Free Exit And Social Inefficiency

Linus Wilson, University of Louisiana at Lafayette, USA

ABSTRACT

Mankiw and Whinston (1986) show that free entry is socially excessive when firms have fixed costs and produce identical goods. That is because rival firms fail to externalize the business stealing costs they impose on their rivals. This paper extends that model by assuming that there are two states of demand. It is proven that weakly too few firms exit voluntarily when demand realizations are low and some of the fixed costs are recoverable. If there is any voluntary exit, social welfare could strictly rise by forcing more firms to exit the industry.

Keywords: entry, exit, market structure, oligopoly

INTRODUCTION

Since Maniw and Whinston (1986), it 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 Scherer (1979) with breakfast cereal varieties; Berry and Waldfogel (1999) with radio stations; Hsei and Morretti (2003) with real estate agents; Hortacsu and Syverson (2004) with index funds; and Davis (2006) with movie theaters. These studies broadly support the contention that there is evidence of socially excessive competition in many industries.

The author knows of only one other theoretical study that has identified the problem of insufficient exit due to the business stealing externality. 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).

Moreover, Amir and Lambson's (2003) results are not entirely driven by the business stealing externality. In most versions of the game of Amir and Lambson (2003), demand could improve because there are multiple productive periods. Thus, it pays for firms to stay in the industry even when they are unprofitable. Thus, the real option of waiting, as outlined in Dixit (1989), may be partially driving the insufficient exit result in Amir and Lambson (2003). In contrast in the present paper, because there are no future productive periods, the real option to wait has no value. Thus, the present paper separates the business stealing effect as driving insufficient exit, while Amir and Lambson (2003) combines both the business stealing and the option to wait to show that there are tendencies towards insufficient exit.

There is some similarity between this study and the studies of Ghemwhat and Nalebuff (1985; 1990). Those studies assume that competitors have identical cost functions for a given capacity. Yet, unlike this study, competitors have different productive capacities. Here we assume all competitors have identical cost functions and market share. Ghemwhat and Nalebuff (1985; 1990) find that larger competitors in a declining industry are the first to exit in the former study or are the first to shrink in the latter study. Thus, the focuses of Ghemwhat and Nalebuff (1985; 1990) are not on insufficient exit problems, but rather on the prediction that big versus small firms strategically exit first. Those papers do not directly address the business stealing externality identified by Mankiw and Whinston (1986). Moreover, unlike this paper, Ghemwhat and Nalebuff (1985; 1990) have a discrete number of competitors.

This paper shows that welfare under free exit is strictly lower than the social optimum when markups over marginal cost are strictly positive; the probability of the low demand state is non-zero,; and 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.

MODEL

Let N_0 be the initial number of firms in the industry in period 0. In period 0, the 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.

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 $\dfrac{\partial P}{\partial Q^s} < 0.$

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:

---

[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.

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, $\dfrac{\partial Q^s}{\partial N^s} > 0.$

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

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

These assumptions seem reasonable when firms face fixed costs and are producing homogenous goods.

ANALYSIS

Our presentation here differs from Mankiw and Whinston (1986). We are 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:

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

where $q^H \equiv q(H, N^H)$, $q^L \equiv q(L, N^L)$.

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 of 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 omitted. 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:

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

This occurs in cases when γ is large and when 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:

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

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.

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 $N^L \le N^H$. N^{L*} may sometimes exceed N^H, especially when recovery values are very low—γ is close to zero. In this case, if $\pi(L, N^H) \ge 0$, 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 free exit is socially optimal. 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

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

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

A numerical example may help illustrate proposition 2(b). Suppose that the industry faces a linear inverse demand curve of $P(L, Q^L) = \$(8 - Q^L)$ in the low demand state. Competitors have a constant marginal cost of \$2 per unit. They paid \$3 to enter and can recover half of that, \$1.50, if they exit. All competitors play a Cournot game, but we will allow fractional competitors to enter or exit. If the number of firms entering is greater than 3.90, then $N^{FX} = 3.90$ will be the free exit number of firms. Under free exit, social welfare is \$11.40, which is entirely composed of consumer surplus. (According to equation (4), profits net of opportunity costs are zero for all competitors under free exit.) The socially optimal number of competitors, indicated by the first order condition in equation (6), is $N^{L^*} = 1.88$ competitors. With only 1.88 competitors, consumer surplus would be \$7.68 and profits net of opportunity costs would be \$5.33. That would generate a total welfare of \$13.01, which is \$1.61 higher than welfare under free exit.

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, the probability of a low demand realization is non-zero and some firms exit voluntarily, then exit is strictly insufficient. This is the first paper to show that business stealing externalities alone weakly lead to insufficient exit in homogenous goods industries with a continuum of competitors.

AUTHOR INFORMATION

Linus Wilson is an assistant professor of finance at the University of Louisiana at Lafayette. He has written papers on bank bailouts, warrant pricing, the Madoff Ponzi scheme, CEO pay, capital structure in professional partnerships, industrial market structure, entrepreneurship, and mutual funds. Prior to coming to Lafayette, he has taught at the University of Cincinnati and the University of Oxford. His work on the banking sector rescue, the Troubled Asset Relief Program (TARP), has been cited in numerous news articles in media outlets such as *The New York Times*, *The Wall Street Journal*, and *The Financial Times*.

REFERENCES

1. 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.
2. Berry, S., and Waldfogel, J. (1999). 'Free entry and social inefficiency in radio broadcasting', *RAND Journal of Economics*, vol. 30, pp. 397-420.
3. 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.
4. Dixit, A., 1989, 'Enty and exit decisions under uncertainty,' *Journal of Political Economy*, vol. 97, pp. 620-638.
5. Ghemawat, P., and Nalebuff, B. (1985). 'Exit', *The Rand Journal of Economics*, vol. 16, pp. 184-194.
6. Ghemawat, P., and Nalebuff, B. (1990). 'The devolution of declining industries,' *The Quarterly Journal of Economics*, vol. 105, pp. 167-186.
7. 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.
8. 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.
9. Mankiw, G., and Whinston, M. (1986). 'Free entry and social inefficiency', *Rand Journal of Economics*, vol. 17, pp. 48-58.
10. Spence, M. (1976a). 'Product differentiation and welfare', *American Economic Review Papers and Proceedings*, vol. 76, pp. 407-414.
11. Scherer, F. M. (1979). 'The welfare economics of product variety: an application to the ready-to-eat cereals industry', *The Journal of Industrial Economics*, vol. 28, pp. 113-134.
12. Spence, M. (1976b). 'Product selection, fixed costs, and monopolistic competition,' *Review of Economic Studies*, vol. 43, pp. 217-235.

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 that the marginal entrant finds it optimal to liquidate itself. That is, $\pi(H, N^0) < \gamma K$. Moreover, we assumed that $\pi(H, N^0) > \pi(L, N^0)$. Therefore, $\pi(L, N^0) < \pi(H, N^0) < \gamma K$. 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.