it as the growth rate in the customer base per firm that is needed to sustain at least zero profits when the number of firms increases from $n-1$ to n. Market entry affects firms' strategic behavior and thus the markup per unit. Therefore, g_n measures the rate at which profits fall when the number of firms increases, which depends on both markup and quantity.

In the current example, the per-firm ETR equals

$$g_n = \frac{s_n}{s_{n-1}} = \frac{n-1}{n^2}\frac{(n+1)^2}{n}. \tag{3}$$

Note that the value of g_n does not evolve monotonically. Specifically, we have

$$g_2 = 1.125 \quad g_3 = 1.185 \quad g_4 = 1.171 \quad g_5 = 1.152 \quad g_6 = 1.134$$

Fig. 1 shows the evolution of ETRs in the linear demand case with continuous n. The ratio first increases with the number of active firms n, for $n \leq 3$. From $n \geq 3$ onwards it decreases monotonically and converges to one. We state this result formally.

Proposition 1. *In the canonical oligopoly model of competition in quantities with n firms, constant marginal costs, and linear demand, the (per-firm) entry threshold ratio* g_n *is hump-shaped in n. Specifically, we have* $g_2 < g_3$, $g_n > g_{n+1}$ *for all* $n \geq 3$, *and* $g_n \to 1$ *for* $n \to \infty$.

The intuition for this result can be seen from the fact that the ETR is identical to the ratio between industry variable profits when there are $n-1$ firms and profits when there are n firms: $g_n = \Pi(n-1)/\Pi(n)$. Industry variable profits are highest in a monopoly situation. They decrease in the number of firms, but the decline is not monotonic in n. Starting from the monopoly quantity q_1^*, a small quantity increase only has a second-order effect on industry variable profits. Thus, when the number of firms increases from $n = 1$ to $n = 2$, the drop in profits is smaller than when the number of firms increases from $n = 2$ to $n = 3$. We therefore obtain $g_2 < g_3$.

The evolution of industry variable profits with the number of active firms is determined by the combined effect of declining markups and rising output. In Table 1 we show both components separately. The factor that scales the markup per unit $(p_n^* - c)$ equals $1/(n+1)$, while the factor that scales total quantity Q equals $n/(n+1)$. The markup decreases in n

Fig. 1. Entry threshold ratio for linear demand.

Table 1
Evolution of industry variable profits in the number of firms.

number	markup per unit		total quantity		industry variable profits
of firms	level	change	level	change	change
n	$\frac{1}{n+1}$	$\log \frac{n}{n+1}$	$\frac{n}{n+1}$	$\log \frac{n^2}{(n-1)(n+1)}$	$\log(\Delta^{mu}) + \log(\Delta^{Q})$
1	$\frac{1}{2}$	–	$\frac{1}{2}$	–	–
2	$\frac{1}{3}$	$\log \frac{2}{3} = -0.405$	$\frac{2}{3}$	$\log \frac{4}{3} = 0.288$	-0.117
3	$\frac{1}{4}$	$\log \frac{3}{4} = -0.288$	$\frac{3}{4}$	$\log \frac{9}{8} = 0.118$	-0.170
4	$\frac{1}{5}$	$\log \frac{4}{5} = -0.223$	$\frac{4}{5}$	$\log \frac{16}{15} = 0.065$	-0.158
5	$\frac{1}{6}$	$\log \frac{5}{6} = -0.182$	$\frac{5}{6}$	$\log \frac{25}{24} = 0.041$	-0.141

Note: The changes are calculated as the log-ratio for the values for n and $n-1$.

at a gradually declining rate. The change drops to zero as the price asymptotes to the marginal cost. The increase in quantity with n is especially pronounced for the first few entrants, but the rate of increase rapidly falls for higher n.

The change in industry profits depends on the relative strength of these two effects and is shown in the last column of Table 1, which exactly equals $\log g_n$. As total output increases most (in percentage terms) when the industry goes from a monopoly to a duopoly, it compensates a large fraction of the decrease in markups and as a result industry profits only fall by 11.7 percent. For the change from $n = 2$ to $n = 3$, the markup decrease is more robust than the output increase, and industry profits fall more, by 17.0 percent. For even more entrants, the markup eventually goes to zero and the output to the competitive quantity, and both rates of change slowly converge to zero. As their absolute differences shrink, the percent change in industry profits also converges to zero.

A possible reason why the ambiguous relationship between changes in competition and changes in ETRs has not been noted before is that the numerical illustration in the original paper, in Table 1 of Bresnahan and Reiss (1991), shows ETRs that are not normalized by the number of firms. The statistics correspond to the S_{n+1}/S_n ratio, while the table heading mistakenly indicates the normalized ratio s_{n+1}/s_n. Note that Berry and Reiss (2007) p. 1858 use the correct statistics when discussing the same example. Moreover, Fig. 4 in Bresnahan and Reiss (1991) shows the evolution of estimated ETRs normalized by s_5, i.e., s_5/s_n rather than s_{n+1}/s_n, while the latter ratio has become a focal point in most of the applications of the framework.

Note that Proposition 1 holds for any linear demand model, i.e., the pattern is independent of the intercept a, slope b, marginal costs c, and fixed costs F. This raises the question how general the result is, which we address in the next subsection.

2.2. A general result for Cournot oligopolies

We now generalize the result on the entry threshold ratios in a Cournot oligopoly to more general forms of demand. Let market demand be given by $SQ(p)$, where S is again total market size and $Q(p)$ is a downward-sloping demand curve. Let $p(Q)$ be the corresponding inverse demand function. To ensure the existence of a unique, well-behaved symmetric equilibrium, we adopt the following assumptions on the inverse demand function: (i) $p(Q)$ is twice differentiable, $p'(Q)$ is strictly negative and bounded from below, (ii) demand is bounded, i.e., there exists a K such that if $Q \geq K$, then $p(Q) = 0$, and (iii) for all q_i and total rival quantity Q_{-i}, we have $q_i p''(Q_{-i} + q_i) + p'(Q_{-i} + q_i) < 0$.

There are n identical firms that produce quantities $q_1, \ldots, q_n$. Firm i chooses its individual quantity q_i to maximize

$$\pi_i = S(p(Q) - c)q_i - F.$$

Let q_n^* be the quantity that each firm produces in the symmetric equilibrium when there are n firms. The first-order condition for q_n^* to be optimal for an individual firm is

$$p(nq_n^*) + p'(nq_n^*)q_n^* - c = 0. \tag{4}$$

The industry variable profits in this equilibrium are

$$\Pi(n) = nSq_n^*(p(nq_n^*) - c).$$

The ETR can be derived directly from the zero profit condition as

$$g_n = \frac{s_n}{s_{n-1}} = \frac{n-1}{n}\frac{q_{n-1}^*[p((n-1)q_{n-1}^*) - c]}{q_n^*[p((n)q_n^*) - c]}. \tag{5}$$

Note that g_n equals $\Pi(n-1)/\Pi(n)$, the ratio between industry variable profits when there are $n-1$ and n firms, respectively. If we treat n as a continuous variable, it again holds that ETRs are hump-shaped. We state this result formally.

Proposition 2. *In the canonical oligopoly model of competition in quantities with n firms and constant marginal costs, the (per-firm) entry threshold ratio g_n is hump-shaped in n. Specifically, we have $1 < g_2 < g_{2+\Delta}$ if Δ is sufficiently small, and $g_n \to 1$ for $n \to \infty$.*

Proof. By applying implicit differentiation to the first-order condition in (4) we get

$$\frac{\partial q_n^*}{\partial n} = -\frac{q_n^*}{n+1}\frac{p'(nq_n^*) + q_n^* p''(nq_n^*)}{p'(nq_n^*) + q_n^* p''(nq_n^*)\frac{n}{n+1}}.$$

By the assumptions on $p(Q)$ this derivative is strictly negative (Ruffin, 1971). Using (4) we calculate how industry variable profits vary in n as

$$\Pi'(n) = -(n-1)Sq_n^*\left[p(nq_n^*) - c - n\frac{\partial q_n^*}{\partial n}p'(nq_n^*)\right].$$

Note that $\Pi'(n)$ (i) is continuous on $[1, \infty)$, (ii) equals zero at $n = 1$, and (iii) is strictly negative for $n > 1$. Given the equality between g_n and the variable profit ratio, we can write it as

$$g_n = \left[1 + \frac{\int_{n-1}^{n}\Pi'(\tilde{n})d\tilde{n}}{\Pi(n-1)}\right]^{-1}. \tag{6}$$

Note that $\Pi(n-1+\Delta) < \Pi(n-1)$ for any $\Delta > 0$. Hence, $g_n < g_{n+\Delta}$ will be satisfied if

$$\left|\int_{n-1+\Delta}^{n+\Delta}\Pi'(\tilde{n})d\tilde{n}\right| > \left|\int_{n-1}^{n}\Pi'(\tilde{n})d\tilde{n}\right|.$$

For $\Delta \le 1$ we can rewrite this inequality as

$$\left|\int_{n}^{n+\Delta}\Pi'(\tilde{n})d\tilde{n}\right| > \left|\int_{n-1}^{n-1+\Delta}\Pi'(\tilde{n})d\tilde{n}\right|.$$

For $n = 2$ this inequality is implied by the properties (i) to (iii) of $\Pi'(n)$. Hence, we have $g_2 < g_{2+\Delta}$ if Δ is sufficiently small. From property (iii) it also follows that $1 < g_2$. Next, observe that

$$\frac{\Pi'(n)}{\Pi(n)} = -\frac{n-1}{n}\left[1 - \frac{n}{n+1}\frac{q_n^* p'(nq_n^*)}{p(nq_n^*) - c}\frac{p'(nq_n^*) + q_n^* p''(nq_n^*)}{p'(nq_n^*) + q_n^* p''(nq_n^*)\frac{n}{n+1}}\right].$$

From the first-order condition in (4) we get $q_n^* = (p(nq_n^*) - c)/p'(nq_n^*)$, so for $n \to \infty$ we get $\Pi'(n)/\Pi(n) \to 0$. By equation (6), this directly implies that $g_n \to 1$ for $n \to \infty$. □

Proposition 2 establishes that the ETR is strictly greater than one for the second entrant and that it rises in n initially. Given that it converges to one in a mature market with many active firms, this implies that it reaches a maximum somewhere, before declining. However, if we limit attention to discrete numbers of firms, $n \in \mathbb{N}$, ETRs are not necessarily hump-shaped. In particular, if the maximum is reached before $n = 3$ and g_n declines rapidly in n, it might be the case that $g_2 > g_3$. We now consider an example where the hump might or might not show up with discrete numbers of firms, depending on the elasticity of demand.

Consider a Cournot oligopoly with constant elasticity demand. The aggregate demand function is given by $Q = \alpha S p^{-\beta}$, where $\alpha > 0$ and β is the constant price elasticity of demand parameter with $\beta > 1$. The inverse demand function is then

$$p = \left(\frac{\alpha S}{Q}\right)^{\frac{1}{\beta}}.$$

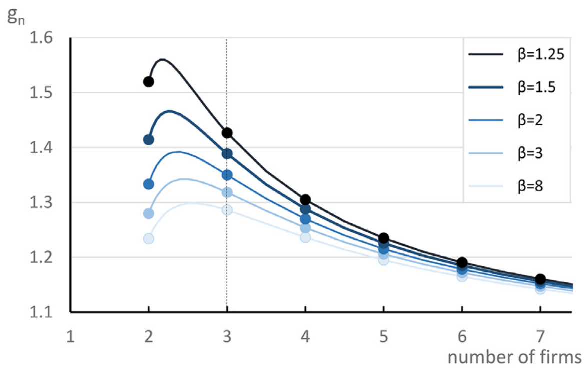


Fig. 2. Entry threshold ratios for various constant elasticity demand curves.

Firm i chooses its quantity q_i to maximize its profit

$$\pi_i = \left(\frac{\alpha S}{Q}\right)^{\frac{1}{\beta}} q_i - cq_i - F.$$

The unique symmetric equilibrium is characterized by the following values:

$$\text{quantity per firm:} \quad q_n^* = \frac{\alpha S}{n}\left(1 - \frac{1}{\beta n}\right)^{\beta} \frac{1}{c^{\beta}}$$

$$\text{total quantity:} \quad Q_n^* = S\alpha\left(\frac{\beta n - 1}{\beta n}\right)^{\beta} \frac{1}{c^{\beta}}$$

$$\text{equilibrium price:} \quad p_n^* = \frac{\beta nc}{\beta n - 1}$$

$$\text{markup per unit:} \quad p_n^* - c = \frac{c}{\beta n - 1}$$

$$\text{industry variable profit:} \quad n \times q_n^* \times (p_n^* - c) = \frac{\alpha S}{c^{\beta-1}} \frac{(\beta n - 1)^{\beta-1}}{(\beta n)^{\beta}}$$

$$\text{firm profit:} \quad q_n^* \times (p_n^* - c) - F = \frac{\alpha S}{nc^{\beta-1}} \frac{(\beta n - 1)^{\beta-1}}{(\beta n)^{\beta}} - F$$

We again examine the evolution of ETRs. We now have

$$g_n = \left(\frac{n}{n-1}\right)^{\beta} \left(\frac{\beta(n-1) - 1}{\beta n - 1}\right)^{\beta-1}. \tag{7}$$

Unlike in the linear demand case, the ETR now depends on a parameter of the demand function. In Fig. 2, we show the evolution of the ETRs for a range of demand curves that vary in terms of elasticity: lighter colors correspond to higher elasticities. The familiar hump-shaped pattern appears for each curve, but the value of n at which each curve reaches its maximum increases with the demand elasticity. The ETR declines especially rapidly for relatively less elastic demand curves.

Note that for $\beta \to 1$, industry profits converge to $\alpha S/n$. This expression is a function that declines convexly in n over its entire range. In that extreme case, the hump disappears entirely as the drop in industry profits is largest for the first entrant. For low values of β, for example for $\beta = 1.25$ or $\beta = 1.5$ in Fig. 2, there is still a hump, but it is situated entirely before the third firm enters. For integer numbers of firms, which are indicated by the solid markers, it holds that $g_2 > g_3$, and the hump is not empirically relevant. Only when the price elasticity of demand β is sufficiently large do we find again that $g_2 < g_3$ and $g_n > g_{n+1}$ for all $n \geq 3$, as in the linear demand case.[5] Thus, we obtain the following result.

Proposition 3. *Consider the canonical oligopoly model of competition in quantities with n firms, constant marginal costs, and constant elasticity demand. There is a threshold price elasticity* $\beta^* \approx 1.723$ *such that the following holds: If* $\beta > \beta^*$*, the entry threshold ratio* g_n *is hump-shaped in n and we have* $g_2 < g_3$*; if* $\beta \leq \beta^*$*, it is not hump-shaped for integer values of n.*

The intuition for the threshold price elasticity β^* is as follows. If demand is relatively inelastic, the profit maximizing strategy for a monopolist is to sell few units for a very high price. Note that for $\beta \to 1$ it holds that $Q_1^* \to 0$ and $p_1^* - c \to \infty$. With a second firm in the market, this strategy is no longer profitable. The total quantity produced will be too large

[5] Numerically, we can show that the critical threshold for β is approximately 1.723, in which case g_3 exactly equals g_2.

to charge very high prices. Conduct and industry profits change significantly when the second firm enters. In contrast, if demand is more elastic, the monopolist already produces a relatively large quantity. Note that $p_1^* - c \to 0$ for $\beta \to \infty$. Hence, the change in conduct and firm profits is less pronounced when the second firm enters.

As can be seen from Fig. 2, the mechanism that produces the hump-shaped result operates for all values of β. The negligible impact of an increase in n on industry profits at $n + \epsilon$ lowers the entry threshold at low n. Whether this effect extends to the second entrant and produces a hump-shaped pattern for discrete numbers of firms, depends on the rate of decline of industry variable profits, which in turn depends on the elasticity of demand.

For the most elastic demand curve shown in Fig. 2, the hump shape even extends to the fourth entrant. For $\beta = 8$, not only $g_2 < g_3$, but also $g_2 < g_4$. The relative change in market size needed to support four rather than three firms, is larger than the corresponding increase when the market goes from one to two firms. It is however the case that $g_3 > g_4$, such that the evolution of ETRs from $n = 3$ onwards straightforwardly maps into the change in competition.

3. Other modes of competition

So far, we have analyzed entry threshold ratios in Cournot oligopolies and found that they are hump-shaped. We now evaluate whether this is also the case for other forms of competition. We first consider price competition between firms offering differentiated goods and then we study a repeated Cournot oligopoly where firms can play strategies that lead to collusion. We show that in both environments ETRs can be hump-shaped in an intuitive manner. With only a few active firms, incumbents do not react to entry if the degree of product differentiation is sufficiently large or if the collusive arrangement can accommodate additional firms. However, as more firms enter, eventually local monopolies turn into competitive settings and collusive arrangements can no longer be sustained. These adjustments have been shown before, but we highlight the implications for successive ETRs.

3.1. Product differentiation

We first consider Salop's (1979) model of product differentiation. Consumers are located uniformly on a circle with a perimeter equal to 1. The mass of consumers is given by S. There are n firms that are located around the circle with equal distance between them. Each consumer wishes to buy at most one unit of the good. Her utility from the good equals v and she has to pay unit transport costs t. Thus, if a consumer trades with firm i at price p_i and has to travel distance x (on the circle) to firm i, her payoff equals $v - p_i - tx$. The value of her outside option is normalized to zero.

Firms only select prices $p_1, \ldots, p_n$. Firm i's profit from charging price p_i equals $(p_i - c)q_i - F$, where q_i is the demand for its good. To characterize the symmetric equilibrium with n firms, we have to distinguish between two cases: In the first case, transport costs t are small enough that firms are effectively competing against each other; the equilibrium quantity q_i then depends on p_i and the prices of firm i's neighbors. In the second case, transport costs t are so large that each firm enjoys a local monopoly; the equilibrium quantity q_i then only depends on p_i. We now characterize the equilibrium for both cases.

Assume first that in the symmetric equilibrium the market is covered and firms compete against each other. Suppose firm i charges price p_i, while all other firms charge p. The distance x of consumers who are indifferent between buying from firm i or from one of its neighbors is defined by the indifference condition

$$p_i + tx = p + t\left(\frac{1}{n} - x\right).$$

The demand for firm i's good then equals

$$q_i = S\left(\frac{p - p_i}{t} + \frac{1}{n}\right).$$

Maximizing firm i's profit with this demand generates the optimal price $p_i = (p + c + t/n)/2$. The unique symmetric equilibrium price is then given by $p_n^* = c + t/n$. In this equilibrium, each firm's profit equals $St/n^2 - F$. When the market is covered, the ETR equals

$$g_n = \frac{n}{n-1}. \tag{8}$$

Next, assume that in the symmetric equilibrium the market is not covered such that all firms enjoy a local monopoly. When firm i charges price p_i, the distance x of consumers who are indifferent between buying from firm i or choosing the outside option is defined by the indifference condition

$$v - p_i - tx = 0.$$

The demand for firm i's good then equals

$$q_i = S\frac{2(v - p_i)}{t}.$$

From profit maximization we find the unique equilibrium price $p_n^* = (v + c)/2$; each firm's profit in this equilibrium equals $S(v - c)^2/2t - F$. When the market is not covered, the ETR equals

$$g_n = \frac{n-1}{n}. \tag{9}$$

Note that this ratio is below one. The market size per firm that is needed for break-even shrinks rather than grows when the number of firms increases from $n - 1$ to n. The reason for this is mechanical. As long as the market is not covered, the n'th entrant sells to consumers that opted for the outside option before and were not served by any of the $n - 1$ incumbents. The same total market size, as measured by S, now supports one more firm, and $S/(n - 1) > S/n$. Schaumans and Verboven (2015) study a similar situation where the derivative of industry variable profits with respect to n has two terms: the standard effect that captures the adjustment in price and quantity to greater competition, and a second, market expansion effect. In our stark example, the first term is zero and the sign of $\Pi'(n)$ is determined by the second effect such that $\Pi'(n) > 0$ and $g_n < 1$.

Whether the market is covered or not depends on the transportation costs. Firms do not compete in equilibrium if consumers that are located exactly halfway in between firms, at a distance $x = 1/(2n)$, are indifferent between the monopoly price and the outside option. When there are n firms, this is the case if $v - (v + c)/2 - t/(2n) \leq 0$ or if

$$t \geq (v - c)n.$$

Define by $n^* \in \mathbb{N}$ the largest possible value n such that this inequality holds. If t is large enough, then n^* is well-defined. It holds then that the market is not covered in equilibrium if $n \leq n^*$, and it is covered if $n > n^*$. We calculate the entry threshold g_n for $n = n^* + 1$. Using the expressions for firm profits from the two cases above we get

$$g_{n^*+1} = \frac{n^*(n^*+1)}{2}\left(\frac{v-c}{t}\right)^2 \tag{10}$$

for the ETR that compares situations across both regimes.[6] With this, we obtain the following result.

Proposition 4. *Consider the circle model of product differentiation in this subsection. If transport costs t are sufficiently large, the (per-firm) entry threshold ratio g_n is hump-shaped in the following sense: There exists a threshold $n^* \in \mathbb{N}$ with $n^* \geq 3$ such that $g_{n-1} < g_n < 1$ for all $n \leq n^*$, $g_{n-1} > g_n > 1$ for all $n > n^* + 1$, and $g_n \to 1$ for $n \to \infty$.*

3.2. Collusion

Next, we consider a repeated Cournot oligopoly in which firms may be able to sustain collusion through repeated-game incentives. Let the stage-game be given by the Cournot model with general demand $SQ(p)$ that we studied in Section 2. We now consider the corresponding repeated game with observable actions and common discount factor δ. Firms collude – if possible – through the following implicit agreement: Each firm produces a $1/n$'th share of the monopoly quantity q_1^* unless in a previous period at least one firm produced a different quantity, in which case each firm chooses the quantity q_n^*; i.e., after any deviation firms' behavior reverses permanently to the symmetric equilibrium of the stage game.

Firms can maintain monopoly profits through this collusive agreement. We derive for which values of the discount factor δ and number of firms n unilateral deviations are not profitable. In this case the collusive agreement constitutes an equilibrium. When all firms follow they equilibrium strategy, they all make period profits equal to $S(q_1^*/n)(p(q_1^*) - c) - F$, while if play reverses to the symmetric stage game equilibrium, they only earn $Sq_n^*(p(nq_n^*) - c) - F$. Denote by $\tilde{q}$ a single-period optimal best-response deviation when all rivals produce q_1^*/n. Standard arguments show that the collusive strategy supports an equilibrium if and only if

$$\frac{1}{1-\delta}\frac{q_1^*}{n}(p(q_1^*) - c) \geq \tilde{q}\,(p(\tilde{q} + \tfrac{n-1}{n}q_1^*) - c) + \frac{\delta}{1-\delta}q_n^*(p(nq_n^*) - c). \tag{11}$$

For given n, this inequality is satisfied if δ is sufficiently large, and for given $\delta < 1$ it is violated if n is sufficiently large. Thus, if δ is sufficiently large, there is a threshold $n^* \in \mathbb{N}$ with $n^* \geq 2$ so that inequality (11) holds if the number of firms is weakly below n^*, but not if the number of firms exceeds n^*.

The implication of collusion for ETRs is as follows: If there are only a few active firms, $n < n^*$, the arrival of another firm does not change their conduct, and we have $g_n = 1$. However, if the number of firms exceeds the critical value n^*, collusion breaks down and the ETR equals $g_n = \Pi(1)/\Pi(n^* + 1)$ at $n = n^* + 1$ and $g_n = \Pi(n - 1)/\Pi(n)$ at $n > n^* + 1$. It then again exceeds 1. We therefore get a hump-shaped pattern for g_n that is based on the well-known effect that collusion collapses once there are too many firms enter the market (e.g., Selten, 1973). Selten (1973) We summarize this in the following result.

Proposition 5. *Consider the repeated quantity-setting game in this subsection. If the discount factor δ is sufficiently large, the (per-firm) entry threshold ratio g_n is hump-shaped in the following sense: There exists a threshold $n^* \in \mathbb{N}$ with $n^* \geq 2$ such that $g_n = 1$ for all $n \leq n^*$, $g_n > 1$ for all $n > n^*$, and $g_n \to 1$ for $n \to \infty$.*

[6] If we only consider discrete numbers of firms, $n \in \mathbb{R}$, the ETR in (10) applies only to $n = n^* + 1$. With continuous n, it applies to all $n \in (n^*, n^* + 1]$.

4. Implications for empirical work

We have shown that in a range of theoretical models entry threshold ratios display a hump-shaped pattern in the number of firms. We now consider what this means for their practical use. One solution to the ambiguous implication of an increase or the absence of change in the ETR could be to impose more structure on the reduced form model of profits. For example, Toivanen and Waterson (2005) illustrate how one can distinguish whether the presence of rival firms primarily means that competition is stronger or that the unobservable component in demand is higher. Note, however, that their extension relies on functional form and exclusion assumptions. Distinguishing the effect of entry on the price-cost margin and firm output would require even more assumptions, for example on the exact shape of demand. Such a solution seems unattractive as the strength of the framework lies exactly in the weak assumptions needed to calculate ETRs.

The main takeaway from our analysis is that care is warranted when ETRs are calculated in markets with few firms. In the case of a one-shot Cournot oligopoly, there is no problem when comparisons are limited to g_3 and up, i.e., comparing the s_3/s_2 ratio to the s_4/s_3 ratio, etc. For $n \geq 3$, a higher value for the ratio corresponds to a diminished strength of market competition (a higher price-cost margin), and vice versa. In contrast, the absence of a decline from g_2 to g_3 should not be interpreted as evidence that the third entrant did not strengthen competition or that the incumbents did not adjust their pricing behavior.

In some empirical applications the number of firms observed in the market never exceeds three, for example in Feinberg (2008), Manuszak (2002), or Pfann and Van Kranenburg (2003). In such a case it is still informative to know whether ETRs are estimated to be larger than unity or not, as such a value implies that an additional entrant strengthens competition. However, the absolute magnitude of g_3 or the difference between g_2 and g_3 is not a reliable gauge of how competition changed from the second to the third entrant. In particular, an increase in the ETR for the third entrant should not be interpreted as a less competitive situation.

In other applications, for example Dranove et al. (2003) or Cleeren et al. (2010), the observed firms can be classified into two distinct types—implicitly defining two market segments—and competition is conjectured to be stronger between firms of the same type. As each segment necessarily contains only a subset of the total number of firms, it heightens attention to the ETRs for the first few entrants. For example, Cleeren et al. (2010) calculate a separate ETR for within-type competition, holding the number of other type competitors constant. Their markets contain up to 7 retail stores, discounters or supermarkets, but they can only calculate the ratios up to $n = 5$, the maximum number of discounters in a market.

It would also be a mistake to consider the $n = 3$ threshold an absolute limit on the range of ambiguous ETRs. It is well known that the framework cannot identify the level of competition and cannot distinguish between a stable cartel and perfect competition. The product differentiation model and the repeated game illustrate that situations that are stable with few firms, might not be once there is enough entry. Market expansion and stable cartels are certainly more relevant when n is low, but there is no a priori limit. The results with constant elasticity demand also showed that the hump-shape will be more important if demand is less elastic and that it might extend to $n = 4$. Bresnahan and Reiss (1991) showed that the decline in ETRs depend on some of the maintained assumptions. For example, if marginal costs increase in quantity, entry generates an additional downward effect on the markup and industry profits decline more rapidly. Whether this eliminates the hump or not depends on the exact shape of the marginal cost function, but the range over which it is relevant diminishes.[7]

Like Nevo and Whinston (2010), our analysis highlights the difficulty of drawing inferences about the strength of competition from a reduced form analysis. Different models of competition can generate wildly different evolutions of the price-cost margin. In Table 2 we show this measure of competition for the three models that we studied, as well as the ETR and the corresponding ratio without normalizing the market size thresholds by the number of firms (S_n/S_{n-1}). In the Cournot oligopoly, the strength of competition rises monotonically with the number of firms, which is reflected in the declining ratio of price-cost margins.[8] The other two models display a similar evolution beyond the threshold $n^* + 1$, but the price-cost margins (and competition) does not respond to entry when the number of firms is small. Given that there is an intermediate region where the mode of competition changes and the comparison between n and $n - 1$ is a comparison across two competitive regimes, the price-cost ratio can even show a greater decline for later entrants.

It is difficult to capture these radically different evolutions with a single sufficient statistic. Not normalizing the market size thresholds by the number of firms appears to be a good idea in the first two models. The hump disappears in the Cournot case and the statistic mimics the invariance of the price-cost margin over the range where the (unobservable) market expansion supports the entry of differentiated firms. However, in the repeated Cournot game the ratio of aggregate entry thresholds erroneously interprets the larger market needed to support entry as a strengthening of competition, while this is merely driven by the need to recoup fixed cost. The price-cost margin remains constant before n^* and the usual ETR appropriately captures this. Given that in each case the change in price-cost margin lies in between the change in the per-firm and aggregate entry threshold ratios, calculating both would provide complementary information.

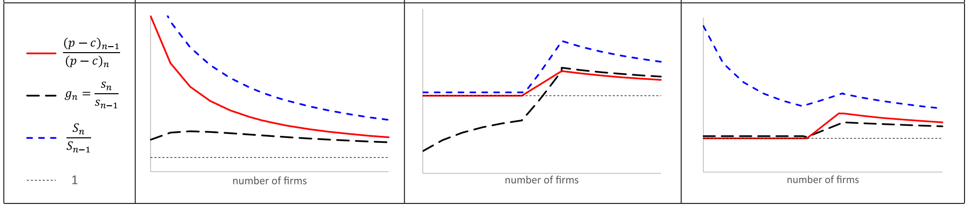
---

[7] If potential entrants differed in fixed costs and they entered in reverse order, i.e., the firm with lowest costs entered first, the change would intuitively be similar. Industry profits would decline more rapidly with entry, but the exact impact on the rate of change of ETRs would depend on the difference in fixed costs for successive entrants.

[8] Note that the relevant $p - c$ ratio divides the value for $n - 1$ by the corresponding margin for n, as was the case in the equality $g_n \equiv s_n/s_{n-1} = \Pi(n-1)/\Pi(n)$.

Table 2
Relationship between entry threshold ratios and changes in price-cost margin.

	Cournot oligopoly	Product differentiation		Repeated interaction	
		$n \le n^*$	$n > n^* + 1$	$n \le n^*$	$n > n^* + 1$
$\frac{(p-c)_{n-1}}{(p-c)_n}$	$\frac{n+1}{n}$	1	$\frac{n}{n-1}$	1	$\frac{n+1}{n}$
$g_n = \frac{s_n}{s_{n-1}}$	$\frac{(n-1)(n+1)^2}{n^3}$	$\frac{n-1}{n}$	$\frac{n}{n-1}$	1	$\frac{(n-1)(n+1)^2}{n^3}$
$\frac{S_n}{S_{n-1}}$	$\left(\frac{n+1}{n}\right)^2$	1	$\left(\frac{n}{n-1}\right)^2$	$\frac{n}{n-1}$	$\left(\frac{n+1}{n}\right)^2$



Note: The upward-sloping sections in the middle and right graphs run from n^* to $n^* + 1$.

5. Conclusion

We have shown that the entry threshold ratio, i.e., the increase in the minimum market size needed per firm to sustain one additional firm in the market, does not fall monotonically for additional entrants. Even though the first entrant in a monopoly market has the largest impact on the price-cost margin—which is one way to define the strength of the competitive effect of entry—it does not translate into the highest ETR. For a range of models and irrespective of the shape of demand, we find that the ETRs display a hump-shaped pattern: they first rise with the number of active firms, but after this initial increase they decline monotonically and converge to one.

Especially in the one-shot Cournot oligopoly, this finding is unexpected and interesting from a theoretical perspective. The intuition is that starting from the monopoly situation which maximizes industry profits, initial entry has only second order effects on aggregate profit. It makes entry of the second firm particularly easy, requiring an unusually small increase in the necessary market size for break-even. Only from the third entrant onwards does the evolution of ETRs correspond to the intuitive pattern, namely that each successive entrant has a gradually smaller effect in terms of strengthening competition. That market expansion and the collapse of a collusive arrangement can also increase the ETR is less surprising, but we highlight that both phenomena are most relevant when the number of in incumbents is small. Hence, they also lead to hump-shaped ETRs.

The finding is also relevant for applied work as it calls for caution when interpreting changes in ETRs. When comparing the increases in market size that is needed to support the second and third entrant, a small increase in the entry threshold cannot be interpreted as entrants having only a limited effect on competition or price-cost margins. Only from later entrants onwards (at least $n \ge 3$) should we unambiguously expect smaller ratios to indicate that the competitive effect of entry is diminishing.

References

Abraham, J.M., Drake, C., McCullough, J.S., Simon, K., 2017. What drives insurer participation and premiums in the federally-facilitated marketplace? Int. J Health Econ. Manag. 17, 395–412.
Abraham, J.M., Gaynor, M., Vogt, W.B., 2007. Entry and competition in local hospital markets. J. Ind. Econ. 55 (2), 265–288.
Asplund, M., Sandin, R., 1999. The number of firms and product capacity in relation to market size. J. Ind. Econ. 47 (1), 69–85.
Berry, S., Reiss, P., 2007. Empirical Models of Entry and Market Structure. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization, 3. North-Holland, pp. 1845–1886.
Bresnahan, T.F., Reiss, P.C., 1991. Entry and competition in concentrated markets. J. Polit. Econ. 99 (5), 977–1009.
Carree, M., Dejardin, M., 2007. Entry thresholds and actual entry and exit in local markets. Small Bus. Econ. 29, 203–2012.
Chevalier, J., Harrington, D. E., Scott Morton, F., 2009. Differentiated to death?Working Paper.
Cleeren, K., Verboven, F., Dekimpe, M.G., Gielens, K., 2010. Intra- and interformat competition among discounters and supermarkets. Market. Sci. 29 (3), 456–473.
Dranove, D., Gron, A., Mazzeo, M.J., 2003. Differentiation and competition in HMO markets. J. Ind. Econ. 51 (4), 433–454.
Feinberg, R.M., 2008. Explaining the credit union entry decision, and implications for performance. Rev. Ind. Organ. 33, 81–91.
Feinberg, R.M., Reynolds, K.M., 2010. An examination of entry and competitive performance in rural banking markets. South Econ J 76 (3), 624–637.
Gayle, P.G., Harrison, T.D., Thornton, J., 2017. Entry, donor market size, and competitive conduct among nonprofit firms. Int. J. Ind Organ. 50, 294–318.
Gil, R., Marion, J., 2018. Residential segregation, discrimination, and african-american theater entry during jim crow. J. Urban Econ. 108, 18–35.
Itin-Shwartz, B., 2017. Competition and market power in deregulated fertilizers in India. Working Paper.
Lee, G.S., 2007. Germany: Entry and Competition for Bavaria, Germany Notary Profession. In: Schmid, C.U., Sebastian, S., Lee, G.S., Fink, M., Paterson, I. (Eds.), Conveyancing Services Market. Zentrum für Europäische Rechtspolitik, Bremen.

Manuszak, M.D., 2002. Endogenous market structure and competition in the 19th century american brewing international journal of industrial organization industry. Int. J. Ind Organ. 20, 673–692.
Nevo, A., Whinston, M.D., 2010. Taking the dogma out of econometrics: structural modeling and credible inference. J. Econ. Perspect. 24 (2), 69–82.
Nishida, M., Gil, R., 2014. Regulation, enforcement, and entry: evidence from the spanish local TV industry. Int. J. Ind Organ. 32, 11–23.
Pfann, G.A., Van Kranenburg, H., 2003. Tax policy, location choices, and market structure. J. Law Econ. 46, 61–83.
Ruffin, R.J., 1971. Cournot oligopoly and competitive behaviour. Rev. Econ. Stud. 38 (4), 491–502.
Salop, S.C., 1979. Monopolistic competition with outside goods. Bell J. Econ. 10 (1), 141–156.
Schaumans, C., Verboven, F., 2015. Entry and competition in differentiated product markets. Rev. Econ. Stat. 97 (1), 195–209.
Selten, R., 1973. A simple model of imperfect competition, where four are few and six are many. Int. J. Game Theory 2 (1), 141–201.
Tirole, J., 1988. The Theory of Industrial Organization. MIT Press.
Toivanen, O., Waterson, M., 2005. Market structure and entry: where's the beef? RAND J. Econ. 36 (3), 680–699.
Varela, M.J., 2018. The costs of growth: accelerated growth and crowd-out in the mexican supermarket industry. Int. J. Ind Organ. 61, 1–52.
Xiao, M., Orazem, P.F., 2011. Does the fourth entrant make any difference? entry and competition in the early u.s. broadband market. Int. J. Ind Organ. 29, 547–561.