Free Exit and Social Inefficiency

Linus Wilson
Assistant Professor of Finance
University of Louisiana at Lafayette
B. I. Moody III College of Business
Department of Economics & Finance
214 Hebrard Boulevard, Moody Hall 253
P. O. Box 44570
Lafayette, LA 70504-4570
Phone: (337) 482-6209
Fax: (337) 482-6675
E-mail: linuswilson [at] louisiana [dot] edu
Web: http://www.linuswilson.com

October 2008

Journal of Economic Literature Codes: D43, D62, & L13

Keywords: entry, exit, market structure, oligopoly

Free Exit and Social Inefficiency

Abstract

Mankiw and Whinston (1986) shows that free entry is socially excessive entry when firms have fixed costs and produce identical goods. Here it is shown that weakly too few firms exit voluntarily when some of the fixed costs are recoverable.

1. Introduction

Since Maniw and Whinston (1986) its has been known that free entry often leads to socially wasteful investment in homogeneous goods industries. That study shows how, under reasonable conditions with the lost profits of rival firms, the business stealing externality exceeds the gains to consumer surplus under free entry when firms produce identical products. Whether or not free entry is socially efficient is ambiguous when firms have differentiated products as in Spence (1976a; 1976b). Measuring this tendency for excessive entry has been pursued in empirical work by Berry and Waldfogel (1999), radio; Hsei and Morretti (2003), real estate agents; Hortacsu and Syverson (2004), index funds; and Davis (2006), movie theaters.

The author knows of only of one other theoretical study that has identified the problem of insufficient exit. Amir and Lambson (2003) first identified this tendency in the homogenous goods case with discrete competitors. In contrast, this paper finds that there is insufficient exit when competitors are continuous. The present paper's structure has the benefit of being directly comparable to the calculus-based analysis in Mankiw and Whinston (1986).

This paper shows that welfare under free exit is strictly lower than the social optimum if the following conditions hold:

1. Markups over marginal cost are strictly positive.
2. The probability of the low demand state is non-zero.
3. The number of firms exiting under free exit is greater than zero.

When no firms voluntarily exit under free exit, then free exit is sometimes socially optimal. Yet, when some firms do exit in the low demand state, social welfare could be raised if more were pushed out of business.

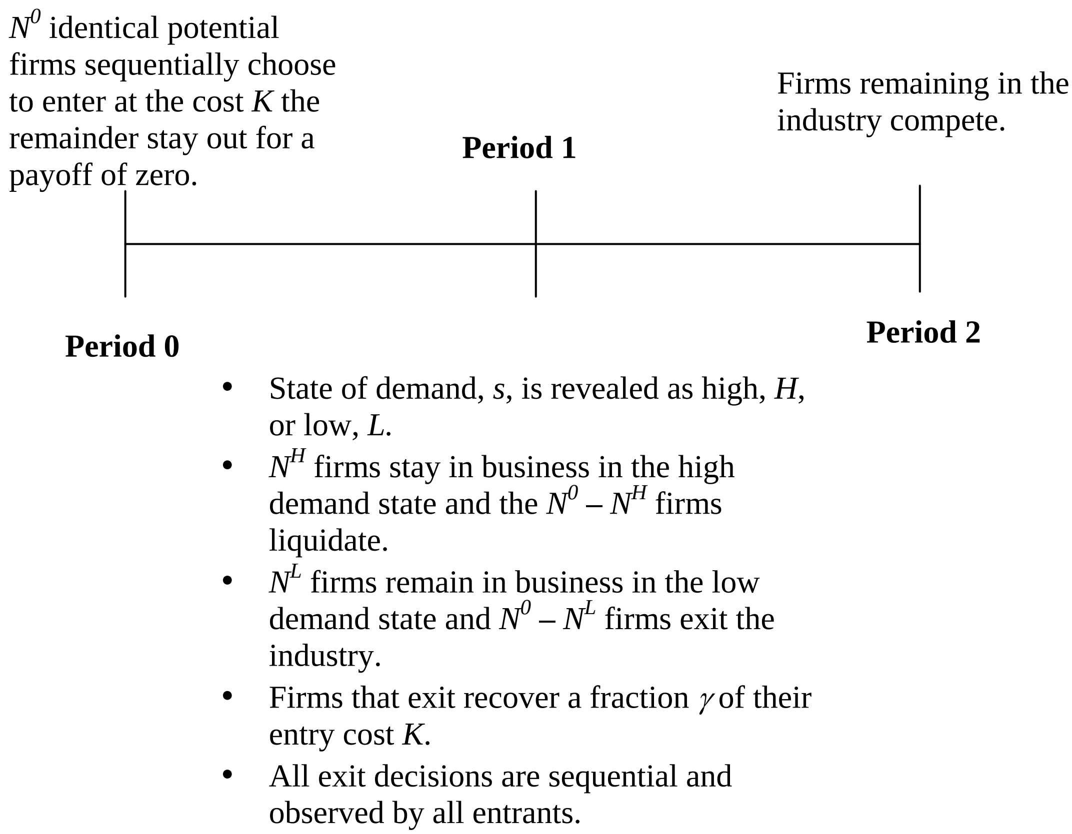

Figure 1

2. Model

Figure 1 above outlines the game. In period 0, N_0 identical firms sequentially choose to enter or stay out of the industry. Since firms are assumed to be identical, we do not index individual entrants. In period 1, firms sequentially choose to exit the industry or stay in to compete in period 2. All decisions are "now or never," therefore, no results rely on the real option value to waiting as described in Dixit (1989).

All firms have identical cost structures. All firms pay a cost of entry K. Some of this cost can be recovered upon exit $\gamma \in [0,1)$. Likewise, a fraction of this cost is unrecoverable, $1 - \gamma$, or sunk. Firms have variable cost functions $c(q^s)$, which does not include the entry cost, K. It is assumed that all firms weakly have diseconomies of scale after they have entered the industry. That is, $c(0) = 0, c'(q) \geq 0,$ and $c''(q) \geq 0.$

There are two states, high and low. That is, $s = H$ or L. The probability of the high demand state is h where $0 \leq h \leq 1$. In the high demand state, consumers are willing to pay a higher price for every quantity sold to the market. The inverse demand schedule given by $P(s, Q^s)$ is a function of the state, s, and of aggregate equilibrium output in state s, Q^s. The inverse demand is increasing in the state. That is, $P(H, \hat{Q}) > P(L, \hat{Q})$ for a given $\hat{Q}$. In addition, the market price is falling in aggregate output $\frac{\partial P}{\partial Q^s} < 0.$

Further, the number of firms in operation is given by the superscript s. That is N^s firms operate in period 2 in the state of demand s. By assumption, no firms are able to enter after the state is revealed because there are lags between the initial investment and a firm's ability to bring its product to market. In particular, no new firms can enter after period 0.

Aggregate industry output is just the outputs of the N^s identical firms producing an individual output $q^s \equiv q(s, N^s)$. That is, $Q^s \equiv N^s q^s$. The firms competing in period 2 produce the same output $q(s, N^s)$ in equilibrium. Further, it is robust to assume that an individual firm's output is increasing in the state. That is, $q(H, \hat{N}) > q(L, \hat{N})$ for any given $\hat{N}$ size of the industry.

The profits before entry costs for a single firm that is in operation in a given state is the following:

$$\pi(s, N^s) \equiv P(s, Q^s)q(s, N^s) - c(q^s) \tag{1}$$

We will assume that per firm producer surplus is increasing in the state. That is, the equilibrium output response leads to rising per firm producer surplus for a given industry size. That is, for any given industry size $\hat{N}$, $\pi(H, \hat{N}) > \pi(L, \hat{N})$.[1]

We can combine our entry assumption and our profit assumption to conclude the following:

Proposition 1

No firms exit in the high demand state. Further, the number of firms in the high demand state weakly exceeds the number of firms in the low demand state. That is, $N^0 = N^H \geq N^L$.

A proof is left for the appendix.

Following Mankiw and Whinston (1986) we will make 3 more assumptions. That paper proved that these fairly innocuous conditions will guarantee excess entry:

1. Total industry output is rising in the number of firms, $\frac{\partial Q^s}{\partial N^s} > 0$.

2. Individual firms' outputs are falling in the number of firms, $\frac{\partial q^s}{\partial N^s} < 0$.

3. All firms price at or above marginal cost, $P(s, Q^s) - c'(q^s) \geq 0$.

[1] The model is more general than Cournot competition. Yet, Cournot competition is a special case of the model. All the output, price, and per firm producer surplus assumptions are consistent with an industry composed of Cournot competitors facing a linear inverse demand curve.

These assumptions seem reasonable when firms weakly face diseconomies of scale and are producing homogenous goods.

3. Analysis

Our presentation here differs from Mankiw and Whinston (1986). We are also concerned with how the exit behavior affects the investment incentives of firms. Mankiw and Whinston (1986) only considers entry behavior because in that model there is only one state.

Social welfare function is the following:

$$
\begin{aligned}
W(N^H, N^L) &= h\left[\int_0^{N^H q^H} P(H,v)dv - N^H c(q^H)\right] \\
&+ (1-h)\left[\int_0^{N^L q^L} P(L,v)dv - N^L c(q^L) + \gamma \int_{N^L}^{N^H} Kdu\right] \\
&- \int_0^{N^H} Kdu, \\
&\text{where } q^H \equiv q(H, N^H),\ q^L \equiv q(L, N^L).
\end{aligned}
\tag{2}
$$

The top term is the expected total surplus in the high demand state. This is the price that consumers are willing to pay, less the total variable costs of producing the output. The second term is the expected total surplus generated in the low demand state plus the expected scrap value of the firms that are liquidated. Finally, the last term is the total entry costs for the industry.

Because welfare is a function two variables if there is an interior optimum, it will be a stationary point where the first derivative of welfare with respect to both the number of firms entering, N^H, and the number of firms remaining in the low state, N^L, are equal to zero.

Analysis of the optimal number of firms in the high demand state is analogous to the entry results derived in Mankiw and Whinston (1986). For this reason this analysis has been left to a supplemental appendix 6. The original contribution of this paper deals the first order condition with respect to the number of firms in the low demand state.

Let us differentiate welfare in equation (2) with respect to N^L.

$$\frac{dW}{dN^L} = (1-h)[\pi(L,N^L) - \gamma K] + (1-h)\frac{dq(L,N^L)}{dN^L} N^L [P(Q^L) - c'(q^L)], \tag{3}$$
where $Q^L \equiv N^L q^L$.

If firms are free to exit, they will do so when their revenues less variable costs are less than or equal to the recovery value of their fixed costs. Given that any firm exits, the following condition must be met:

$$\text{If } N^H > N^L, \text{ then } \pi(L,N^L) - \gamma K = 0. \tag{4}$$

This occurs in cases, when γ is large and some firms benefit from recovering a portion of their costs, γK. If there is a non-zero number of firms exiting, we can simplify the first-order condition. Let us define the free exit number of firms N^{FX} as the number of firms in the low demand state, N^L, when the condition in equation (4) is met. That is, by inserting equation (4) into equation (3) above, the FOC with respect to N^L reduces to the following:

$$\left.\frac{dW}{dN^L}\right|_{N^L=N^{FX}} = (1-h)\frac{dq}{dN} N^{FX}[P(Q^{FX}) - c'(q^{FX})] \leq 0. \tag{5}$$

We know this is weakly negative because per firm output falls in the number of firms, markups are weakly positive, and the probability of the low state occurring—1 – *h*—is non-negative. Therefore, equation (5) implies that weakly too few firms, $N^H - N^L$, exit in the low demand state, given that any firms voluntarily exit. When $h < 1$ and markups are strictly above marginal cost (that is, when $P(Q^{FX}) - c'(q^{FX}) > 0$) then welfare would strictly rise if more firms would exit. Amir and Lambson (2003) found that there are too few firms exiting when firms are discrete. In contrast, we have shown that this is the case when we relax the integer constraint and assume that firms are continuous. Thus, the results of this paper are more directly comparable to the results in Mankiw and Whinston (1986) which focuses primarily on the continuous case. Further, by only having one productive period, we have separated out the option value of waiting to identify the pure incentive towards insufficient exit in the continuous case.

Evaluated at the optimum, equation (3) reduces to the following expression:

$$\left.\frac{dW}{dN^L}\right|_{N^L=N^{L*}} = [\pi(L, N^{L*}) - \gamma K] + \frac{dq(L, N^{L*})}{dN} N^{L*}[P(Q^{L*}) - c'(q^{L*})] \equiv 0. \tag{6}$$

We know that the second term, $\frac{dq}{dN}N^{L*}[P(Q^{L*})-c'(q^{L*})],$ is negative when markups are positive. That is, positive markups imply that $\pi(L, N^{L*})-\gamma K>0,$ at the optimum. This can only be true if no firms choose to exit. That is, $N^H = N^L$. Therefore, an interior optimum implies that there is a tendency for too few firms to exit, given that any firms exit at all.

Further, the first-order condition in (6) does not depend on the number of firms entering, N^H. The only limit to this is that $N^L \leq N^H$. N^{L*} may sometimes exceed N^H, especially when recovery values are very low—γ is close to zero. In this case, if $\pi(L, N^H) \geq 0$, the N^H-th firm might do the socially optimal thing by not exiting in the low demand state. In this case, no firms will exit, and this will weakly coincide with the social optimum because technological constraints prevent further entry in period 1. This anomaly comes from the fact that welfare could be improved if firms could enter the industry at the liquidation value. Therefore, the constraint that N^L cannot exceed N^H must bind and $N^H = N^L$ when $N^{L*} > N^H$. Since zero firms optimally exit in this case, then we can say that sometimes exit is socially optimal. In particular, free exit is socially optimal when N^{FX} implied by the equality in (4) is greater than or equal to N^H. Yet, it still must be the case that $N^{L*} < N^{FX}$. Despite the technical constraints, we can make the following statement from our analysis above:

Proposition 2

- *There are weakly too few firms that exit in the low demand state.*
- *There are strictly too few firms exiting when $N^H > N^{L*}$.*

The proposition above follows from equation (5) and the preceding discussion.

This result should not be surprising in light of Mankiw and Whinston (1986)'s result. Suppose that we interpret the game in period 1 as an entry game. Exiting with partially sunk costs is very much like entering. A firm pays a fee γK to stay in the industry in period 1. Too many firms pay this fee. An equivalent interpretation is that too few firms exit. Therefore, the excessive entry result of Mankiw and Whinston (1986) also implies insufficient exit when demand is revealed. Yet, the author knows of no other paper, besides the present paper, that proves the pure insufficient exit incentive for the continuous case. Since exit is almost as important as entry in determining market structure, this omission in the literature is surprising.

4. Conclusion

This paper has shown that private incentives for firms to exit a homogenous goods industry when demand is low is weakly insufficient from the point of view of social welfare. When markups over marginal cost are positive, and the probability of a low demand realization is non-zero, and some firms exit voluntarily, then exit is strictly insufficient. This paper shows that business stealing externalities alone weakly lead to insufficient exit in homogenous goods industries with a continuum of competitors.

References

Amir, R., and Lambson, V.E., (2003). ‘Entry, exit, and imperfect competition in the long run’, *Journal of Economic Theory*, vol. 110, pp. 191-203.

Berry, S., and Waldfogel, J. (1999). ‘Free entry and social inefficiency in radio broadcasting’, *RAND Journal of Economics*, vol. 30, pp. 397-420.

Davis, P. (2006). ‘Measuring the business stealing, cannibalization and market expansion effects of entry in the U.S. motion picture exhibition market’, *Journal of Industrial Economics*, vol. 54, pp. 293-321.

Dixit, A., 1989, ‘Enty and exit decisions under uncertainty,’ *Journal of Political Economy*, vol. 97, pp. 620-638.

Hortacsu, A., and Syverson, C. (2004). ‘Product differentiation, search costs, and competition in the mutual fund industry: a case study of S&P 500 mutual funds’, *Quarterly Journal of Economics*, vol. 119, pp. 403-456.

Hsieh, C. and Moretti E., (2003). ‘Can free entry be inefficient? fixed commissions and social waste in the real estate industry’, *Journal of Political Economy*, vol. 111, pp. 1076-1121.

Mankiw, G., and Whinston, M. (1986). ‘Free entry and social inefficiency’, *Rand Journal of Economics*, vol. 17, pp. 48-58.

Spence, M. (1976a). ‘Product differentiation and welfare’, *American Economic Review Papers and Proceedings*, vol. 76, pp. 407-414.

Spence, M. (1976b). ‘Product selection, fixed costs, and monopolistic competition,’ *Review of Economic Studies*, vol. 43, pp. 217-235.

5. Appendix: Proof of Proposition 1

Proof of $N^0 = N^H \geq N^L$:

Let us prove this by contradiction. Suppose that $N^H < N^0$ and $N^L > N^0$.

It must be true that in the high demand state the marginal entrant, with the highest liquidation value of all firms, finds it optimal to liquidate itself. That is, $\pi(H, N^0) < \gamma I(N^0)$. Moreover, we assumed that $\pi(H, N^0) > \pi(L, N^0)$. Therefore, $\pi(L, N^0) < \pi(H, N^0) < \gamma I(N^0)$. The marginal entrant will be liquidated in both states. Expected returns are $(\gamma - 1)K < 0$ because $0 < \gamma < 1$. The marginal entrant must make non-negative profits or else it will not enter. Therefore, this is a contradiction. Further, we assumed that no entry will be permitted in period 1 after the state is revealed. Therefore, it is impossible that $N^L > N^0$. *Q.E.D.*

6. Supplemental Appendix: Excessive Entry with Two States

Let us differentiate this with respect to the initial number of firms in the market, N^H. After rearranging some terms, the first derivative can be expressed as the following:

$$\begin{aligned}\frac{dW}{dN^H} &= \left[h\pi(H, N^H) + (1-h)\max\{\gamma K, \pi(L, N^L)\} - K\right] \\ &+ h\,N^H \frac{dq(H, N^H)}{dN}(P(H, Q^H) - c'(q^H)),\end{aligned} \tag{7}$$

The terms in square brackets should be equal to zero under free entry. The marginal entrant's expected returns in both the high and low state minus its entry cost should just equal that firm's entry cost. If this were not the case, another firm would want to enter. Let us define the free entry equilibrium (FE) number of firms as

$$h\pi(H, N^{FE}) + (1-h)\max\{\gamma K, \pi(L, N^{FE})\} - K \equiv 0, \tag{8}$$

where the superscript "*FE*" signifies the initial number of firms entering the unregulated market.

The free entry (FE) condition in (8) implies that the first-order condition in equation (7) can be simplified considerably. The first-order condition must be weakly negative, and unregulated industry has excessive entry:

$$\left.\frac{dW}{dN^H}\right|_{N^H = N^{FE}} = h\, N^{FE}\, \frac{dq(H, N^{FE})}{dN^s} (P(H, Q^{FE}) - c'(q^{FE})) \leq 0 \tag{9}$$

We know that $dq/dN^s < 0$ from assumption 2. The term in brackets is the difference between price and marginal cost for an individual firm. Price must weakly exceed marginal cost if firms are profit maximizing. Therefore, the whole quantity must be weakly negative. If the first order condition is weakly negative, then this implies that entry is weakly excessive. When markups are strictly positive, free entry is strictly excessive.[2]

The social optimum is in part characterized by the following first order condition:

[2] Positive markups are a necessary condition for firms to make zero profits with positive entry costs.

$$\left.\frac{dW}{dN^H}\right|_{N^H=N^{H*}} = h\pi(H,N^{H*}) + (1-h)\max\{\gamma K, \pi(L,N^{H*})\} - K$$
$$+ h\,N^{H*}\frac{dq(H,N^{H*})}{dN^s}(P(H,Q^{H*}) - c'(q^{H*})) \equiv 0 \qquad (10)$$

When markups are positive, we know that the bottom term is negative. This implies that the top term on the left-hand side must be positive. This is impossible under free entry (FE). Firms cannot make strictly positive expected profits without encouraging more firms to enter. Since we assumed that entry was unregulated, (10) is not achievable. Later on we will consider if taxes and bankruptcy costs can move us closer to the optimum, N^{H*}. Yet, for now, this discussion leads us to the conclude the that free entry leads to excess entry when firms have positive markups, produce homogenous products, and must pay entry costs. This is a restatement of Mankiw and Whinston (1986)'s result.