INVESTING IN LONG-TERM CUSTOMER RELATIONSHIPS

ARTHUR FISHMAN

ZIV HELLMAN

Department of Economics, Bar Ilan University, Israel

AVI WEISS

Department of Economics, Bar Ilan University, Israel

Taub Center for Social Policy Studies in Israel

IZA

ABSTRACT. In many markets consumers form long-term relationships with firms. In such settings, a firm's existing customers are valuable assets whose 'loyalty' must be maintained through continued investment. In this paper we assume that consumer loyalty is strengthened with repeated buying but may erode if the relationship is interrupted. In this context we show how a firm's history of costs and sales and the size of its customer base determine the extent to which it invests in maintaining its long term customer relationships by satisfying demand even when this involves a short-term loss. In particular, our model shows that very young firms with small customer bases will prefer losing customers in the short run to absorbing losses in high cost periods, middling sized firms will take the opposite position, absorbing losses for the sake of continuing to build customer bases, while established firms with very large customer bases revert to a willingness to lose customers in the short run.

email: `ziv.hellman@biu.ac.il` .
1

1. INTRODUCTION

There is considerable evidence that relationships between firms and their customers are characterised by considerable inertia; that is, consumers tend to return to buy from the same seller. The long-term nature of customer relationships may be explicit, as in service contracts for banking, telephone and internet services, or in repeated business-to-business transactions. But implicit long-term customer relationships appear prevalent in other contexts as well: for example Dubé, Hitsch and Rossi (2010) and Bronnenberg, Dubé, and Gentzkow (2012) provide evidence in the context of consumer packaged goods that brand preferences are extremely persistent over the consumer life-cycle. This implies that existing customers are valuable intangible assets that raise firm value above the value of physical capital.

A related body of evidence shows that the rate at which firms can expand by acquiring new customers is limited. For example, Foster, Haltiwanger and Syverson (2009) show that even in the context of relatively homogeneous manufacturing goods, the gradual endogenous build-up of demand-side capital limits expansion in sales. Similarly, the marketing literature emphasizes time-of-entry as a key determinant of market share, with early entrants capturing a larger share of the market (see Bronnenberg, Dhar, and Dubé (2009) and Kalyanaram et al. (1995)).

As has been argued in several studies (Fishman and Rob (2003), Rob and Fishman (2005), Gourio and Rudanko (2014)), these dual features of consumer inertia, on the one hand, and a limited rate of firm expansion, on the other, are consistent with the presence of search frictions. In particular, consumers who have sunk search capital with a particular seller prefer to keep buying from that familiar seller than to search for a new one. This, in turn, implies that it is difficult for firms to attract customers from competitors, which limits expansion.

The importance of a firm's base of past customers for its profitability is consistent with the large investments in advertising, introductory offers, etc. that firms typically make to attract new customers. However, merely attracting a new customer does not ensure that the relationship will endure. As in any relationship, once initiated, it is necessary to keep investing in it to maintain it. As the popular adage goes, it is cheaper to keep an existing customer than to acquire a new one. According to a recent study (cited in a UK social media agency Our Social Times), 70% of companies say it is cheaper to retain a customer than acquire one and 49% say that, pound for pound, they achieve better ROI by investing in relationship marketing over acquisition marketing. Thus, while it is important for firms to strive to attract new customers, it is no less important to invest in retaining existing

customers. Accordingly, the focus of this paper is to formulate a model to address this largely unexplored dimension of investment in customer retention.

Specifically, we focus on the extent to which a firm is willing to invest in order to prevent the erosion of its customer base by continuing to satisfy customer demand when, due to temporary supply shocks, doing so involves a short term loss. In our model, when a new consumer enters the market she is randomly matched with a firm. Due to search frictions, such an initial match has the potential to form the basis of a long term relationship. Our main assumption is that customer relationships are strengthened and reinforced with continued interaction, but are weakened if interrupted.

Thus, as long as a firm continues to invest in retaining existing customers, both the size of its customer base and the durability of its customer relationships increase over time. In this setting, we consider how a firm responds to a cost or supply shock that makes it temporarily prohibitively costly to supply the product at an affordable price. In this case, the firm must choose between supplying the product at a current loss or disappointing its customers by being temporarily out of stock. The first option involves a short term loss but constitutes an investment in the longevity of its customer relationships. The latter option is more profitable in the short term, but by eroding its hard earned customer base it may jeopardise long term profit.

We identify three stages of a firm's evolution with respect to its investment in customer retention. Sufficiently young firms whose customer base is still quite small do not find it profitable to incur the short term loss necessary to sustain its customer base. Hence, when such firms are hit by a high cost shock they stop selling until costs come down, even if this jeopardises its fledgling customer base. Medium size firms, by contrast, that have already built up a large customer base but whose customer relationships are still relatively fragile optimally invest to maintain its customer base intact and growing. Finally, once a firm is 'mature' and its customer base is sufficiently large, its customer relationships are sufficiently robust to withstand a temporary interruption of service. Thus, when hit by a high cost shock, the investment behavior of mature firms is more similar to that of very young firms than to firms of medium age.

2. THE MODEL

There are n identical firms that produce and sell an unstorable, homogeneous good. At any period t a firm's unit production cost is denoted c_t. There are two possible unit production costs, c_l and c_h, where $c_h > c_l$. From one period to the next, for each individual firm, these costs are identically

and independently distributed, where $c_t = c_h$ with probability q and $c_t = c_l$ with probability $1 - q$.

We interpret this distribution of costs as follows: assuming that q is relatively low, most of the time a firm enjoys stably low production costs. Occasionally, however, firms' costs temporarily increase as a result of exogenous supply disruptions, machinery breakdown, labour disputes and the like. At any period t a firm learns its production cost for that period before producing and posting its price.

Consumers are identical. Each consumer demands at most one indivisible unit of the product at each period from which she derives utility u, with $c_h > u > c_l$.

Consumers know only the equilibrium price distribution in the market, but do not know which firm charges what price at any period. At each period, a consumer costlessly observes the price of one firm and may pay search costs to observe the prices of additional firms. Specifically, if the consumer is new to the market, she is costlessly matched with a firm at random and views its price. She may either buy from that firm at that price or may incur a search cost, $s > 0$, randomly sampling the price of a second firm. A consumer may also sequentially sample any number of additional firms, at a cost s per firm. After sampling one or more firms she buys from the cheapest one or decides not to buy a unit that period. If a consumer bought from some firm i in the preceding period, she may costlessly return to that firm and observe its current price, or costlessly observe the cost of a randomly chosen firm. After this first search, the consumer may sequentially sample the prices of any number of firms at a cost of s per firm. As a tie breaking rule, we assume that a consumer buys if she is indifferent between buying and not buying.

As the preceding paragraph implies, and as is more fully discussed below, in our setting consumers' search frictions imply that a firm only has access to a subset of consumers, which is referred to as the firm's customer base. We assume that each firm's customer base increases by Y consumers at each period. At this point we assume an exogenous flow of newly born consumers into the market at each time period, Y of whom are randomly matched with each firm. In section 4, Y will be reinterpreted to represent an endogenous flow of consumers between firms.

Consumers are habit forming in the sense that the more often a consumer buys from a particular firm, the more likely she is to return to that firm again. Specifically, at any period, the consumers in a firm's customer base are characterized by their (loyalty) '*grading*': A customer who has bought from firm i for k consecutive time periods, up to the current period, is said

to be k-graded (at firm i). If $k < L$ then a k-graded customer is said to be unattached; if $k \geq L$ then a k-graded customer is said to be loyal. The difference between the two is explained in the following paragraph.

At any period, if the expected utility of a consumer with grading k at firm i from first searching at firm i is greater than or equal to her utility from first searching at a different firm, then she first searches at firm i that period with probability ρ, $0 < \rho < 1$ (and with probability $1 - \rho$ does not return to firm i, and instead searches at a different, randomly chosen firm). If she also buys from firm i at period t, her grading at firm i increases to $k + 1$.

However, if she doesn't buy from firm i in period t, then her behaviour in period $t+1$ depends on her grading. If the consumer is unattached at period t, i.,e., $k < L$, then at period $t + 1$ she does not return to firm i (i.e., her grading at firm i goes down to zero, and she samples a different firm with probability 1). In contrast, if the consumer is loyal, i.e., $k > L$, her grading decreases to $k - 1 \geq L$ and she again returns to firm i at time $t + 1$ with probability ρ. If she does not buy from firm i for two consecutive periods, her grading decreases to $k - 2$ and so on, until her grading drops below L, at which time she no longer returns to firm i. In the borderline case that $k = L$, when seeing a high price posted, the consumer's grading is reduced to $L - 1$ and she no longer returns to firm i.

Our interpretation of this is that habit formation is gradual and is increasingly reinforced with repetition. Hence, when k is small, the habit of buying from firm i is relatively weak and a single interruption suffices to breaks the habit. When k is larger, the consumer's habit is sufficiently entrenched to 'survive' a limited interruption, at least until the consumer's grading erodes below L.

Relatedly, our assumptions can also be interpreted in terms of consumer switching costs. Specifically, here the switching cost is the psychological cost of breaking a habit or routine. This cost depends on how ingrained the habit is, i.e., how long it has been going on. If $k < L$, the cost of changing past behaviour erodes after a relatively short interruption – indeed, in one period in our formulation. Thus in the case of an unattached customer, if she stops buying from a firm for one period, no cost is borne in switching to a different seller (that is, in this case indifference between firms leads to a random choice of a new seller). In the case of a loyal customer, since the habit is more ingrained, a consumer still incurs switching costs after one or more periods in which the consumer does not buy from the firm, until the grading erodes below L.

Our formulation of these ideas is, of course, highly stylised, in that the probability of a customer returning to the same firm changes discontinuously. This is done for tractability, but the intuition behind our analysis also applies if the probability of return changes more continuously.

Comment: ρ may be interpreted as the probability that the consumer survives to the following period. According to this formulation the probability of return (as long as the grading is positive) is 1 if she survives to the next period, but with probability $1 - \rho$ she exits the market before then.

Denote, at each time t, the history of prices that a firm has posted to date by H, i.e., H is a t-long sequence of prices. (We denote the history of a firm at the first period of its existence by $H = 0$). The significance of H for a firm's pricing strategy and profit will become clear presently. At the beginning of each period every firm learns its cost, either c_l or c_h. Its type is defined by its current cost and its history H. We will call a firm with history H and marginal cost c a (c, H)-firm and refer to (c, H) as that firm's type. A firm with cost c_l at period t is called a low cost firm at period t and a firm with cost c_h at period t is said to be a high cost firm at period t.

The firm's pricing strategy determines what price to charge as a function of its type, (c, H). A consumer's search strategy is where to search first (if she has previously bought) and a stopping rule, which, after each price sampled, specifies whether to accept that price, reject it and continue searching, or leave the market without buying that period. The objective of a firm is to maximize its discounted profits, where we assume a constant discount factor of λ that is used by all firms in the market to calculate the present value of expected future profits (and losses). To attain its objective, it must decide what price to charge in each period.

An equilibrium is a strategy for each firm and a search strategy for consumers such that each firm chooses prices to maximize discounted expected future profits given the strategies of all other firms and consumers' search strategies, while consumers search to maximise utility given firms' pricing strategies.

We consider equilibria characterized by two prices, p_l and p_h, where $p_h > u \geq p_l > c_l$, such that all low cost firms post the price p_l and the strategies of high cost firms is yet to be determined. To be specific, we consider equilibria in which $p_l = u$. Note that since in such an equilibrium u is the lowest price of any firm at any date, it is optimal for consumers to accept this price without further search. Conversely, given that consumers accept u without search, and since no consumer will pay more than u, it is indeed optimal for low cost firms to post that price. We comment below on possible alternative equilibria in which $p_l < u$. The qualitative properties

of the equilibrium which will be derived do not depend on which equilibrium value of p_l obtains. Notice that a firm that posts the price $p_h > u$ makes no sales that period. Hence this strategy is equivalent to a strategy of not offering the product for sale that period (i.e., being temporarily 'out of stock').

2.1. **Accumulating a Customer Base.** As long as its price is not greater than u, a firm retains at least part of its customer base, and, in addition, its customer base increases by a fixed number of Y customers per period. In contrast, whenever its price is greater than u it makes no sales that period, and the size of its customer base decreases. The extent to which its customer base decreases depends on its customers' loyalty grading and its price history H.

A firm that has a non-empty set of loyal customers, as specified in the previous section, will be termed an *established firm*. Otherwise it is a *junior firm*.

A necessary (but not sufficient) requirement for being an established firm is having at some point accumulated at least L consecutive time periods of offering a low price in the market.

2.2. **Losing Customers.** A firm with *pricing history* (or simply history, for short) H will be said to be H-*tenured*. Prior to the first period of its existence, a new firm has history H that is the empty set. We call such a firm 0-tenured.

For junior firms (who have no loyal customers), for all practical purposes the only datum that is relevant in their history H is how many consecutive periods of low prices they have posted since the last time they posted a price above u. This is the only thing that determines the number of (unattached) customers they have in their customer base. A junior firm that posts a high price loses all its customer base in the next period and essentially 'starts again' as if it is a new firm; we will therefore say in such a case that it returns to being a 0-tenured firm even if its history H is non-empty. The next period (if it sells) it may be regarded as being 1-tenured again, and so forth.

For established firms matters are more complicated. Consider, for example, a firm whose initial history is $L + 3$ consecutive periods of low prices, enabling it to build a customer base that includes some loyal customers. If in the next period it posts a high price, it loses all customers with a loyalty grading lower than L, but its customer base is not reduced to zero. If it then has two consecutive periods of low prices followed by a period of a high

price, it loses the customers that it gained in the two intervening periods between the high prices, but again retains some customer base, and so on.

Denote the size of the customer base of a firm with history H by B_H. It is possible to calculate an upper bound for the size of a customer base for any history H of length τ. Clearly, amongst all such histories, the ones that give a firm its largest possible customer base are those in which the firm has posted prices less than or equal to u in all τ periods. Label such a history as H_τ^*. Then it follows from the assumptions of the model that:

$$B_{H_1^*} = Y$$
$$B_{H_2^*} = Y + \rho Y$$
$$\vdots$$
$$B_{H_\tau^*} = Y + \rho Y + \rho^2 Y + \ldots + \rho^{\tau-1} Y$$
$$= Y(1 + \rho + \rho^2 + \ldots + \rho^{\tau-1})$$

Denoting
$$\eta_\tau := 1 + \rho + \rho^2 + \ldots + \rho^{\tau-1}$$

we can write

(1)
$$B_{H_\tau^*} = \eta_\tau Y.$$

In particular, we can calculate from this the maximal number of unattached consumers, $B_{H_L^*} = \eta_L Y$.

Even though $\eta_\tau > 1$ for all τ, as ρ is bounded away from 1, η_τ is asymptotically bounded in the limit, hence $\overline{B} := \lim_{\tau \to \infty} B_{H_\tau^*}$ will also be a bounded real number, since $\overline{B} = \lim_{\tau \to \infty} \eta_\tau Y = Y/(1 - \rho)$. It follows that for any history H of any firm, $B_H < \overline{B}$.

Equation 1 can be generalised to any history H. Suppose initially that a firm has a pricing history H_τ^*, in which it has posted prices less than or equal to u for $\tau > L$ periods, and hence has customer base $B_{H_\tau^*} = \eta_\tau Y$. If that firm at time $\tau + 1$ posts a high price, it will lose the $\eta_L Y$ unattached customers and at time $\tau + 2$ gain Y new customers. But much the same reasoning applies to *any* history H for an established firm: posting a high price means losing at most $\eta_L Y$ unattached consumers. In any case, there is some unambiguous positive parameter η_H such that the customer base of an H-tenured firm can be given as $B_H = \eta_H Y$.

Since firms have no fixed costs, a firm can always earn positive profit in low cost periods by selling to whatever customers come its way. In addition, by staying in the market, it retains the option of selling when costs are low in the future at profit $u - c_l$ per customer. Thus, it can never be optimal for a firm to exit the market.

3. STRATEGIC ANALYSIS

Firms in the model as presented do not interact or compete. We can therefore focus attention on a single-firm decision process.

For a low cost firm at time t, whatever its history H up to that point, it is obviously optimal to charge the price u. By doing so it earns positive profit in that period and in addition grows its customer base. In contrast, if its price is greater than u it earns zero that period and in addition reduces the size of its customer base.

Matters are less obvious in the case of a high cost firm. If its price is u, it takes a loss that period but preserves and grows its customer base, while if its price is greater than u it avoids a current loss, but reduces the size of its customer base. A high cost firm must choose between absorbing temporary losses in the expectation that customer base retention will enable it to offset those losses in the future and avoiding immediate losses, even at the cost of losing some or all (if it is a junior firm) of its current-period customer base.

3.1. **Dynamic Programming Analysis.** To simplify notation, we will denote by H^- (respectively, H^+) the tenure of an H tenured firm in the next time period if it posts a high price (respectively, low price) in the current period. In moving from H to H^+, a firm increases its customer base in the subsequent period, while moving from H to H^- reduces its customer base. Hence it is always the case that $B_{H^+} > B_H \geq B_{H^-}$. If a firm is a junior firm, at H^- it has lost its customer base and its situation is equivalent to that of a newly starting 0-tenured firm with the empty set for its history. An established firm will have a more complicated relationship between H and H^-.

Denote by $R(c, H)$ the maximum discounted profit of a (c, H)-firm, i.e., its maximal discounted profit if it implements an optimal pricing strategy for all future time periods. We may cast the firm's decision in dynamic programming format, in which (c, H) is the state variable:

$$(2) \quad R(c, H) = \max \Big\{ \lambda \big[qR(c_h, H^-) + (1 - q)R(c_l, H^-) \big],$$

$$B_H(u - c) + \lambda \big[qR(c_h, H^+) + (1 - q)R(c_l, H^+) \big] \Big\}$$

Equation (2) reflects the decisions facing a firm at the beginning of each period, after it has discovered its costs for that period. In detail, it may take a decision to:

- Post a price above u this period and then become an H^--tenured firm in the next period. In that case, there are zero profits in this

period, and the discounted return in the next period will be either $\lambda R(c_h, H^-)$, with probability q, or $\lambda R(c_l, H^-)$ with probability $1 - q$.

- Post price u and become an H^+-tenured firm in the next period. In that case, the discounted return in the next period will be either $\lambda R(c_h, H^+)$, with probability q, or $\lambda R(c_l, H^+)$ with probability $1 - q$. The firm's profit in this period is $B_H(u-c)$; note that if $c = c_l$, the profit is a positive quantity, whilst if $c = c_h$, the profit is a negative quantity.

By dynamic programming techniques there is a solution to Equation (2), which, moreover, is a stationary solution, meaning that the action chosen at each time period depends only on the state variable (c, H) at that time period and not on what actions were chosen in prior time periods. But we can glean much more than simply a statement of solution existence. Several interesting characteristics may be derived from the form of the solution.

Note first that given the basic parameters, i.e. q, λ, c_h and c_l, which are considered fixed, for any junior firm the term $\lambda\big[qR(c_h, H^-) + (1 - q)R(c_l, H^-)\big]$ equals a constant, namely, $\lambda\big[qR(c_h, \emptyset) + (1 - q)R(c_l, \emptyset)\big]$. Furthermore, the quantities $u - c_l$ and $u - c_h$ are also fixed throughout. For ease of exposition, we will denote these by the following constants:

$$
\begin{aligned}
G &:= u - c_l \\
D &:= u - c_h \\
Z &:= \lambda\big[qR(c_h, \emptyset) + (1 - q)R(c_l, \emptyset)\big]
\end{aligned}
$$

Note that $G > 0 > D$.

For established firms, the size of the customer base at H^- is a variable that may depend very much on the history. Denote

$$
Z_{H^-} := \lambda\big[qR(c_h, H^-) + (1 - q)R(c_l, H^-)\big].
$$

Lemma 1.

(a) *If $B_H > B'_H$ for two histories H and H', then $R(c, H) \geq R(c, H')$.*
(b) $R(c_l, H) > R(c_h, H)$, *for all H.*
(c) $R(c_l, H) > 0$, *for all H.*
(d) $Z_{H^-} > 0$, *for all H.*
(e) $R(c, H) \geq Z_{H^-} > 0$, *for all types (c, H).*
(f) $R(c_l, H) > Z_{H^-} > 0$, *for all H*

Proof. (a) This can be proved by considering a 'replication' strategy. That is, for any strategy used by the H'-tenured firm for all subsequent time periods and production costs, one can imagine the H-tenured firm copying, one for one, every action chosen by the H'-tenured firm, but starting with a larger customer base. Then the H-tenured firm will enjoy larger profits in low cost periods in which it sells at a low price (because of the larger customer base). If (under the optimal strategy) the firm with a smaller starting customer base chooses to post a low price in a high cost period, it is because it calculates that it will make up the loss in later periods with a larger customer base; but this same reasoning holds *a fortiori* for the firm with the larger starting customer base. Hence, the return is greater than or equal under all conditions.

(b) $u - c_l$ is a positive quantity and $u - c_h$ is negative. This means that in Equation (2), the terms $B_H(u - c)$ and $B_H(u - c) + \lambda\big[qR(c_h, H^+) + (1 - q)R(c_l, H^+)\big]$ are always greater when $c = c_l$ compared to the case when $c = c_h$. At the same time, $\lambda\big[qR(c_h, H^-) + (1-q)R(c_l, H^-)\big]$ is always less than or equal to $\lambda\big[qR(c_h, H^+) + (1 - q)R(c_l, H^+)\big]$ (this follows from (a) above, since $B_{H^+} > B_{H^-}$). Hence the maximum for c_l always dominates the maximum for c_h.

(c) At worst, a firm with low costs can sell at a positive profit at the current time period and then exit the market. The maximum of Equation (2) for $c = c_l$ is therefore always positive.

(d) Note that $R(c_h, \emptyset) \geq 0$ because of zero profit from always charging a high price. On the other hand, we have already seen in (c) above that $R(c_l, H) > 0$ for all H, hence, in particular, $R(c_l, \emptyset) > 0$. By assumption $q < 1$, hence $1 - q > 0$. Putting this all together, one concludes that $Z = \lambda\big[qR(c_h, \emptyset) + (1 - q)R(c_l, \emptyset)\big] > 0$. Since $Z_{H^-} \geq Z$ (as the customer base can never be a negative quantity for any history), the conclusion follows.

(e) This follows immediately from (d), since a firm of any type can always choose Z_{H^-} as an option.

(f) Putting together (b) and (e) yields $R(c_l, H) > R(c_h, H) \geq Z_{H^-}$.

∎

3.2. **High Cost Firms.** We have already seen that a low cost firm always charges the low price. We seek conditions under which a high cost firm also behaves this way, i.e., in which a high-cost H-tenured firm will choose to bear a one-period loss and continue as an H^+-tenured firm.

Recall the assumption that ρ is bounded away from 1, which implies that $\overline{B} = \lim_{\tau \to \infty} B_{H^*_\tau}$ is bounded from above. As this is the asymptotic value of the trajectory of monotonically increasing maximal customer-base

accumulation, $B_H < \overline{B}$ for any H. The best that an H-tenured firm can expect in terms of future profits is to have low costs for the rest of time from the present onwards. This translates into future profits of $B_H G + \lambda B_{H^+} G + \lambda^2 B_{H^{++}} G + \ldots$ in the rosiest of possible worlds. It follows in this case that the discounted expected value of a (c, H)-type firm is always bounded from above, in particular that $R(c, H) < \frac{1}{1-\lambda} \overline{B} G$.

Now, suppose (recalling that D is a negative quantity) that $|D|$ is sufficiently large such that for some history H the inequality $|DB_H| > \frac{1}{1-\lambda} \overline{B} G$ holds. Then a (c_h, H)-firm will always choose to lose customers and be an H^--tenured firm in the next period, as its one-period losses by selling at a price u could never be recouped, even if it would forever-more have only low marginal production costs from that point on.

The previous paragraph contains a negative result, presenting a condition under which firms firms never sell at a loss. Are there sufficient conditions that yield the opposite behaviour, that is, under which a high cost firm charges the low price, taking a current loss in order to preserve its customer base? The next proposition provides just such a condition.

Proposition 1. *For H such that*

(3) $$B_H \Big[D + \frac{1}{1-\lambda}(G - q(G - D)) \Big] > Z_{H^-},$$

an H-tenured firm, of any type, will post a low price.

Proof. The solution to Equation (2) must be a strategy that prescribes taking the action that grants maximal future expected return at each state (c, H). Suppose, by way of contradiction, that in particular this strategy calls for a (c_h, H)-type firm to become an H^--tenured firm, yet one can show that there is an alternative strategy that calls for posting a low price and grants greater future expected return. Then this contradiction would suffice to conclude that the solution to Equation (2) must also call for a (c_h, H)-type firm to continue on to being an H^+-tenured firm.

We therefore consider the following strategy that is available to a (c_h, H)-type firm: commit to selling at price u in every time period from now on, including the current period t. This strategy is equivalent to accepting a definite loss $B_H D$ now and then forever-more playing a lottery in each time period $\kappa > t$, with that lottery returning an expected payoff of $qB_{H_\kappa} D + (1 - q)B_{H_\kappa} G$ at each time period κ, where H_κ denotes the history H followed by $\kappa - t$ consecutive periods of posting a low price.

The discounted expected payoff from playing such a lottery in every time period $\kappa > \tau$ is either greater than or equal to $\frac{1}{1-\lambda} B_H[(1 - q)G + qD]$, if $(1 - q)G + qD > 0$, or less than or equal to $\frac{1}{1-\lambda} B_H[(1 - q)G + qD]$,

otherwise (since $B_{H_\kappa} > B_H$). However, since by assumption

$$B_H D + B_H \frac{1}{1 - \lambda}[G - q(G - D)] > Z_{H^-} > 0$$

and $B_H D < 0$, it must be the case that $(1 - q)G + qD > 0$, hence the discounted expected payoff of this strategy is definitely greater than or equal to $\frac{1}{1-\lambda} B_H[(1 - q)G + qD]$.

It follows that by adopting the strategy of selling at a price u in every time period in the future, a (c_h, H)-type firm gains more than it loses in the single period by absorbing a loss $B_H D$ due to high costs, and that this is preferable to Z_{H^-}. ∎

In Theorem 1 we present a sufficient condition that guarantees that any junior firm that has a 'big enough customer base', in a way we shortly make precise, will always choose to suffer short term losses in exchange for further increasing the customer base in later time periods.

Note that for junior firms (i.e., those without loyal customers) the only relevant aspect of their history is the number of consecutive time periods of low prices immediately prior to the present. This is because one high price posting reduces their customer base back down to zero. Hence, for these firms we can denote by B_t the customer base of a junior firm of history H that has posted low prices in the previous $t < L$ time periods (where t, of course, is simply an integer) without needing to specify any further details of H.

Theorem 1. *If for some $k < L$,*

(4) $$B_k\left[D + \frac{1}{1 - \lambda}(G - q(G - D))\right] > Z$$

then for all $k \leq \tau < L$, a τ-tenured firm will absorb single period losses by charging the low price and continue on to become a $\tau + 1$-tenured firm.

Proof. Equation (4) is the special case of Equation (3) in which $Z_{\tau^-} = Z$, hence satisfying Equation (4) is sufficient for a k-tenured firm to absorb single period losses.

At the same time, Z is fixed for time periods less than L, and B_τ is an increasing function of time. Hence, for all $k \leq \tau < L$ a τ-tenured firm will absorb single period losses. ∎

Theorem 1 exhibits a condition that guarantees that a junior firm, once it has accumulated a sufficiently large customer base, will post low prices even in high cost periods. This occurs when Equation (4) is sastified, i.e., if

and when B_k grows sufficiently large to satisfy

$$(5) \qquad\qquad B_k > \frac{Z}{D + \frac{1}{1-\lambda}(G - q(G - D))}.$$

The left hand side of Equation (5) is a function of ρ, the customer retention rate, and of Y, the expected number of new customers a firm picks up in a single time period. We can therefore roughly say that the closer ρ is to 1, i.e., the greater customer loyalty is, and the larger the potential new customer base is, the likelier we are to see young firms absorb temporary losses (although some care should be used in this interpretation, as the right hand side is also indirectly dependent on ρ and Y). In the extreme case, it is possible for the initial customer base and expected customer loyalty values to be so large that even freshly minted new firms with $k = 1$ will post low prices in high cost periods, as will all k-tenured firms with $k < L$.

However, once a firm becomes established and it begins drawing a loyal customer base, Equation (3) is the relevant factor and matters become less simple. This is because, while B_H might grow with time for a firm consistently posting low prices, so does Z_{H^-}. Does this mean that behaviour can be reversed, with a sufficiently established firm passing on high prices to consumers? It turns out that this is indeed the case. For the next result, recall that $B_{H_L^*}$ is the maximal number of unattached consumers a firm can have, that $\overline{B}$ is the asymptotically maximal customer base, and that D is a negative quantity.

Theorem 2. *If $|\overline{B}D| > B_{H_L^*}(u - c_l)/(1 - \rho)$, then a firm of history H with sufficiently large customer base B_H will prefer to pass on high costs to its customers, and hence to lose some of its customer base, rather than absorbing losses by posting low prices at high cost periods.*

Proof. The question that an H-tenured firm faces when it encounters a high cost period is a relatively simple one: which course of action is more remunerative, absorbing a loss of $B_H D$ and then gaining the expected value of an H^+-tenured firm, or instead choosing the expected value of an H^- firm without paying any price? Using the notation $R(H)$ for the maximal discounted profit of an H-tenured firm, this translates into the question: which is greater, $\lambda R(H^-)$ or $\lambda R(H^+) + B_H D$?

Suppose first that an established H-tenured firm suffers a high cost period at time t, but then afterwards enjoys the ideal situation of subsequent low cost periods forever more, for all $\tau > t$. In this case the optimal strategy is clearly always to post low prices subsequently, and one can then ask what is better at time t, to move to H^- or $H+$. Posting a high price means losing unattached customers, hence $B_{H^-} < B_{H^+}$. Since, from here on, in

either case low prices are always posted and the customer base grows, the discounted future profits are either $B_{H^+}(u-c_l)/(1-\rho)$ or $B_{H^-}(u-c_l)/(1-\rho)$, and the choice at time t is between $B_{H^-}(u-c_l)/(1-\rho)$ (by moving to H^-) or $B_{H^+}(u-c_l)/(1-\rho) - |B_H D|$, if a one-period loss is taken (again, recalling that D is negative, hence so is $B_H D$).

Now, $B_{H^+} - B_{H^-}$ is bounded by the maximal number of unattached customers, while $|B_H D|$ can potentially grow (in absolute value) asymptotically to $|\overline{B}D|$, hence, given the assumptions of the statement of the theorem, a firm with a sufficiently large customer base will come to the conclusion that absorbing a one period loss in the hopes of making up the difference in later time periods is not profitable.

The above analysis was based on the supposition that low cost periods always follow H, but we can drop that assumption and still come to the same conclusion. This is because for any subsequent history of costs and any pricing strategy adopted by a firm choosing H^+, the same pricing strategy can be adopted by a firm choosing H^-. It will then be the case that the differences in future profits between choosing H^+ and H^- will be even less than the above calculated $(B_{H^+} - B_{H^-})(u-c_l)/(1-\rho)$ (because there will be periods of no income and losses of customers, and periods of absorbing losses), while at the same time $|B_H D|$ can still potentially grow to the same maximal value. The conclusion follows. ∎

Note that if ρ is sufficiently close to 1 then the condition in Theorem 2 is almost guaranteed to hold.

It might seem at first glance that Proposition 1, along with Theorem 1, and Theorem 2 are somewhat contradictory: one contains sufficient conditions for firms to post low prices in high cost periods from a certain time period onwards, while the other describes firms that do exactly the opposite, posting high prices in high cost periods, absorbing losses for future gains.

There is, of course, no contradiction here. Junior firms that are just starting out have small customer bases. They have little to lose but much to gain in choosing temporarily to exit the market and 'go back to square one' when faced with high prices. At a certain age, junior firms may accumulate a sufficiently large customer base to cause them to change this behaviour and tolerate a single period of losses due to posting low prices despite high costs, based on the calculation that continuing to accumulate a customer base going forward offsets the temporary loss.

Sufficiently established firms, in contrast, have such large loyal customer bases that they can afford to lose a few unattached consumers and still profitably proceed forward in time without needing to bear the loses by posting low prices in high cost periods (although it is also always possible that a

very established firm will experience a sufficiently long run of high cost periods to reduce it back to being a junior firm). The model here accommodates differing behaviour, depending on the size of customer base accumulated by a firm and the expected loyalty of consumers.

We summarise the conclusions of this section as follows. Assuming that the model satisfies the assumptions in the statements of Theorems 1 and 2, then junior firms of tenure ages 0 and k prefer losing their customer bases to absorbing losses by posting low prices when costs are high. Junior firms beyond tenure age k, and established firms with up to a certain size of customer base prefer absorbing losses by posting low prices when costs are high to suffering losses to their customer bases. Established firms of sufficiently large customer base, in contrast, have accumulated a sufficiently large pool of loyal customers to enable them to afford to lose some customers by posting high prices when costs are high.

Thus, under these assumptions, our analysis is consistent with the prediction that sufficiently large firms change prices more often than smaller ones, as the former increase their price in response to cost shocks more frequently than smaller ones. Buckle and Carlson (2000) present evidence consistent with this feature.

We have derived the equilibrium when the low price is u. There may also exist equilibria in which the low price $p_l < u$. At these equilibria, low cost firms are prevented from posting prices greater than p_l by the consumers' credible 'threat' not to return to a firm whose price is greater than p_l but less than or equal to u. It is easy to ascertain that none of our results are affected in any meaningful way if the equilibrium value of p_l differs from u.

4. The Multiple Firm Perspective: Birth and Death

In the model as presented in the previous section, the total number of firms was an exogenous quantity. Here we extend the model to endogenise the number of firms as an equilibrium quantity under the assumption of barrier-free entry and zero expected profit from entry.

To accomplish this, in this section we add firm birth and death to the model. In detail, our new assumptions are that:

(1) In every time period, a quantity Θ of new firms are 'born', newly entering the market (details on how Θ is determined are explained later in Theorem 3). New market entrants bear fixed entry costs. The new firms entering the market at the same time are said to belong to a cohort.

(2) The number of time periods that have elapsed since the firm was born is the 'cohort age' of the firm. Note that this is distinct from

the firm's number of tenure periods, which depends on the number of consecutive time periods that the firm has sold at a price at or below u, whereas the age is chronologically determined with no relation to any actions taken by the firm.

(3) Shocks causing firms to experience high cost time periods continue to be i.i.d. over different time periods. Here we also add the assumption that such shocks are independent in each time period between firms; for example, a firm might suffer from internal pipeline disruptions or face a strike in its labour force, but these shocks are idiosyncratic to each firm. In other words, a general market shortage in, say, the raw materials used simultaneously by all firms is excluded here.

(4) In addition to high cost shocks, at the end of every time period t, each firm faces an i.i.d. probability $0 < \delta < 1$ of 'dying', meaning that it goes bankrupt due to an exogeneous shock and forever leaves the market. All such dying firms release their entire accumulated customer base.

(5) There is a finite number C of total consumers in the market.

One immediate effect of this is to change the effective discount rate used in the previous sections. The death rate itself is a form of discount factor. However, all we need do is to assume that λ already incorporates the death rate; all the previous calculations using λ then follow as before, unchanged.

At any given moment t, there are Θ newly born firms, constituting a cohort. In parallel, there are in expectation $(1 - \delta)\Theta$ firms remaining from the cohort of time $t - 1$, $(1 - \delta)^2\Theta$ firms remaining from the cohort of time $t - 2$, and so forth. It follows that the expected total long run number of firms in the market at any given time is $\frac{1}{\delta}\Theta$.

Knowing the total number of firms in the market and the distribution of their chronological ages is not sufficient; we also need to know the distribution of customer bases, which in turn depends on the pricing history H of the firms, as the determinant of B_H. For junior firms, this task is simplified by the fact that, as mentioned above, the customer base is determined entirely by the total number τ of immediately preceding periods of low cost subsequent to the last high cost period.

If we suppose that firms implement the optimal strategy detailed in the previous sections, then the pricing history is entirely determined by the cost history encountered by a firm. Recall that the probability at any time of a high cost c_h is q and of a low cost c_l is $1-q$. We introduce the notation π_t^H to denote the probability of a firm at t having customer base B_H, conditional on surviving to cohort age t. For junior firms, we will write π_t^τ for the

probability of the firm at t having had τ of immediately preceding periods of low cost subsequent to the last high cost period, again conditional on surviving to cohort age t.

For example, if $\tau = 1$ and $t = 1$, then clearly $\pi_1^1 = 1 - q$, as this is simply the probability of a 0-aged firm drawing low production cost in its first period. Somewhat more generally, for $\tau < k$ (where k is the cutoff at which which junior firms begin absorbing losses in high cost time periods), one has $\pi_t^\tau = q(u - c_l)^t$ for $t - 1 \geq \tau$ (calculated as the probability of any cost history of length $t - \tau + 1$ followed by a high cost period reducing the tenure age to 0 followed by τ consecutive periods of low costs) and $\pi_t^\tau = (u - c_l)^t$ for $t = \tau$. In principle, similar considerations can be applied to calculate π_t^H for all pairings of t and histories H, taking into account the probabilities of the various costing histories, which further determine pricing histories and thus customer base size.

Denote the expected number of firms of pricing history H and cohort age t by F_t^H. Since the probability of a firm surviving to cohort age t is $(1 - \delta)^t$ and there are Θ initial firms in each cohort, one has

$$F_t^H := (1 - \delta)^t \Theta \pi_t^H.$$

With this, it is possible to count up the total expected number of firms with customer base of size B_H in the economy, which we denote by β^H and calculate as:

$$(6) \qquad \beta^H := \sum_{t=0}^{\infty} F_t^H$$
$$= \sum_{t=0}^{\infty} (1 - \delta)^t \Theta \pi_t^H$$
$$= \Theta \sum_{t=0}^{\infty} (1 - \delta)^t \pi_t^H.$$

β^H is, of course, bounded, since the set of firms with pricing history H is a subset of the total set of firms, whose expected number is $\frac{\Theta}{\delta}$.

Now, recalling that there is a well-defined value η_H such that $B_H = \eta_H Y$ and that the total number of consumers in the market is C, it follows from

Equation (6) that

$$(7) \qquad C = \sum_{\{\text{histories } H\}} B_H \beta^H$$

$$= \sum_{\{\text{histories } H\}} Y \eta_H \beta^H$$

$$= Y \sum_{\{\text{histories } H\}} \eta_H \beta^H$$

$$= \Theta Y \sum_{\{\text{histories } H\}} \sum_{t=0}^{\infty} (1-\delta)^t \eta_H \pi_t^H$$

It might seem from Equation (7) that to calculate Y, the number of new customers attracted by a firm in each time period, one could simply set $Y = C/\sum_{\{\text{histories } H\}} \eta_H \beta^H$, with β^H regarded as a function of Θ. But matters are not this straight-forward. The value of β^H (via the value of π_t^H) is also a function of Y, since Y determines profits, which in turn affects strategic calculations of firms, and hence determines several important elements, such as the critical point beyond which firms become willing to absorb short-term losses instead of going back to 0-tenured, etc. Therefore β^H is, in part, a function of Y as well as of Θ. The equation

$$(8) \qquad Y \sum_{\{\text{histories } H\}} \eta_H \beta^H (Y, \Theta) = C,$$

(with β^H written as a function of Y and Θ) needs to be regarded as implicitly determining Y for a given level of Θ.

This leaves us two degrees of freedom: assuming that C and δ are given exogenously, a given Θ determines Y, while conversely a given Y will determine the equilibrium Θ. A degree of freedom can be removed if we add fixed costs to the model and enable barrier-free entry to the market.

We first need to consider the expected profit a firm can have, conditional on attaining a cohort age t before succumbing to an exogenous shock sending it to bankruptcy. Denote this expected profit by ζ_t.

Calculating ζ_t in general is, like π_t^H, possible but tedious. We present the method of calculating it in detail for $t = 2$. A firm could see a run of two low cost periods, $c_l c_l$, with probability $(1-q)^2$, in which case it will have sold in two consecutive periods, seeing total customer bases $B_0 + B_1$. Alternatively, the run could be $c_l c_h$ or $c_h c_l$, each with probability $q(1-q)$, with total customer base B_0, or $c_l c_l$ with probability q^2 and zero profit. It

follows that

$$\zeta_2 = [(B_0 + B_1)(1 - q)^2 + B_0 2(1 - q)q](u - c_l).$$

This can in principle be generalised to any t, by taking into account all the possible cost histories of length t, their respective probabilities of occurring, and the pricing strategy of the firms.

Theorem 3. *If a fixed cost is incurred as a price of entry to the market, then there is a minimal level of initial customer base $\underline{Y}$ below which no firm will enter the market. This will determine, at equilibrium, the number of new firms entering the market at each time period.*

Proof. Suppose that a fixed cost of Q is incurred as the price of entry into the market within a cohort. Since the probability of attaining cohort age t is $(1 - \delta)^t$, it follows that the expected profit of a firm entering the market with a cohort is

(9)
$$E := \sum_{t=0}^{\infty} (1 - \delta)^t \zeta_t$$

As ζ_t depends on the size of the expected customer base at time t, which in turn depends on Y, it follows from Equation (9) that there is some minimal $\underline{Y}$ below which $E < Q$, in which case no firm will wish to enter the market. That, in turn, determines Θ, the number of new firms entering the market at equilibrium, using Equation (8).   ∎

REFERENCES

BRONNENBERG, B. J., S. DHAR, AND J. P. DUBÉ (2009), Brand History, Geography, and the Persistence of Brand Shares, *Journal of Political Economy*, 117, 87-115.

BRONNENBERG, B. J., J. P. DUBÉ, AND M. GENTZKOW (2012), The Evolution of Brand Preferences: Evidence from Consumer Migration, *American Economic Review*, 102 (6), 2472–2508.

BUCKLE, R. AND J. A CARLSON (2000), Menu costs, Firm Size and Price Rigidity, *Economics Letters*, 66 (1), 59–63.

DUBÉ, J. P., G. J. HITSCH, AND P.E. ROSSI (2010), State Dependence and Alternative Explanations for Consumer Inertia, *RAND Journal of Economics*, 41, 417-45.

FISHMAN, A., AND R. ROB (2003), Firm Growth and Industry Dynamics, *Journal of Economic Theory*, 109, 24–38.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2009), The Slow Growth of New Plants: Learning about Demand?, *unpublished manuscript*, University of Chicago.

GOURIO, F., AND L. RUDANKO (2014), Customer Capital, *Review of Economic Studies*, 81 (3), 1102–1136.

KALYANARAM, G., W. T. ROBINSON AND G. L. URBAN (1995), Order of Market Entry: Established Empirical Generalizations, Emerging Empirical Generalizations, and Future Research, *Marketing Science*, 14, G212–221.

ROB, R., AND A. FISHMAN (2005), Is Bigger Better? Customer Base Expansion through WordofMouth Reputation, *Journal of Political Economy*, 113 (5), 1146–1162.