Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 02-034

"Consumer Inertia, Firm Growth and Industry Dynamics"

by

Arthur Fishman and Rafael Rob

http://ssrn.com/abstract_id=335742

Consumer Inertia, Firm Growth and Industry Dynamics

Arthur Fishman
Bar Ilan University
Ramat Gan
ISRAEL
afishman@mail.biu.ac.il

Rafael Rob
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA. 19104-6297
USA
rrob@econ.sas.upenn.edu

January 10, 2002

Abstract

We develop a model of firm size, based on the hypothesis that consumers are "locked in," because of search costs, with firms they have patronized in the past. As a consequence, older firms have a larger clientele and are able to extract higher profits. The equilibrium of this model yields: (i) A downward sloping density of firm sizes. (ii) Older firms are less likely to exit than younger firms. (iii) Larger firms spend more on R&D. *Journal of Economic Literature* Classification numbers:

Key words: Consumer inertia, Firm Growth, Industry dynamics, R&D, Firm size distribution

Suggested running title: Consumer Inertia

1. Introduction

Firms, like people, are born small. Even a McDonald's or a Microsoft are very small at their inception and require the passage of time to grow and achieve their spectacular size. This paper develops an equilibrium model of firm growth, explores its aggregate implications and applies it to empirical and theoretical issues in industrial organization.

Our analysis emphasizes the role of intangibles such as "brand loyalty" or "consumer inertia." That is, we consider markets where consumers might not switch from one brand to another although the two are functionally identical and one is selling for a lower price. This brand loyalty gives firms market power over their repeat purchasers, and implies that a firm's market share determines its profit.

In our model, consumer inertia arises because consumers must incur search costs to learn about the prices of new sellers with whom they have not previously transacted. New consumers and firms continuously enter the market. A newly arrived consumer is randomly matched with a firm. Subsequently, the cost of searching for a new firm and the prices that firms charge (in equilibrium) lock the consumer in with her original firm. Thus established firms enjoy a proprietary relationship with repeat purchasers and compete with new entrants only for first-time buyers, who are as yet unattached to any firm. As it acquires successive generations of new consumers, a firm's stock of repeat-customers grows. Thus, the longer its tenure in the market, the greater is the firm's market share.

We explore the implications of this model and, especially, its ability to account for various facts about industry dynamics. An extensive empirical literature (e.g., Dunne, Roberts and Samuelson (1988, 1989) or Davis and Haltiwanger (1992)) finds the extent of firm turnover (or entry and exit) to be quite large even in mature and narrowly-defined industries. These studies suggest that firm turnover is *not* the consequence of events that are external to firms, for instance, business fluctuations, technological innovations, sectoral shifts, or changes in consumers' demands. Rather, they are triggered by internal, firm-specific changes; for example, losing a successful manager. Firm size and age have been identified among the characteristics most strongly associated with turnover: larger and older firms are less likely to exit than younger and smaller firms. Our model of firm growth accounts for these facts in a very natural way. We assume that individual firms' production costs are subject to idiosyncratic shocks. The effect of an idiosyncratic cost shock on a firm's exit decision depends negatively on its size. The reason for this is that the value of remaining operative in the wake of an adverse cost-shock is determined by the option value of a cost turnaround, which is higher for older firms (which have already accumulated a large customer base) than for younger firms (which have yet to do so). For the same reason, firms invest more in cost reduction measures the older (and larger) they are. The older is a firm the more it stand to gain if those measures are effective. Therefore, to an outside observer, the probability that a firm exits (known in the empirical literature as the "hazard rate") is decreasing in age. First, because an old firm invests more in R&D, it stands a greater chance of containing costs. Second, even if its cost goes up, the firm may still weather the increase because its large clientele base implies large profit upon a cost turnaround, which is not the case for a young firm.

The relationship of this paper to the industry dynamics literature is as follows. Jovanovic (1982) and Hopenhyn (1992) analyze models where the only characteristic of a firm is its cost of production. Here, by contrast, we introduce customer base, show how it interacts with costs, and show what effect they have on firms' decisions to enter and exit and spend money on R&D. Ericson and Pakes (1995) consider R&D, but, again, there is no such thing as customer base. We introduced customer base in two previous papers (Fishman and Rob (1995, 1999)), and analyzed their effect on pricing, but did not analyze entry and exit. Burdett and Coles (1997) consider the evolution of customer bases and its effect on pricing. However, they impose exogenous exit and do not consider R&D. Therefore, the contribution of this paper is to introduce customer base as a firm characteristic, and consider a fully dynamic model where firms enter, exit and do R&D, and where a distribution over firm sizes is endogenously determined.

2. Model

Time is discrete, indexed by $t = 1, 2, ..$. There is a continuum of firms producing an identical product, and a continuum of consumers who buy it. The population of consumers is subjected to turnover. A constant flow of measure ν enters each period, and a fraction δ of the existing consumers exits. The probability of exit is the same across consumers and is independent of age. We consider a steady state in which the stock of consumers is $\frac{\nu}{\delta}$.

2.1. Consumers

All consumers are identical and demand either zero or one unit at each period. A consumer's utility from one unit is $\overline{p}$. Thus $\overline{p}$ is the monopoly price.

Upon entering the market, a consumer is randomly and costlessly matched with a seller. Subsequently, she can costlessly return to the seller from which she bought in the previous period. Switching to a new seller, however, is costly. It is assumed that consumers know only the distribution of prices in the market, but not the prices of particular sellers. To learn the price of, and buy from, a new seller costs $\sigma > 0$. We call σ the **search cost.** At the beginning of each period, a consumer may sequentially sample the prices of an unlimited number of new sellers at the constant cost of σ per seller, before buying.

2.2. Firms

A firm bears four types of costs. First, to enter the market, the firm pays K, which is sunk subsequent to entry. Second, at the beginning of each period the firm pays a fixed cost F. This cost can be saved by exiting the market; re-entry, however, requires paying K once more. Third, the firm pays a constant per-unit cost, c. The per-unit cost c is determined stochastically, as described immediately below, and can assume one of m values: c_1, c_2,..., c_m, where $c_m > c_{m-1} > ... > c_1$, and $c_m > \overline{p} \geq c_{m-1}$. We refer to a firm with a current cost c_i as a **c_i-firm**. (Empirically, the unit cost is sometimes observed as Total Factor Productivity, firms with a high TFP being firms with low unit costs).

The fourth cost is R&D expenditures, aimed at containing or reducing per-unit costs. Assume a c_i-firm expends w dollars on R&D. Then - in the next period - it becomes a c_j-firm with probability $g_{ij}(w)$, i, $j = 1, ..., m$. We assume that $g_{mm} = 1$ and $0 < g_{ij} < 1$, for $i = 1, ..m - 1$, $j = 1, ..., m$, and all $w \geq 0$. That is, c_m is an "absorbing state," whereas all other c_i's are "transient states."

The assumption that $g_{mm} = 1$ and that $c_m > \overline{p}$ ensures that c_m-firms always exit and thus ensures (in a simple way) that continual entry and exit of firms persist in the steady state.[1] This assumption is motivated by the empirical findings that significant and continual entry and exit occur even in mature and narrowly defined industries. The assumption that $c_{m-1} \leq \overline{p}$ implies that firms make higher profit the more customers they have (although the profit may be negative because of the fixed cost, F). This assumption simplifies the analysis since it eliminates the possibility that a firm might want to turn customers away.[2]

Let G_i be the c.d.f. corresponding to g_i: $G_{i,k}(w) \equiv \sum_{j=1}^{k} g_{i,j}(w)$. We assume that $G_{i,j}$'s are twice continuously differentiable in w, and let $G'_{i,j}(w)$, $G''_{i,j}(w)$ denote their first and second derivatives. G is taken to satisfy the following assumptions:

Assumption A:

[1] In a previous paper we have analyzed the case where firms remain in the market forever; see Fishman and Rob (1995).

[2] Even if some c_i were $> \overline{p}$ a firm's value would still be increasing in its customer stock, which is the key to our results. This is because a firm has always the option of turning away customers, but does not have the option of attracting more customers than what the market dictates (see below).

1. $G'_{i,j}(w) > 0 > G''_{i,j}(w)$.
2. $G'_{i,j}(w) > G'_{i+1,j}(w)$.

Assumption A.1 is stochastic dominance and diminishing returns with respect to R&D expenditures: When a firm increases its R&D expenditures it stands a better chance of achieving lower cost in the next period. At the same time, R&D expenditures are subject to diminishing returns. Assumption A.2 implies stochastic dominance with respect to present-period cost: The lower is the present-period cost the better is the chance of achieving lower cost in the next period: $G_{i,j}(w) > G_{i+1,j}(w)$. However, A.2 is somewhat stronger because we assume the *derivative*, $G'_{i,j}(w)$, is larger than $G'_{i+1,j}(w)$ for every w. What is being assumed, therefore, is a form of complementarity between R&D expenditures and current cost: The lower is the current cost the bigger is the return on the marginal R&D dollar. The role of this complementarity will be seen below.

There is an infinite pool of potential entrants. We assume that entrants pay K and then draw their initial unit cost from a distribution $\alpha \equiv (\alpha_1, ..., \alpha_m)$, i.e., their initial unit cost is c_i with probability α_i.

At each date, a firm decides whether to exit or not. If it does not exit it has to choose R&D expenditures, and a price to charge for the present period. The objective of each firm is to maximize its discounted profit. The discount factor is denoted β.

Firms are distinguished by current marginal cost, c, and age, τ, the time elapsed since entry. A firm that just entered is considered an age-1 firm, in the next period it is age 2, etc. We call a firm with marginal cost c and age τ a **(c, τ)-firm** and refer to (c, τ) as a **firm type**. The age of a firm is one determinant of its payoff since, as will be shown below, firms accumulate customers only gradually, so the longer its tenure in the market the larger is a firm's clientele base and the larger is the revenue it can collect.

We denote the relative frequency of (c, τ)-firms by $f(c, \tau)$.

2.3. Equilibrium

We seek firm exit strategies, prices, R&D expenditures, consumer search strategies, flows of firm entry and exit, and a distribution over firm types which give rise to a **free-entry, steady-state equilibrium**. Formally:

(E.1) Each firm chooses a price, $p(c, \tau)$, and R&D expenditures, $w(c, \tau)$, to maximize its profit.

(E.2) Each consumer has a surplus-maximizing reservation-price strategy, $r(c, \tau)$, so that the consumer buys from a (c, τ)-firm whenever the price it quotes him is no higher than $r(c, \tau)$.

(E.3) There is a constant flow of entry, y, so that each entering firm makes zero expected profit.

(E.4) There is an exit strategy $\tau^*(c)$ so that all (c, τ)-firms with $\tau < (\geq)\ \tau^*(c)$ exit (stay, respectively), and this decision is optimal.[3]

(E.5) There is a distribution over firm types, $f(c, \tau)$, which is invariant under the entry rate in (E.3) and the exit strategy in (E.4).

3. Existence of Equilibrium

Since consumers have identical unit demand, have identical positive search costs, and search sequentially, it is well known (Diamond, 1971) that the equilibrium price of each firm is the monopoly price, $\overline{p}$, which consumers accept without search, i.e., $r(c, \tau) = \overline{p}$ for each type of firm.[4] Since switching to a new seller is costly and since each firm offers each consumer zero surplus, a consumer will never switch to a new seller. He remains with his first seller (as long as its price is not greater than $\overline{p}$) and exits the market if that firm exits.

Thus, as firms accumulate stocks of "locked in" customers, new entrants can only access newborn consumers. Let us fix the number of new consumers each firm receives in a period, and call it x. Let z be the number of customers a firm has, which is determined by the flow of consumers it gets per period and by its age, τ. Every period that the firm is in business, it gets a flow of x consumers but, at the same time, is losing a fraction, δ, of its existing customers. Hence, z grows as the sum of geometric series, the formula for it being equation (3.1). We refer to z as the firm's **customer stock** (or **customer base**). z is related to x and τ as follows:

$$z(\tau) = x + ... + x(1-\delta)^{\tau-1} = \frac{x[1-(1-\delta)^\tau]}{\delta}. \tag{3.1}$$

[3]As will be seen, a firm's (future discounted) profit is higher the older it is. Therefore the equilibirum exit rule is indeed given by a cutoff age.

[4]In Burdett and Coles (1997) small firms charge less than the monopoly price so as to attract more consumers. Their model, though, is predicated on "noisy search," whereby a consumer may get several price quotations at the same time. In our model, the consumer proceeds sequentially, receiving only one price quotation at a time. With sequential search a firm cannot attract more consumers by lowering its price, which is why all firms in our model charge the monopoly price.

Recall that (c, τ) is called firm type. With x fixed we can, equivalently, call (c, z) the firm's type. Accordingly, let $R_i(z; x)$ be a (c_i, z)-firm's maximum discounted-profit given a customer arrival-rate x. (c_i, z) are considered "state variables," while x is considered a "parameter." Written in dynamic programming format, the firm's maximization program reads as follows:

$$R_i(z;x) = Max\{0, \quad -F + z(\overline{p} - c_i)$$

$$+\beta \underset{w \geq 0}{Max}\{-w + \sum_{j=1}^{m} g_{ij}(w) R_j[x + z(1-\delta)]\}\}, i = 1, ..., m. \tag{3.2}$$

Equation (3.2) reflects two decisions the firm makes at the beginning of each period.

- *Exit Decision:* Consider a (c_i, z)-firm. The tradeoff it faces is the following. If it exits, the firm loses all its accumulated customers but saves the fixed cost, F, attaining a profit of zero. If it stays it pays F, which enables it to retain (and gradually increase) its customer base, make variable profit on its current customers, and face the possibility of lower costs (and, hence, higher profits) in the future. The firm chooses the larger of the two. Since c_m is an absorbing state, $R_m = 0$ and a c_m-firm immediately exits.

- *R&D Decision* If the firm stays, it chooses optimal R&D expenditures to maximize its value net of the expenditures, w.

By dynamic programming techniques (Stokey, Lucas and Prescott (1993), ch. 4):

(i) There exists a unique solution to equations (3.2).
(ii) The solution is continuous and increasing in z and x.
(iii) $R_{i-1}(z;x) > R_i(z;x)$ (this follows from the stochastic dominance of $G_{i,\bullet}$ over $G_{i-1,\bullet}$).

By (E.3) the expected profit of entrants is zero. Since entrants receive x customers in their initial period this implies that the equilibrium x - call it x^* - satisfies:

$$\sum_{i=1}^{m} \alpha_i R_i(x^*; x^*) = K. \tag{3.3}$$

Given this x^*, since R_i is increasing in z, there exists a z_i^* so that:

$$R_i(z_i^*; x^*) = 0 \text{ (if } R_i(0; x^*) > 0 \text{ we let } z_i^* = 0). \tag{3.4}$$

z_i^* is the **critical size** so that firms with cost c_i remain operative if and only if they have accumulated at least z_i^* customers. Since R_i is decreasing in i, $z_i^* > z_{i-1}^*$, that is, a high-cost firm requires a larger customer base to remain operative than a low-cost firm.

Equivalently, we can relate the exit decision of a firm to its age. Let t_i^* be the **critical age** so that firms of age t_i^* will have accumulated z_i^* customers. t_i^* is the smallest integer which satisfies:

$$\frac{x^*}{\delta}[1-(1-\delta)^t] \geq z_i^* \text{ (if } z_i^* > x^*/\delta, \text{ set } t_i^* = \infty) \,. \tag{3.5}$$

A firm exits if and only if it turns c_i - or higher - less than t_i^* periods after entering the industry (i.e., $\tau^*(c_i) = t_i^*$). Again, by monotonicity, $t_i^* > t_{i-1}^*$.

We have now shown that the per-firm flow of arriving consumers, x^*, and the exit strategy, (t_i^*), can be chosen so that (E.1) (E.2) and (E.4) are satisfied. It remains to show that a flow of entry y can be found which is consistent with x^* and (t_i^*), to compute a corresponding equilibrium distribution, f, and to show that y and f satisfy conditions (E.3) and (E.5). The next Proposition, proven in the Appendix, states that this can indeed be done.

Proposition 3.1. *An equilibrium exists.*

Proof. The strategy of the proof is to fix a y, and compute the steady-state distribution of firm types corresponding to it. In doing so, we account for firms' optimal exit and R&D policies. Once we express this distribution as a function of y we are able to choose a y is so that new entrants expect to make zero discounted profits. The details of this computation are worked out in the appendix. ■

This shows the existence of an equilibrium. We turn now to study its properties and relate them to empirical literature.

4. Properties of the Equilibrium and their Relation to Empirical and Theoretical Literature

4.1. Firm Size and the Probability of Exit

Empirical studies of industry dynamics find that exit probabilities are decreasing in both size and age (see Dunne, Roberts and Samuelson (1988, 1989) or Davis and Haltiwanger (1992)). Our model is consistent with this finding. On the basis of the preceding analysis, we may distinguish three possible cases. At one extreme, $z^*_{m-1} \leq x$. Then, only c_m-firms ever exit. At the other extreme, $z^*_2 \geq x^*/\delta$ ($z^*_1 \leq x$ or else entry would be unprofitable). Then all but c_1-firms exit, irrespective of age. In both these cases, the exit probability is independent of size (equivalently, age). When $z^*_{m-1} \leq x$, the exit probability of any firm is g_{im}, regardless of its size. When $z^*_2 \geq x^*/\delta$, all but c_1-firms exit, regardless of size, while c_1-firms exit only if they turn c_m.

The third, and most interesting, possibility is that for some c_j, $x^*/\delta > z^*_j > x^*$. In this case, a firm's exit probability depends on its size. Specifically, let c_h and c_k, $m \geq k > h > 1$, be the smallest and the largest j (respectively) for which $x^*/\delta > z^*_j > x^*$ holds. Then we need only consider firms with cost $\leq c_k$, since firms with cost $> c_k$ are not in the market. Consider a firm with cost $c_j < c_h$. If this firm is of size less than z^*_h, its probability of exit is $g_{j,h} + g_{j,h+1...}$, while if its size is greater than z^*_h, the corresponding probability is at most $g_{j,h+1} + g_{j,h+2...}$. For firms with cost $c_j, c_h \leq c_j < c_k$, of size z, $z^*_j < z < z^*_{j+1}$, the exit probability is $g_{j,j+1} + g_{j,j+2} + ...$, while for z between z^*_{j+1} and z^*_{j+2}, the exit probability is $g_{j,j+2} + g_{j,j+3} + ...$, and so on. Thus, for all firms whose cost is less than c_k, the probability of exit is decreasing in size. Finally, for c_k-firms, the probability of exit is $g_{k,k+1} + g_{k,k+2} + ...$, independent of size.

Thus, on average, considering all types of firms, the exit probability is decreasing in size, which is in accordance with the empirical literature cited above.

This property results from the fact that a firm's value increases with age. The reason for this is twofold. First, the current profit increases in age, since, in our model, the passage of time increases the firm's customer stock and hence its current profit; the unit profit applies to a larger volume of sales.

Second, the continuation value is increasing in age. This is for two reasons. First, a firm with a larger stock of customers gets a bigger boost from turning low-cost in the future. This is seen most clearly in the case of a firm with $c_i = \overline{p}$. Such a firm's operating profit is $-F$, regardless of size. However, the benefit from turning low-cost is increasing in size, because a larger firm materializes a bigger profit from achieving a cost turnaround. Therefore, in this case, the cost of staying in business is the same (F), but the benefit is increasing in size. This implies that larger firms have a stronger incentive to weather adverse cost shocks, i.e., they are less likely to exit.

The second reason that the continuation profit is increasing in size is that, as is shown below, firms' investment in cost-reducing innovations (R&D expenditures,

w) increase in firm size. Thus a larger firm not only is better able to weather adverse shocks but has a higher probability of achieving low future costs, which further increases its option value. Moreover, as is shown below, R&D expenditures are bigger for firms with lower per-unit costs. This results from assumption (A.2) that current costs and R&D expenditures *complement* each other: If a firm already has low unit cost it gets a bigger return for its marginal R&D expenditures. This further reinforces the value-enhancing role of size; larger size leads to greater investment, which leads to lower future costs, which will lead to even larger future investments.

The practical implication of all the above is that there is a polarizing tendency in the market: low costs and/or a large customer bases boost firm R&D expenditures which tends to further lower costs and to enhance the probability of survival. Hence, in our model, success breeds more success.

Relation to literature: It is instructive to compare this reasoning with the one in Jovanovic (1982) and Hopenhyn (1992). In those models, the exit probability is decreasing in size because large firms have lower marginal-cost than small firms and, hence, are less likely to exit. Here, as discussed below, large firms do not necessarily have lower marginal cost than small ones. In fact - as we point out below - they may have *higher* marginal cost. However, their competitive advantage is in having accumulated a large customer stock, which is a time-consuming process, and which they are reluctant to forego by exiting the market. Put differently, in Jovanovic (1982) and Hopenhyn (1992) firm size is not a characteristic of a firm, merely a reflection of its cost, whereas here firms are distinguished by cost *and* size.

4.2. Firm Size and R&D expenditures.

The next Proposition shows that a firm with a large customer base or a firm with low cost will do more R&D.

Proposition 4.1. *(i) R&D investments - the RHS maximizer of (3.2) - increase in z. (ii) R&D investments decrease in i.*

Proof. Holding x constant, consider the RHS of (3.2) as an operator on the space of continuous bounded functions in z. Assume further that the function we insert into the RHS of (3.2) is such that $R_i(z;x) - R_{i+1}(z;x)$ is increasing in z, and call this property (P). We prove that the function we obtain on the LHS satisfies

(P). Therefore, if we keep iterating this operator, the function we get in the limit, which is the firm's value function, R, also satisfies (P).
Consider some initial R^0 satisfying (P). Plug it into the RHS of (3.2) and call the resulting LHS R^1:

$$R_i^1(z;x) \equiv Max\{0, \quad -F + z(\overline{p} - c_i)$$

$$+\beta \underset{w\geq 0}{Max}\{-w + \sum_{j=1}^{m} g_{ij}(w) R_j^0[x + z(1-\delta)]\}\}$$

$$= \quad Max\{0, \quad -F + z(\overline{p} - c_i) + \beta \underset{w\geq 0}{Max}\{-w + \sum_{j=1}^{m} G_{ij}(w)\{R_j^0[x + z(1-\delta)] - R_{j+1}^0[x + z(1-\delta)]\}\},$$

using summation by parts. Since G is strictly concave this implies there is a unique RHS maximizer, call it $w_i(z)$, which can be found from the first order condition. Since $R_j^0[x + z(1-\delta)] - R_{j+1}^0[x + z(1-\delta)]$ is increasing in z and since $G' > 0$ (assumption A.1), $w_i(z)$ is increasing in z. Likewise, by assumption A.2, $w_i(z)$ is decreasing in i.

Consider now $R_i^1(z;r,x) - R_{i+1}^1(z;r,x)$. To show it is increasing in z we need to consider three cases: (a) where the RHS maxima for i and $i+1$ are 0, (b) where the RHS maximum for $i+1$ is 0, but the RHS for i is positive; and (c) where both RHS maxima are positive. Let's consider the third case since the other two are easier. Note first that $z(\overline{p} - c_i) - z(\overline{p} - c_{i+1})$ is increasing in z. Therefore it suffices to prove that the difference between the maximized values is increasing in z. By the envelope theorem, the z derivative of this difference is $\sum_{j=1}^{m}[G_{i,j}(w_i(z)) - G_{i+1,j}(w_{i+1}(z))]\{R_j[x+z(1-\delta)] - R_{j+1}[x+z(1-\delta)]\}$. But this is positive since $w_i(z) > w_{i+1}(z)$ (which we just proved), since $G_{i,j}(w) > G_{i+1,j}(w)$ (which A.2 implies) and since G is increasing in w. Thus, $R_i^1(z;r,x) - R_{i+1}^1(z;r,x)$ is indeed increasing in z.
While proving that the operator on the RHS of (3.2) preserves property (P), we have also shown that (i) and (ii) hold for any R satisfying property (P). Therefore, since R satisfies (P) it also satisfies (i) and (ii). ■

The proof relies on assumption A.2, which says that R&D is such that if a firm starts out with a low-cost it is more likely to remain low-cost than a firm that starts out with a high-cost. If R&D means reverse engineering, this assumption need not apply. Indeed with reverse engineering "all" that a firm needs to do is copy a technologically advanced firm. In that sense, a high cost firm obtains a higher return on its R&D investment. Thus, Proposition 4.1 applies when cost-reduction has to be done in-house and is not easily imitable.

Relation to literature: Larger firms invest more in R&D for the same reason that the exit probability is increasing with size: The larger the current market share, the greater the future market share to which the cost saving is expected to apply and, hence, the higher the return on the investment. This reasoning is familiar from the literature.[5] For instance, this argument is made in Arrow (1962) and analytical examples are worked out in Wilson (1975) in the context of a *single firm*. For that reasoning to apply in a multi-firm context, however, market share must be a distinct characteristic of the firm, and that is *not* the case for competitive or oligopolistic models. In those models, since a firm can access as many customers as it wants to, costs and market shares are perfectly correlated and, thus, there is no natural relationship between market share and the incentive to invest in cost reduction: A currently small, high-cost firm can achieve as large a profit by becoming a low-cost firm as that of a currently large, low-cost firm. Therefore, a small firm has no less of an incentive to invest in lowering costs. By contrast, in our model, market share *is* a characteristic which evolves independently of cost and, thus, is an independent variable which affects the incentive to innovate.

4.3. The Size Distribution of Firms

An immediate consequence of the preceding analysis is that in our model, the density of the size distribution of firms is decreasing, i.e., the proportion of firms of size equal to z is decreasing in z. This follows directly from the fact that for a firm to reach size z_2^* it must be uninterruptedly c_1 for a period of t_2^*. In order to reach size z, $z_2^* < z < z_3^*$, it must be uninterruptedly c_1 for a period of t_2^* *and* be either c_1 or c_2, for an additional period of at least $t_3^* - t_2^*$, i.e., it must enjoy a longer run without adverse cost shocks. More generally, the bigger is the size the longer must the firm have escaped adverse cost shocks, which occurs with a smaller probability. Consequently, the proportion of firms of a given size is decreasing in size.

[5] Proposition 4.1 is also consistent with a large empirical literature, surveyed by Cohen and Lewis (1989), which suggests that absolute R&D expenditures are positively correlated with size.

Relation to literature: There is an old literature, originating with H. Simon's work (see Ijiri and Simon (1977)), which tries to find the best fit between a statistical distribution and the actual size distribution of firms. The "theory" in this literature is to assume some law of firm growth, usually Gibrat's law of proportional growth (see next section), and then crank out the steady-state which corresponds to it. However, the growth law itself, which is the key element in such an exercise, is exogenously specified, i.e., it is not related to consumers' maximizing behavior. In contrast, this paper models explicitly consumers' search problem, and from it derives how firms grow over time.[6]

4.4. Are small firms leaner and meaner?

A popular conception is that small firms are more cost efficient than large ones. For example, recently there have been numerous articles in the popular press as well as policy proposals, suggesting subsidies to small firms as means of generating new jobs. One rationale that these articles suggest is that small firms are more "nimble" and "dynamic", and that they are bound to become the large corporation of tomorrow. The preceding discussion implies one sense in which our model is consistent with such a view. In equilibrium, the marginal cost of all firms of size less than z_j^* is less than c_j (since firms with cost greater than c_j of this size exit), while the marginal cost of some firms of size bigger than z_j^* exceeds c_j. In this sense, smaller firms have lower marginal costs in our model. (On the other hand, in the class of firms of size greater than z_j^*, marginal costs are more likely to decrease over time because these firms invest more in cost reduction measures. So the net relationship between size and costs is indeterminate in our model.)

This flexibility of the model contrasts the models of Jovanovic (1982), Lambson (1992) or Hopenhyn (1992) in which large firms unambiguously have lower marginal costs than small ones; in fact, the marginal cost in those models is perfectly (and negatively) correlated with size.

4.5. Technological Progress, Entry and the Stock Market Value of Firms

Consider technological progress, which raises firms' TFP, or equivalently, lowers their unit costs. For example, advances in computer technology are said to have had this effect. To fix ideas let us assume that the initial distribution α as well as

[6] Apart from that literature, none of the "maximizing" models of industry dynamics generates concrete predictions about the size distribution of firms.

the intertemporal R&D function g_{ij} shift to the left, resulting in lower costs (on average). What effect might this have on entry and the stock market valuation of existing firm?

The initial impact of the cost reduction is to increase firms' profits and values. This increases the rate of entry, and reduces the flow of new customers that each firm gets. The stock market valuation of new firms remains unchanged; they earn just enough to cover the cost of entry, K. On the other hand, the stock market valuation of old firms goes up. The logic is as follows. Suppose, for example, that the average unit cost faced by older firms is the same as the average unit cost that new entrants face (this is the case if the stochastic process governing unit costs is a martingale). An old firm derives value from its locked in customers and from the future arrival of new customers. With respect to the latter, an old firm is on the same footing as a new entrant: Since, in equilibrium, the value of this flow equals K to new entrants, both before the cost-reducing shock and after it, it is also worth K to old firms. On the other hand, the cost reduction increases the value of the old customer stock. So the valuation of old firms must increase. These two predictions of the model - increased entry and increased stock market valuation - are consistent with the stock market behavior in the U.S. during the late 1980's and the 1990's. By contrast, in more conventional models, cost savings of all firms, old or new, must be competed away by increased entry.

5. Model Extensions

The theme we focus on - customer loyalty - can be combined with other prominent themes that are used in the industrial economics literature. This would generate richer and more realistic models. Rather than develop a full-blown model for each theme, we suggest in the following subsections how these themes might bear on customer loyalty.

5.1. The Rate of Growth and Gibrat's Law

Our assumption that firms grow by a fixed number of consumers each period implies that the growth rate is inversely related to size. This is at variance with Gibrat's law, according to which the growth rate is independent of size.[7] Our

[7] It is worth stressing, however, that Gibrat's "Law" is a conveniet assumption, and not an empirical law. Several studies have tried to verify the empirical validity of this law, resulting in mixed results; see Ijiri and Simon (1977). One study which accords with the assumption of this

model could equally - and perhaps even more plausibly - be reformulated to accommodate Gibrat's law by assuming that the number of new consumers a firm attracts is proportional to its size, say because a first-time buyer is more likely to hear of a large firm than a small one. For example, if newborns locate a firm by asking around, then large firms capture more newborns, due to the fact that large firms are being "advertised" by more people. Such a formulation would *reinforce* the advantage that large firms have over small ones and, hence, would retain the property that large firms are less likely to exit.

5.2. Learning by doing

Another reason that larger firms may have lower costs (in addition to the reason that larger firms invest more in cost reducing measures) is because of learning by doing and on the job experience. Incorporating this into our model would reinforce our preceding analysis. Namely, older firms have a higher survival probability because they have moved further down the learning curve.

5.3. Takeovers and Mergers

We have assumed that bankrupted firms - firms with a small clientele base and high costs - exit the market. Since in our framework, a firm's customer stock is a valuable asset, a more interesting and realistic assumption is that such firms are acquired by low cost firms (assuming that the latter are able to transfer their technology to the acquired firm). This is an intriguing topic left for future research.

5.4. More Active Competition

In our formulation, all consumers have the same search cost, so there is no active search in equilibrium. As a result, competition between firms is somewhat indirect, since consumers patronize all firms equally. An alternative formulation is that consumers have different search costs. Then, in equilibrium, firms with sufficiently high costs would "specialize" in high search-cost consumers, charging prices which consumers with low search-costs reject to search for low-cost/low-price firms, while low-cost firms would sell to all types of consumers. Consumers would then search in equilibrium, and competition between low- and high-cost firms would be more direct, with low-cost firms charging low prices to attract low search-cost consumers away from high-price firms.

---

paper, that the rate of growth is negtively related to size, is Vining (1976).

Another feature under search-cost heterogeneity is that the correlation between age and size is less than perfect. For instance, an old firm with a run of high-cost shocks might have less customers than a young firm with a run of low-cost shocks. So in that instance the customer base of a firm (and, hence, its profit) is history-dependent.

5.5. Shakeout

Our model focuses on the steady state of an industry, intending to capture facts about mature industries. In addition to such facts, there are facts on the *growth-patterns* of industries, in particular the fact that at some point in the evolution of an industry, massive exit of firms takes place to be followed by much more stable pattern of entry and exit. A good documentation of this phenomenon can be found in Gort and Klepper (1982). A variant of the model presented here can be used to capture this phenomenon. Assume that a new market opens (as a result of innovation, say) and that new consumers start to arrive in waves. This raises the clientele base of existing firms and brings about the entry of new firms. Now if the arrival of consumers stops or slows down (which, it must, at some point), some firms - namely, the youngest - will find themselves with too small a clientele base (and with limited future opportunities to increase it) to continue operating and, as a result, will exit. A similar thing will occur if there are learning-by-doing effects at the industry level (i.e., all firms are learning independent of their age/experience), which at some point come to an end. Then, again, the firms that will be adversely affected by this are the youngest, and a whole bunch of them will exit. Therefore, the prediction of such model is “last in first out,” which is what the data show (see Gort and Klepper (1982)).

References

[1] Arrow, K. J., Economic Welfare and the Allocation of Resources for Invention, Reprinted in Collected Papers of K. J. Arrow, Volume 5, 104-119. Cambridge: Harvard University Press.

[2] Burdett, K., and M. G. Coles, Steady State Price Distributions in a Noisy Search Equilibrium, *Journal of Economic Theory* 72 (1997), 1-32.

[3] Cohen, W. M. and R. C. Lewis, Empirical Studies of Innovation and Market Structure, in Handbook of Industrial Organization, Schmalensee, R. and R. D. Willig (eds.), Amsterdam: North Holland (1989).

[4] Davis, S. and J. Haltiwanger, Gross Job Destruction and Employment Reallocation, *Quarterly Journal of Economics* 107 (1992), 819-863.

[5] Diamond, P., A Model of Price Adjustment, *Journal of Economic Theory* 3 (1971), 156-168.

[6] Dunne, T., M. Roberts and L. Samuelson, Patterns of Firm Entry and Exit in US manufacturing Industries, *Rand Journal of Economics* 19 (1988), 495-515.

[7] Dunne, T., M. Roberts and L. Samuelson, The growth and failure of US manufacturing plants, *Quarterly Journal of Economics* 104 (1989), 671-698.

[8] Ericson, R. and A. Pakes, Markov Perfect Industry Dynamics: A Framework for Empirical Work, *Review of Economic Studies* 62 (1995), 53-82.

[9] Fishman, A. and R. Rob, The Durability of Information, Market Efficiency and the size of Firms, *International Economic Review* 36 (1995), 19-36.

[10] Fishman, A. and R. Rob, The Size of Firms and R&D Investment, *International Economic Review* 40 (1999), 915-931.

[11] Gort, M. and S. Klepper, Time Paths in the Diffusion of Product Innovations, *Economic Journal* 92 (1982), 630-653.

[12] Hopenhayn, H., Entry, Exit and Firm Dynamics in Long Run Equilibrium, *Econometrica* 60 (1992), 1127-1150.

[13] Ijiri, Y. and H. A. Simon, Skew Distributions and the Size of Business Firms, Amsterdam: North Holland Publishing Company (1977).

[14] Jovanovic, B., Selection and the Evolution of Industry, *Econometrica* 50 (1982), 649-670.

[15] Lambson, V., Competitive Profits in the Long Run, *Review of Economic Studies* 59 (1992), 125-142.

[16] Stokey, N., R. Lucas and E. Prescott, Recursive Methods in Economic Dynamic, Cambridge: Harvard University Press (1989).

[17] Vining, D. R., Autocorrelated Growth Rates and the Pareto Law: Further Analysis, *Journal of Political Economy* 84 (1976), 369-380.

[18] Wilson, R., Informational Economies of Scale, *The Bell Journal of Economics* 6 (1975), 184-195.

A. Proof of Proposition 3.1

As the text shows the construction of equilibrium can be separated into 2 steps. First, we determine the flow of customers each firm receives, call it x^*, and the exit rules that firms follow, (t_i^*), so that conditions (E.1), (E.2) and (E.4) are satisfied. Second, one has to determine the flow of firm entry, call it y^*, which gives rise to x^*, and to compute the associated distribution over firm types. In other words, one has to show that (E.3) and (E.5) are satisfied as well. Here we complete this second step.

To that end, consider a constant flow of entry, y, and assume that firms exit according to some prespecified dates, (t_i^*). Based on these two we compute the steady-state over firm types.

Since firms are subjected to independent shocks and since there is a continuum of firms, the number, n_{ij}, of c_i-firms of age $< t_j^*$ is proportional to y:

$$n_{ij} \equiv a_{ij} y. \tag{A.1}$$

Let n_i be the number of c_i-firms at the beginning of some period and let $n_i^{'}$ be the number of c_i-firms at the beginning of the next period. Then

$$n_i^{'} = \alpha_i y + \sum_{j \geq i} g_{ji} n_j + \sum_{j<i} g_{ji}(n_j - n_{ji}). \tag{A.2}$$

That is, n_i receives a flow of new firms, a flow of old firms which used to have cost c_j but have changed to c_i, and losses a flow of c_i-firms which change to c_j. A steady-state is defined by $n_i^{'} = n_i$. Substituting this and (A.1) into (A.2), we get a system of linear equations

$$B \begin{pmatrix} n_1 \\ \bullet \\ \bullet \\ n_m \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \bullet \\ \bullet \\ \alpha_m \end{pmatrix} y, \tag{A.3}$$

where B is some $m \times m$ matrix. These equations admit a unique solution which is linear in y:

$$n_i = (B^{-1}\alpha)_i \equiv a_i y, \qquad i = 1, ..., m. \tag{A.4}$$

This gives the steady-state distribution of firms according to cost. The distribution according to age is as follows. The number of age one firms with cost c_i is 0 or $\alpha_i y$ - depending on whether c_i-firms stay or not. Assume we have the number of

c_i-firms of age $\tau < t$, for some $t > 1$. Then the number of c_i-firms of age t is as follows

$$n_{i,t} = \begin{cases} \sum_j g_{ji} n_{j,t-1}, & \text{if } t_i^* \leq t \\ 0, & \text{otherwise} \end{cases}.$$

Let $n \equiv \sum n_i$ and $a \equiv \sum a_i$. Then $n = ay$. Therefore, we have shown that if a constant flow of firms, y, enters, then there is a unique steady-state stock of firms, n, which is proportional to y: $n = ay$. The steady-state number of consumers that each firm receives each period, x, is then $x = \frac{\nu}{n} = \frac{\nu}{ay}$.

Consider now the x^*, which satisfies (3.3). Then if we set

$$y^* = \frac{\nu}{ax^*},$$

each firm would get exactly x^* new consumers per period. Inserting this y^* into (A.4) we get the equilibrium firm-type distribution. ■