Economics Division
School of Social Sciences
University of Southampton
Southampton SO17 1BJ UK

Discussion Papers in Economics and Econometrics

Explaining the Shakeout Process: A Successive Submarkets Model

Jian Tong

No. 0703

This paper is available on our website
http://www.socsci.soton.ac.uk/economics/Research/Discussion_Papers

ISSN 0966-4246

Explaining the Shakeout Process: A 'Successive Submarkets' Model[1]

Jian Tong[2]
University of Southampton

August 28, 2007

[1]The author wishes to thank Marcus Asplund, Bruce Lyons, Jusso Valimaki, Ken Hendricks, Joseph E. Harrington Jr., Leonardo Felli, two anonymous referees, and particularly John Sutton for helpful comments and suggestions. Any remaining errors are the author's sole responsibility.

[2]Correspondence address: Economics Division, School of Social Sciences, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom. Tel: +44-23-8059 6801, Fax: +44-23-8059 3858, E-mail: j.tong@soton.ac.uk

Abstract

This paper explains contemporaneous exit and entry in a new industry with a diffusion process across submarkets. It allows a re-interpretation of the shakeout process in some industries in a novel way. The industry is a collection of initially inactive independent submarkets; the timing of their activation is determined by an exogenous aggregate diffusion process. New submarket opening attracts new entry. However, the post-entry endogenous sunk investment requirement induced by innovations also forces much exit to follow entry. The aggregate market thus has overlapping exit and entry; and has a shakeout if the aggregate diffusion process follows a typical S-shape.

Key words: industrial dynamics, diffusion of product innovation, submarkets, market definition, shakeout, barriers to survival

JEL classification: D21, L11, L13, O33

1 Introduction

The shakeout process, first documented for a wide set of U.S. manufacturing industries[1] by Gort and Klepper (1982), involves excessive initial entry followed by a process of rapid and substantial exit. Later work has focused on giving a more precise empirical description of the process, particularly details about gross entry, exit and firm survival patterns (Klepper and Miller (1995); Klepper and Simons (2000), Klepper and Simons (2001); Klepper (2002)). A surprising finding of the recent related literature is the fact that at the (narrowly-defined) industry level the occurrences of entry mostly overlap with exits. This does not square easily with the textbook model, which predicts that within each market entry should occur in response to positive net profit at the margin while exit should occur in response to net loss, but they should never occur at the same time.[2] In some industry shakeouts, the overlapping entry and exit are very dramatic, making them particularly puzzling.

This paper intends to explain contemporaneous exit and entry in a new industry with a diffusion process across submarkets. It allows a re-interpretation of the shakeout process in some industries in a novel way. The main idea is that the aggregate market data typically used in studies of shakeouts can be interpreted as coming from aggregation of (oligopolistic) submarkets linked by some specific (exogenous) process of diffusion of the major product innovation.

The notion of submarket is central in the current investigation. It highlights the importance of market definition[3] to understanding market structure.[4] Applying the standard textbook delineation of market definition, a submarket (as opposed to the aggregate market) of an industry (ideally) should identify a relevant *economic market* which "consists of a set of products, a set of buyers, a set of seller, and a geographic region in which the buyers and sellers interact and determine prices for each product" (Church and Ware (2000)). Under ideal market definition any two products offered in the same submarket should be close substitutes, while any two products traded in separate submarkets ought to be distant substitutes. Given that the definition of submarkets has both a product dimension and a geographical dimension, any two separate submarkets should share a 'marked gap in the chain of substitutes' (Robinson (1954)) in either of these two dimensions.

Following the pioneering work of Sutton (1997) and Sutton (1998), we base our analysis on the key assumption that the industries (of interest to the current

[1] These are very disaggregated categories of industries, which correspond to 5 or 7-digit SIC codes.

[2] This contradiction has been previously noted by Geroski (1995) and Caves (1998) in their surveys of the broad literature on firm entry, exit and growth. The findings cited there, however, are based on more aggregated industry data, at 2 to 4-digit SIC code levels.

[3] The issue of market definition is not new (see Schmalensee (1982)). It has been central in the area of enforcement of competition policy. In fact it is often the case that competition laws have a statutory requirement for identification of market in which there is a substantial lessoning or prevention of competition.

[4] See Sutton (1997) and Sutton (1998) for the role of submarkets in affecting aggregate market structure, particularly firm size distribution.

study) consist of multiple (approximately) independent submarkets. We can then attribute the determination of the evolution of industry-wide market structure to the connection between innovation diffusion process and submarkets. To elaborate how the notions of submarkets and exogenous diffusion process work together, a summary of the theoretical model is in order. The industry is a collection of initially inactive independent submarkets with identical demand structure. The only link among submarkets is an exogenous aggregate diffusion process that determines when a particular submarket becomes active and when production starts (as a function of how many submarkets are already in operation). Correspondingly there is an aggregate timing relative to the industry as a whole and the timing in each particular submarket. The complete timing in a submarket is: (i) first a new product is exogenously invented in submarket j (in this case submarket j is the first market to be activated in the industry) or it is introduced from a different submarket into submarket j by the exogenous diffusion process. Firms in submarket j have to decide if entering production of such new good: in this case a forward looking free entry condition applies. This is time $t_j = 0$ for submarket j. (ii) After entry, all firms initially play a symmetric Cournot game of quantity competition. This is from time $t_j = 0$ to time $t_j = T_j$ for submarket j. (iii) At some random point of time T_j (determined by the outcome of a random search process) an innovation arrives in the submarket and presents all existing firm with the option to incur an endogenous sunk cost to increase the quality of their products. The firms' decisions on whether making the sunk investments or not are in fact determining if staying or exiting: a second zero profit condition applies for them and a shakeout might take place. (iv) For times $t_j \geq T_j$ the remaining firms play a repeated static Cournot game with differentiated quality. The game in a submarket is solved with backward induction starting from time T_j. The industry-wide evolution of aggregate quantity of the new good produced and aggregate number of firms can easily be derived by aggregating the submarkets and by taking into account the timing imposed by the exogenous diffusion process.

It is easy to see why this model predicts industry-wide contemporaneous exit and entry while still consistent with the textbook model of entry and exit. Here the relevant economic markets are the submarkets. Within each submarket, the model predicts that exit should follow entry rather than occur at the same time. It is the industry-wide diffusion process that makes the exits in some submarkets to coincide in time with the entries in some later activated submarkets. As we impose a plausible specific structure to the diffusion process, i.e., assuming that the number of active submarkets follows an S-shape time path (implying the submarket activation rate having a Bell-shape profile), the model demonstrates substantial power in explaining the aggregate market data, in particular, of the U.S. automotive tire industry, for which we delineate the submarkets in detail in the empirical part (section 5) of the paper.

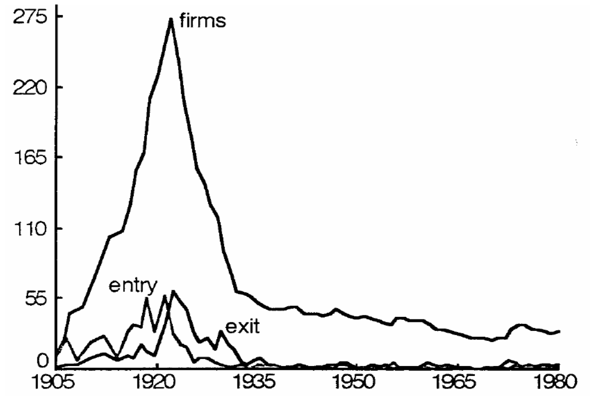

Figure 1: Numbers of entry, exit and firms (Tires, U.S.)
(Reproduced from: Klepper and Simons (2000), Fig. 1)

2 Stylized facts and alternative explanations

Figure 1 shows a striking example of shakeouts, which happened in the U.S. automotive tire industry. This kind of sharp fall in the number of firms earned the phenomenon its name. New evidence allows a disaggregation of the pattern. In Figure 1, the time path of the number of firms can be decomposed as a combination of the time paths of gross entry and gross exit. Two patterns that appear in Figure 1 turn out to hold more generally.

Fact A: For a product that experienced a shakeout, the time path of gross entry roughly had a Bell-shape, which spanned a interval of a few decades, peaked shortly prior to the peak of the number of firms.

Klepper and Miller (1995) examined 16 of the 46 major new products studied by Gort and Klepper (1982). All of the 9 shakeout products in their sample displayed the above described pattern of gross entry. (Also see Agarwal and Gort (1996); Klepper and Simons (2001); Klepper (2002)). As for the pattern of gross exit, Klepper and Miller (1995) found that for all 16 products in their sample, gross exit is positively correlated with lagged gross entry. Using smoothed time series of gross entry and exit, the average correlation between the exit series and the 'optimally' lagged entry series was 0.80 with the average 'optimal' lag in entry being 3 years. This result is affirmative to the findings in the voluminous literature on firm turnovers using data of more aggregate categories of manufacturing industries (e.g., Dunne, Roberts and Samuelson (1988); see Caves (1998) for a comprehensive survey). The latter literature documents a rich body of evidence of positive correlation between gross exit and lagged gross entry; and moreover it closely relates this regularity to another important finding pertaining to firm survival rate over age, that is: "Entrants suffer from high rates of infant mortality" (Caves (1998)). The recent studies of firm survival pattern in the (more narrowly defined) industries that experienced severe shakeouts (see Klepper and Simons (2000), Klepper and Simons (2001); Klepper (2002)) confirm this general finding.

Fact B: There is a high rate of 'infant mortality' among entrants, i.e., a large fraction of new entrants exit at an early age, e.g., more than half drop out within 5 to 10 years; the exit rate gradually eases off after the early years.

In his survey on "What do we know about entry?", Geroski (1995) commented: "[E]ntry appears to be relatively easy, but survival is not. The most palpable consequence of entry is exit, and industries that exhibit high entry rates often also exhibit a high degree of churn at the bottom of the size distribution." His conclusion was: "If one accepts the proposition that the barriers to entry facing small entrants are generally rather modest, then these observations suggest the existence of substantial 'barriers to survival' of some type." The notion of 'barriers to survival' seems particularly relevant to the study of the shakeouts. One implication of fact A above is that there is a period of time when the industry experiences mass entry, this is around the peak of the Bell-shape. Then fact B ('barriers to survival') naturally implies that there is a period of mass exit (that follows the mass entry). This is shakeout (in the classical interpretation of the term).

The above two facts basically challenge any theories of shakeouts to address the following two questions concerning entry and exit: (a) "Why does entry spread out over time, particularly not concentrate on the beginning of the industry?" (b) "What constitute the barriers to survival?" The theory developed in this paper can explain both facts A and B. First, by formulating an S-shape diffusion curve across submarkets, the model predicts that the time path of aggregate market entry follows a Bell-shape. Therefore entry is unlikely to concentrate on the beginning of the industry. Second, the model proposes that some time after entry each firm in a submarket will have to face the requirement of a sunk investment (of endogenously determined amount), which is necessary for increasing the product quality[5] and (hence) remaining competitive in the market. Equilibrium analysis dictates that it would not be profitable for the marginal firm to stay should all existing firms stay, therefore predicts an occurrence of exit, a shakeout, in each submarket. The theory therefore attributes the barriers to survival to the endogenous sunk costs.[6]

An alternative theoretically well-founded explanation for contemporaneous exit and entry (in the same industry) is the 'passive learning' mechanism originated by Jovanovic (1982). In an industry where firms are heterogeneous in their costs, potential entrants are assumed to know the mean and standard deviation of all firms' costs but not their own mean expectations. Upon incurring a sunk entry cost, an entrant starts to receive a noisy signal about its mean

[5] This can be easily extended to include cost reduction. The essence is to incur an endogenous level of sunk investment to acquire some fixed inputs (as opposed to variable inputs) with effect of shifting demand curve upward or marginal cost curve downward.

[6] For a comprehensive account on the role of endogenous sunk costs in determining market structure, see Sutton (1991) and Sutton (1998).

expectation of cost level. An incumbent firm updates its belief about its mean expectation whenever receiving a new signal. Therefore it is possible that some incumbent firms exit as their believes drop below a certain threshold level at the same time as new entrants enter the market with more optimistic believes. The 'passive learning' mechanism thus constitutes a building block which can be used in a model to explain fact B. In this type of models barriers to survival come from the downward adjustments of believes held by some firms about their (innate) efficiency through Bayesian learning.

Hopenhayn (1993) incorporates an akin 'passive learning' mechanism into a perfectly competitive market setting where the proportional growth rate of demand (at each fixed price level) is assumed to be non-negative but decreases over time, i.e., the demand growth rate (at each fixed price level) has a Bell-shape time path. Under the assumption of increasing marginal cost, the model predicts that the demand increase at each point of time can not be satisfied by incumbent firms, therefore induces entry. Since entry rate reflects demand growth rate therefore it does follow a Bell-shape time-path.[7] Therefore, the Hopenhayn (1993) model can indeed explain both facts A and B. There is some similarity between the current paper and Hopenhayn (1993): both ascribe the continual entry flow to demand growth. A key difference, however, is that in our analysis, the relevant economic market is a submarket as opposed to the aggregate market; and we assume that firms engage in imperfect competition as opposed to perfect competition.

The paper by Horvath, Schivardi and Woywode (2001), which also features a 'passive learning' mechanism, seeks to explain the entry pattern leading to industry shakeouts. It attributes the delay and build up in the Bell-shape entry process (fact A) to an information accumulation process. In their explanation, an initial reluctance in entry reflects some uncertainty about the profitability of the market, and a pessimistic prior belief held by potential entrants. The observation of entry and performance by earlier entrants brings more accurate estimation to potential entrants, encouraging more entry and in turn further improving the accuracy of estimate. As a result, entry may have a modest start, gradually accelerated pace over time and an explosive culmination prior to the shakeout phase.

There are a number of models that emphasize the role of technical changes in explaining shakeout (e.g., Jovanovic and MacDonald (1994), Klepper and Graddy (1990) and Klepper (1996)). What these models have in common is that firms have different success in innovations and consequently differ in their efficiency (competitiveness). Efficiency gaps create barriers to survival for the weakest firms who subsequently exit. (This feature is shared by our model.) As for entry pattern, Jovanovic and MacDonald (1994) attributes each occurrence

[7] Hopenhayn (1993) also studied the conditions under which endogenous demand/output growth could be induced by exogenous continuous cost-reducing technical progress under perfect competition, and how this could result in a Bell-shape time path of entry. For the results to hold in this setting, one needs to impose, among other things, some restrictive structure on the demand function (or more precisely, on demand elasticity as a function of output).

of entry to an instance of exogenous technical shock, e.g., invention or refinement, which creates a new entry opportunity. In Klepper and Graddy (1990) and Klepper (1996), technical changes are treated as a gradual endogenous process, the models are silent about what particular pattern the entry process should follow. Therefore, none of these models provide a natural explanation for fact A.

The plan for the rest of the paper is as follows. Section 3 lays out the model. Section 4 characterizes the equilibrium of each submarket, and derives the firm survival pattern, and the aggregation across submarkets. Section 5 calibrates the model to the data of the U.S. tire industry, in particular, to the delineation of the submarkets, and simulates its dramatic shakeout. Section 6 concludes.

3 The model

The model is cast in discrete time. The industry at time $t = 0$ is a collection of b initially inactive independent submarkets with identical demand structure. The only link among submarkets is an exogenous aggregate diffusion process that determines when a particular submarket becomes active and when production starts (as a function of how many submarkets are already in operation). Correspondingly there is an aggregate timing relative to the industry as a whole and the timing in each particular submarket. We proceed to analyze the aggregate timing by describing the aggregate diffusion process in section 3.1. In section 3.2 the timing of a particular submarket is analyzed. In section 3.3 the demand and cost structures, and some related algebra in a submarket is introduced.

3.1 Aggregate diffusion

The aggregate diffusion process is regulated by the following dynamic law of motion:

$$\begin{aligned} k_{t+1} - k_t &= a k_t^{\theta} \left(b - k_t \right) \text{ for } t \geq 1, \\ 0 &< k_1 \leq 1 \end{aligned} \tag{1}$$

which is a generalized logistic equation[8][9], where k_t is the number of active submarkets at time t. Parameter $b > 1$ is the total number of submarkets in the industry[10]. Parameters a and θ affect the rate of diffusion.

[8] Such equation can be derived from some more basic principles.

[9] The logistic equation originated by Verhulst (1838) is a common model of population growth, which has found applications in a range of fields, from biology to economics. This and the related logistic function (S-curve) have been widely used in the studies of diffusion of innovations (for example, see Griliches (1957); Rogers (1962). In this paper we model the life cycle of a new industry as the process of diffusion of a new product invention. The analysis is based on the novel assumption that the basic unit of the diffusion process is an 'approximately' independent submarket.

[10] To ensure $k_t \leq b$, one can use the following modified formula:

$$k_{t+1} - k_t = \min \left\{ a k_t^{\theta} \left(b - k_t \right), b - k_t \right\}.$$

Figure 2 is illustrative of some general features of the diffusion curve (S-shape): (i) it is initially convex (approximately exponential) up to some point, then (ii) it becomes concave, and finally (iii) it becomes flat. Accordingly, the rate of diffusion, as depicted in Figure 3, initially increases up to a peak, then declines, finally converges to zero (Bell-shape). Since the rate of submarket activation follows a Bell-shape curve, this provides an explanation for why the time path of industry-wide entry has a Bell-shape (fact A).

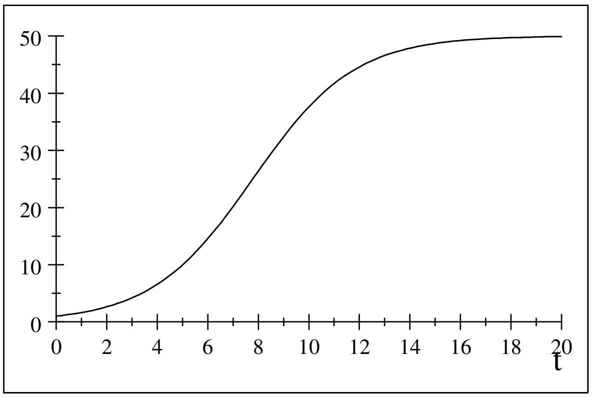

Figure 2: k_t : S-shape

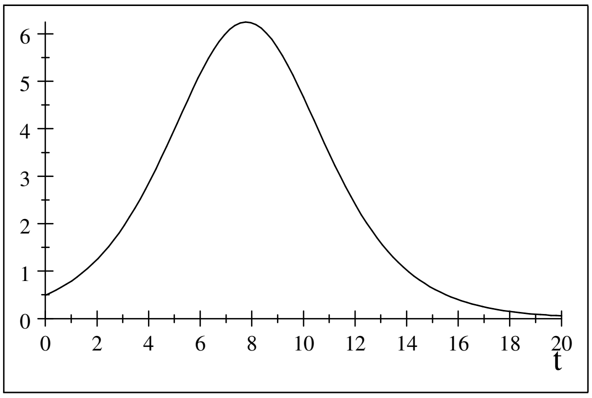

Figure 3: $k_{t+1} - k_t$: Bell-shape

3.2 Timing of a submarket

The complete timing in a submarket is: (i) first a new product is exogenously invented in submarket j (in this case submarket j is the first market to be activated in the industry) or it is introduced from a different submarket into submarket j by the exogenous diffusion process. Firms in submarket j have to decide simultaneously if entering production of such new good: in this case a forward looking free entry condition applies. This is time $t_j = 0$ for submarket j. (ii) After entry, all firms initially play a symmetric Cournot game of quantity competition. This is from time $t_j = 0$ to time $t_j = T_j$ for submarket j. (iii) At some random point of time T_j (determined by the outcome of a random search process) an innovation arrives in the submarket and presents all existing firm with the option to incur an endogenous sunk cost to increase the quality of their products. The firms' decisions on whether making the sunk investments or not are in fact determining if staying or exiting: a second zero profit condition applies for them and a shakeout might take place. (iv) For times $t_j \geq T_j$ the remaining firms play a repeated static Cournot game with differentiated quality. The game in a submarket is solved with backward induction starting from time T_j. The timing of events in a submarket here is similar to Jovanovic and MacDonald (1994). A key difference though is the endogenous sunk cost component which is explained in section 3.3.

3.3 Demand, innovation and sunk costs

To economize on notions, all the subscripts (submarket labels) to endogenous variables are omitted (unless likely to cause confusion). Also all common parameter notations and values are shared across submarkets.

In submarket j, there are a number S of identical consumers. Hence parameter S measures the size of the submarket. The product is vertically differentiable. Each consumer's preference is given by the following static quasi-linear utility function:

$$\begin{aligned} U &= \sum_{i=1}^{N} u_i q_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{l=1}^{N} q_i q_l + y \\ s.t. &: \sum_{i=1}^{N} p_i q_i + y \leq I \end{aligned}$$

where N is the number of active firms that offer a positive quantity of vertically differentiable product, u_i and q_i are the quality and quantity supplied by firm i, y is the numeraire, I is the income.

The first order necessary condition of the above maximization program implies the following inverse demand function

$$p_i = u_i - \sum_{l=1}^{N} q_l. \tag{2}$$

Each firm has a production technology which has constant returns to scale, and the marginal cost is normalized to 0 for all firms. When the number of firms and their respective levels of quality are fixed, all active firms compete *a la* Cournot.[11] Hence, each firm's objective is given by the following maximization program:

$$\max \Pi_i = S \left(u_i - \sum_{l=1}^{N} q_l \right) q_i.$$

The implied reaction function is

$$q_i = \frac{u_i - \sum_{l \neq i} q_l}{2} \text{ for } i = 1, 2, \cdots, N. \tag{3}$$

Summing up the above family of equations and solving for $\sum_{i=1}^{N} q_i$ yields

$$\sum_{i=1}^{N} q_i = \frac{1}{(N+1)} \sum_{i=1}^{N} u_i, \tag{4}$$

Inserting eq. (4) into eq. (3) and (2) gives us the equilibrium output, price and market share at individual firm level as follows

$$q_i = p_i = u_i - \frac{1}{(N+1)} \sum_{l=1}^{N} u_l, \tag{5}$$

$$m_i \left(u_i; u_{-i} \right) = \frac{u_i + \sum_{l=1}^{N} \left(u_i - u_l \right)}{\sum_{l=1}^{N} u_l}. \tag{6}$$

The implied equilibrium profit of each active firm in each period is given by

$$\Pi_i \left(u_i; u_{-i} \right) = S\pi_i \left(u_i; u_{-i} \right) = \frac{S}{(N+1)^2} \left(u_i + \sum_{l=1}^{N} \left(u_i - u_l \right) \right)^2, \tag{7}$$

conditional on all firms: $i = 1, 2, \cdots, N$ being active; where $u_{-i} = \left(u_l \right)_{N-1}$ for $l \neq i$ is the vector of all rivals' levels of quality; $\pi_i \left(u_i; u_{-i} \right)$ is the normalized profit function (for unit market size). A firm's profit flow is discounted by the factor $\frac{1}{1+r}$, where r is the constant interest rate.

At time $t_j = 0$, if a firm enters submarket j, it incurs an sunk entry cost $\sigma > 0$ and obtains the basic level of quality $u = 1$. Meanwhile the submarket (as a whole) also enters a random search for an innovation. In each period, with hazard rate $\rho \in (0, 1)$ the innovation may arrive. There is only one innovation in each submarket. The random searches are i.i.d. across periods. It repeats in each new period until the innovation arrives. This is time $t_j = T_j$. From time $t_j = 0$ to time $t_j = T_j$ all firms in submarket j keep the basic quality level $u = 1$. At time $t_j = T_j$, each firm is given the option to incur a sunk cost $F \left(u_i \right)$ to

[11] For reference on the Cournot game with differentiated quality, see Sutton (1991) and Sutton (1998).

increase quality level to $u_i > 1$. All the firms need to decide simultaneously on their levels of investment (and product quality). If investment occurs, quality level u_i remains for time $t_j \geq T_j$. If no investment is made, a firm's product quality is kept at $u = 1$. The fixed cost (as opposed to marginal cost) function $F(u_i)$ is given by

$$F(u_i) = \sigma u_i^{\beta}, \text{ for } u_i > 1 \tag{8}$$

where $\beta > 2$ is the elasticity of the fixed cost with respect to quality, a measurement of cost (in)effectiveness of the investment[12] in raising quality level.

4 Equilibrium analysis

The important parameters of submarkets are market size relative to entry cost S/σ, and the elasticity of sunk cost w.r.t. quality β. We are interested in the range of these parameters that enables shakeouts. Intuitively, dramatic shakeouts require substantial initial entry, which can only happen if the market size is sufficiently large relative to entry cost. Also shakeouts depend on the product quality difference between investing (surviving) and non-investing (exiting) firms being sufficiently large. This occurs if market size is large because larger market size rationalizes larger endogenous sunk cost and higher quality level. For these two reasons, we assume that the size of each submarket relative to entry cost S/σ is sufficiently large. Quantitatively, we make the following two assumptions:

Assumption 1

$$\frac{S}{\sigma} > (1+r)\left(\frac{\beta}{2}+1\right)^2. \tag{A1}$$

Assumption 2

$$\frac{S}{\sigma} > r\left(\frac{\beta}{2}\right)^2 2^{\beta}. \tag{A2}$$

For a given value of S/σ both assumption 1 and 2 put some (moderate) restriction on the value of β. They request that the sunk investment should be sufficiently cost effective in raising quality level.

The solution concept used in this paper is sub-game perfect Nash equilibrium. For tractability, the analysis is confined to symmetric equilibrium, where all identical active firms in a submarket play identical strategies. For convenience, we also abstract from the integer effect by allowing the numbers of firms to be real numbers, therefore forward looking zero-profit conditions apply to the marginal firms. Using backward induction, we start from the time $t_j = T_j$ subgame.

[12] This can be interpreted as R&D, advertising or other (fixed) marketing expenditures.

4.1 Time $t_j = T_j$

Denote by N_1 the number of existing firms in submarket j from time $t_j = 0$ to $t_j = T_j$, and denote by N_2 the number of firms in the symmetric equilibrium of time $t_j = T_j$ subgame. We proceed by assuming $N_2 < N_1$ so that there exists a marginal firm which is indifferent between 'stay' and 'exit'. (The justification of this assumption will be provided in sections 4.2 and 4.3. See Lemma 2 and Proposition 1.) A representative (surviving) firm i's payoff function is given by

$$\frac{S}{r\left(N_2+1\right)^2}\left(u_i+\sum_{l=1}^{N_2}\left(u_i-u_l\right)\right)^2-\sigma u_i^{\beta}. \tag{9}$$

We characterize the symmetric equilibrium outcome by deriving the following two conditions. The first is the first order condition for maximizing (9) in a symmetric outcome with $u_i = \bar{u}$ for $i = 1, 2, \cdots, N_2$:

$$\frac{2N_2 S}{r\left(N_2+1\right)^2}\bar{u}-\beta\sigma\bar{u}^{\beta-1}=0. \tag{10}$$

The second is a zero-profit condition:

$$\frac{S}{r\left(N_2+1\right)^2}\bar{u}^2-\sigma\bar{u}^{\beta}=0. \tag{11}$$

It is easy to show these two conditions dictate the equilibrium value of N_2 as

$$N_2=\frac{\beta}{2}. \tag{12}$$

From (11) it follows that

$$\bar{u}=\left(\frac{S}{r\sigma\left(\frac{\beta}{2}+1\right)^2}\right)^{\frac{1}{\beta-2}}. \tag{13}$$

4.2 Time $t_j = 0$

Given the fact that the payoff to each firm in playing any time $t_j = T_j$ subgame is zero, the reduced form time $t_j = 0$ game is the following. Firms in submarket j need to decide if to pay the entry cost σ and produce or not. If a firm enters, its net profit per period will be determined by eq. (7) and the game will last until the random point of time $t_j = T_j$.

In a symmetric equilibrium, conditional on $T_j = \tau$, a firm's expected present value (at time $t_j = 0$) of net profit flow is given by

$$\frac{S}{\left(N_1+1\right)^2}\sum_{i=1}^{\tau}\left(1+r\right)^{-i}-\sigma, \tag{14}$$

with probability for $T_j = \tau$ being:

$$\Pr(T_j = \tau) = (1-\rho)^{(\tau-1)}\rho. \tag{15}$$

The implied zero-profit condition then is

$$\sum_{\tau=1}^{\infty} \frac{S}{(N_1+1)^2} \sum_{i=1}^{\tau} (1+r)^{-i} (1-\rho)^{(\tau-1)} \rho - \sigma = 0, \tag{16}$$

which can be simplified to

$$\frac{S}{(\rho+r)(N_1+1)^2} - \sigma = 0.$$

The implied number of firms is

$$N_1 = \sqrt{\frac{S}{(\rho+r)\sigma}} - 1. \tag{17}$$

It is trivial to show:

Lemma 1 *N_1 increases in S, decreases in σ, ρ and r.*

The next lemma compares the numbers of firms and levels of quality prior to and after the innovation.

Lemma 2 *Under Assumptions 1 and 2, $N_1 > N_2$ and $\bar{u} > 1$.*

Proof. From eq. (12) and (17) it is immediate that

$$N_1 > N_2 \text{ if } \frac{S}{\sigma} > (\rho+r)\left(\frac{\beta}{2}+1\right)^2.$$

The condition holds under Assumption 1. Similarly, from (13) it follows that

$$\bar{u} > 1 \text{ if } \frac{S}{\sigma} > r\left(\frac{\beta}{2}+1\right)^2;$$

the condition also holds under Assumption 1. ■

4.3 Submarket shakeout

In this section we begin by showing that after the innovation arrives in a submarket, the investing firms' quality increase is so large that the non-investing firms' market share reduces to zero, i.e., they exit. This causes a submarket shakeout.

Proposition 1 *Under Assumption 2, the $(N_1 - N_2)$ non-investing firms' market share is zero.*

Proof. The market share of a non-investing firm (with $u = 1$) when playing a static Cournot game with N_2 investing firms (with quality level $u = \bar{u}$) (according to (6)) is given by

$$m_i \left(1; (\bar{u})_{N_2}\right) = \frac{1 + N_2 (1 - \bar{u})}{1 + N_2 \bar{u}},$$

which is non-positive if

$$\bar{u} \geq \frac{N_2 + 1}{N_2}. \tag{18}$$

Substituting eq. (12) and (13) into the above inequality and reorganizing yield the following inequality:

$$\frac{S}{\sigma} \geq r \frac{\left(\frac{\beta}{2} + 1\right)^{\beta}}{\left(\frac{\beta}{2}\right)^{\beta - 2}}. \tag{19}$$

Under Assumption 2 we have

$$\frac{S}{\sigma} > r \left(\frac{\beta}{2}\right)^2 2^{\beta}.$$

Given that $\beta > 2$, then it is easy to show that inequality (19) holds. ■

The timing of the shakeout in submarket j is given by the random variable T_j, with probability distribution (see eq. (15)):

$$\Pr (T_j = \tau) = (1 - \rho)^{(\tau - 1)} \rho \text{ for } \tau = 1, 2, \cdots, \infty.$$

Thus the expected timing of a shakeout in submarket j is the mean of T_j:

$$\bar{T}_j = \sum_{\tau=1}^{\infty} \tau (1 - \rho)^{(\tau - 1)} \rho, \tag{20}$$

implying the following proposition.

Proposition 2 *The expected timing of a shakeout in submarket j is given by*

$$\bar{T}_j = \frac{1}{\rho}. \tag{21}$$

Proof. Given eq. (20), the result follows immediately from the algebra:

$$\bar{T}_j = \sum_{\tau=1}^{\infty} \tau (1 - \rho)^{(\tau - 1)} \rho = -\rho \frac{\partial}{\partial \rho} \left(\sum_{\tau=1}^{\infty} (1 - \rho)^{\tau} \right) = -\rho \frac{\partial}{\partial \rho} \left(\frac{1 - \rho}{\rho} \right) = \frac{1}{\rho}.$$

■

4.4 Firm dynamics

Now we look at a firm's prospect in submarket j on equilibrium path. At time $t_j = 0$, the firm enters the submarket as a low-quality firm, and will stay as a low-quality firm from time $t_j = 0$ to time $t_j = T_j$. It is going to experience a submarket shakeout arrived randomly at time $t_j = T_j$. In that case with probability $\frac{N_2}{N_1}$ the firm will become a high-quality firm and stay active forever after; with probability $1 - \frac{N_2}{N_1}$ the firm will exit the submarket and stay out forever.

The probability distribution of the timing of submarket shakeout described by eq. (15) implies the following proposition:

Proposition 3 *A firm's probability of surviving the age of s is*

$$\Pr(life > s) = 1 - \frac{N_1 - N_2}{N_1} \sum_{\tau=1}^{s} (1-\rho)^{(\tau-1)} \rho, \tag{22}$$

which decreases with age s. Or, equivalently, A firm's hazard rate of exit by age s is

$$\Pr(life \leq s) = \frac{N_1 - N_2}{N_1} \left(1 - (1-\rho)^s\right), \tag{23}$$

which increases with age s.

Proposition 3 provides an explanation for why entrants suffer 'high infant mortality rates' (fact B). See below.

Corollary 3 *If market size S and the hazard rate of innovation ρ are sufficiently large, a firm's hazard rate of exit by the age $s = 2$ can exceed any given threshold value z, i.e., $\Pr(life \leq s) \geq z$ for $0 < z < 1$.*

Proof. It suffices to prove that $\Pr(life \leq 2) \to 1$ as $S \to \infty$ and $\rho \to 1$. Eq. (23) determines that

$$\Pr(life \leq 2) = \frac{N_1 - N_2}{N_1} \left(1 - (1-\rho)^2\right).$$

From Proposition 1 it follows that $N_1 \to \infty$ and hence $\frac{N_1 - N_2}{N_1} \to 1$ as $S \to \infty$. Also, $\left(1 - (1-\rho)^2\right) \to 1$ as $\rho \to 1$. Then it is immediate that

$$\Pr(life \leq 2) \to 1 \text{ as } S \to \infty \text{ and } \rho \to 1.$$

■

4.5 Aggregation across submarkets

In this section we derive the expected values of industry-wide aggregate variables. The approach is to aggregate across submarkets and across firms within each submarket. To aggregate across submarkets we consider one representative firm in each submarket. Since we know each (representative) firm's stochastic dynamics we can aggregate (and take expectation) across the representative firms. To (further) aggregate across firms within each submarket, we can simply scale the aggregation across the representative firms by a factor of N_1. The following are a list of aggregate variables and their definitions:[13]

Gross entry G_t:

$$G_t \triangleq N_1 \left(k_t - k_{t-1}\right), \tag{24}$$

where k_t is the number of active submarkets at time t (see eq. (1));

Gross exit X_t:

$$X_t \triangleq \frac{N_1 - N_2}{N_1} \sum_{i=1}^{t-1} \left(G_i \left(1-\rho\right)^{(t-1-i)} \rho \right); \tag{25}$$

Net entry E_t:

$$E_t \triangleq G_t - X_t; \tag{26}$$

Number of firms C_t:

$$C_t \triangleq \sum_{i=1}^{t} E_i; \tag{27}$$

Number of high-quality firms H_t:

$$H_t \triangleq \frac{N_2}{N_1} \sum_{i=1}^{t} \left(G_i \left(1 - \left(1-\rho\right)^{t-i}\right)\right); \tag{28}$$

Number of low-quality firms L_t:

$$L_t \triangleq \sum_{i=1}^{t} \left(G_i \left(1-\rho\right)^{t-i} \right). \tag{29}$$

From eq. (5) it follows that a low-quality firm's quantity and price are given by:

$$q_L = p_L = \frac{1}{N_1 + 1}, \tag{30}$$

and a high-quality firm's quantity and price are:

$$q_H = p_H = \frac{\bar{u}}{N_2 + 1}, \tag{31}$$

where $\bar{u}$ is determined by eq. (13). It is trivial to show that $q_H > q_L$. To make prices comparable, it is helpful to use quality-adjusted price for high-quality product, as formulated below:

$$p_H - \left(\bar{u} - 1\right) = \frac{1 - N_2 \left(\bar{u} - 1\right)}{N_2 + 1}. \tag{32}$$

[13] All the following aggregate variable names should start with the word "expected" and end with the phrase "at time t", which are omitted for simplicity.

We can now look back to provide an intuitive explanation for non-investing firms' exit after the innovation. It is easy to show that if $\bar{u} > \frac{N_2+1}{N_2}$ (see condition (18)), then

$$p_H - (\bar{u} - 1) < 0,$$

i.e., the quality-adjusted price at which the high-quality firms compete is negative, and hence below the marginal cost of the low-quality firms (which is zero).

For completeness we define two additional aggregate variables as follow:

Industry-wide total output:

$$\begin{aligned} Q_t &= L_t q_L + H_t q_H \\ &= L_t \frac{1}{N_1 + 1} + H_t \frac{\bar{u}}{N_2 + 1}; \end{aligned} \tag{33}$$

Industry-wide (output-weighted) average quality-adjusted price:

$$\begin{aligned} P_t &= \frac{L_t q_L}{L_t q_L + H_t q_H} p_L + \frac{H_t q_H}{L_t q_L + H_t q_H} (p_H - (\bar{u} - 1)) \\ &= p_L + \frac{1}{1 + \frac{q_L}{q_H} \frac{L_t}{H_t}} (p_H - (\bar{u} - 1) - p_L) \\ &= \frac{1}{N_1 + 1} - \frac{\frac{N_2(\bar{u}-1)-1}{N_2+1} + \frac{1}{N_1+1}}{1 + \frac{\frac{\bar{u}}{N_2+1}}{N_1+1} \frac{L_t}{H_t}}. \end{aligned} \tag{34}$$

5 Calibration and simulation

In what follows we calibrate the model to simulate a dramatic instance of industry shakeouts. Particularly, we simulate the time paths of entry, exit and the number of firms at the industry level; the time path of each entry cohort population; and the firm survival rate over age. The model is calibrated to (the particular moments of) the data of the U.S. automotive tire industry for the period 1905-1980 (drawn from French (1991); Klepper and Simons (2000), Klepper and Simons (2001); Klepper (2002)). We proceed to define the submarkets of the U.S. automotive tire industry in section 5.1. In section 5.2 the model is calibrated treating submarkets as perfectly independent. Given the fact that (empirically speaking) independence between submarkets is only an approximation, in section 5.3 the model is reinterpreted and re-calibrated to take into account economies of scope that limit the independence between submarkets.

5.1 Defining submarkets of the U.S. tire industry

A first division of the market for the U.S. automotive tire manufacturers was between the original equipment (OE) market which accounted for about 30 percent of all tires sold and the replacement market where the rest 70 percent was sold. The buyers in the OE market were the automotive manufacturers (e.g., General Motors and Ford) and the tires were sold as parts of the new automotive vehicles. The end buyers in the replacement market were automotive vehicle owners, who bought the tires (and the fitting services) from local retail tire suppliers. The service between manufacturing and retail was wholesale distribution. Tire manufacturers were highly involved in tire marketing. In the replacement market, manufacturers typically invested in advertisement which mainly targeted the end buyers. According to French (1991), since the early history of the U.S. automotive tires industry it had been common that tire manufacturers vertically integrated into the tire wholesale sector. Leading manufacturers started investing in (wholesale) branch stores as early as 1909. Manufacturers also developed their relationship with retailers through dealer programs. Exclusive territories were allocated to dealers according to the size of local markets; advertising material and marketing advice were provided.[14]

There were different types of tires designed for different types of vehicles and different driving conditions. Product differentiation limited (demand-side) substitution and segmented the tires market into several product submarkets. Roughly the market was divided among passenger (and light truck) tires, medium and heavy truck (and bus) tires, and miscellaneous speciality tires (agricultural, industrial, mining, military, aviation, racing, etc.). Tires in the same category might also come in different sizes. The passenger tires market was the biggest product submarket.

Transportation cost reduced the substitutability between tires made available from distant locations. By investing in geographical distribution capacities (e.g., wholesale warehouses), a tire manufacturer could reduce the marginal transportation and logistic costs, and (conditional on sufficiently large quantity of output) could save on average transportation and logistic costs. Despite the fact that tire manufacturers had their factories disproportionately located around Akron, Ohio, it was evident that without adequate investment in distribution capacity in a particular geographical submarket, *ceteris paribus* a firm could not compete in that submarket *vis-a-vis* rivals that did invest adequately. That was one of the reasons why the leading tire manufacturers found it important to invest in (wholesale) branch stores.

Given the tremendous geographical size of the U.S., the transportation and logistic costs factor segmented the replacement passenger tires market into many geographical submarkets. A natural candidate definition of the geographical submarkets roughly matched the 50 states. We consider a state X contained (at least) one geographical submarket of replacement passenger tires if (i) X's land area exceeded $31,400$ square miles[15]; or (ii) X contained at least one entire metropolitan area[16]. According to this criterion, the only two states did not qualify were Delaware and Vermont. In contrast, Texas and California were the two states which had land areas at least four times as large as the threshold

[14] Retail tire dealers (and wholesale tire deals when applicable) could be seen as agents who worked for the tire manufacturers, to undertake part of their tire marketing tasks. Product quality, availability, and price were among the main factors that affected the dealers' loyalty, which the manufacturers competed to maintain.

[15] This is the area of an circle with radius of 100 miles.

[16] The 1950 (the earliest available) official definition of standard metropolitan areas is used.

value, and had multiple entire metropolitan areas. We propose that each of these two states had (at least) two geographical submarkets. Overall, it is safe to say there were (at least) 50 submarkets of replacement passenger tires in the U.S.. The number of geographical submarkets of OE (original equipment) tires were very limited because the automotive manufacturers (the buyers) were concentrated in one location - Detroit. Similarly, the geographical extents of other product submarkets of replacement tires could be different (from the replacement passenger tire submarkets), therefore the number of geographical submarkets thereof could be different (from the replacement passenger tire submarkets), e.g., smaller.

The definition of submarkets relies on applying the idea of "marked gaps in the chain of substitutes" (Robinson (1954)). As part of the market definition exercise one (also) needs to examine the degree of supply-side substitution (in addition to demand-side substitution), i.e., the possibility and readiness for firms that are not currently producing a product, to switch to supplying the product if its price is high enough. For instance, it would be easy for a firm that produces passenger tire of a particular size to switch to producing the same type of tire of another size. Also, it would be easier for a passenger tire producer to become a light truck tire supplier than to become a heavy truck tire seller. Here the same principle applies: close substitutes should be considered belonging to the same submarkets while distant substitutes should be categorized to separate submarkets. A "marked gap" (in the supply-side chain of substitution) exists if there is immobility (of some essential fixed inputs) across the submarket boundaries. To put it in another way, a defining feature of a submarket is its specific sunk investment requirement to entrants. In the example of the tire industry, this investment could be in warehouse and sales office at a particular geographical region or location; local marketing/advertising expenditure; securing (essential) local supply of inputs; R&D on specific tires (design, test, etc.); molds and equipment designed for producing specific tires. Inadequacy in such sunk investments would deprive a firm the ability to compete effectively in the particular submarket in question.

Perfectly independent submarkets are those that share neither demand-side linkage nor supply-side linkage between them. Perfectly independent submarkets do not exist in the real world, where submarkets are at best approximately independent from each other. The most important omission from a model of perfectly independent submarkets is economies of scope. Panzar and Willig (1981) state that economies of scope exist if there is spare capacity in an input of producing a product, which can be used to produce another product. A *sharable*, "quasi-public" input is thereby particularly relevant. Spare capacity of this sort is to what Panzar and Willig (1981) attribute the origin of multiple-product firms.[17] In the U.S. automotive tire industry, some fixed inputs had this "quasi-public" input property. For example, certain knowledge and know-

[17] Some extent of incompleteness or friction of contracting over the use of the spare capacity must have an implicit role in the argument. Otherwise, the firm that owns the spare capacity should be indifferent between leasing it to another firm that produces the other product and producing both products.

how of rubber chemistry and tire building, once acquired, could be applied to producing additional types of tires. Also, excess production capacities, once built, could be used to produce additional types of tires, or supply additional geographical submarkets. And indeed all the major tire manufactures[18] were multiple-product firms, and they all operated in multiple geographical submarkets.

5.2 Calibration I

We set the calendar years 1904-1906 as time $t = 1$ of the model (one time period of the model consists of 3 calendar years), therefore years 1961-1963 is time $t = 20$ (which is the last period of our simulation[19]). The number of entry (exit) per period then corresponds to the three-year sum of annual entries (exits).

We set the total number of submarkets to $b = 50$.[20] From Klepper (2002) we can infer that the total number of entry in the tire industry during 1905-1980 was 616, and the number of firms stabilized around 30 near the end of the sample period. We therefore calibrate the model to generate the total number entry $bN_1 = 616$. It follows from $b = 50$ that $N_1 = 12.32$. To make bN_2 to approach 30 (the number of firms at the end of the sample period), we set $N_2 = 1.1$ and hence $bN_2 = 55$.

We normalize parameter σ to 1 and set the interest rate at $r = 0.05$. By choosing $\beta = 2.2$ we have $N_2 = 1.1$. We choose the values of parameters $\rho = 0.59209$, $a = 0.0257$, $\theta = 0.96$ and $k_1 = 0.7$ such that the simulated gross entry G_t peaks at time $t = 7$ with $G_7 = 57$, and the simulated number of firms C_t peaks at time $t = 7$ with $C_7 = 275$. These simulated values match the data (peak entry of 57 and peak firm number of 275 in 1922, see Klepper and Simons (2000)).

Equation (1) determines the aggregate diffusion process. The calibrated generalized logistic equation is the following:

$$\begin{aligned} k_{t+1} - k_t &= 0.0257 k_t^{0.96} (50 - k_t), \\ k_1 &= 0.7. \end{aligned}$$

The implied firm hazard rate of exit by the age of 2 (i.e., 6 years old) is

$$\Pr(life \leq 2) = \frac{N_1 - N_2}{N_1} \left(1 - (1-\rho)^2\right) = 0.759\,18,$$

[18] The largest four U.S. tire manufacturers during 1926-1980 were Goodyear, Firestone, B.F. Goodrich, U.S. Rubber (later Uniroyal).

[19] What happened between 1964 and 1980 did not have any dramatic impact on entry, exit and the number of firms.

[20] The crude submarket definition exercise in section 5.1 suggests that the number of submarkets in the U.S. tire industry could significantly exceed 50. Given the fact that economies of scope could limit the independence between submarkets, here we deliberately admit a smaller number of submarkets so that the submarkets whose entry process was heavily affected by economies of scope are excluded. In so doing, the assumption of independent submarkets remains a reasonable approximation.

which clearly suggests a high rate of 'infant mortality' among new entrants.

From eq. (17) we can derive

$$S = (N_1 + 1)^2 (\rho + r) \sigma = 113.92.$$

It is straight forward to verify that this parameter of market size satisfies both Assumption 1 and 2.

Tables 1 and 2 summarize all the specified parameter values and the implied variable values.

Table 1: Calibration I

Parameter:	a	b	θ	k_1	σ	r	β	ρ	S
Value:	0.0257	50	0.96	0.7	1	0.05	2.2	0.59209	113.92

Table 2: The implied values of variables (I)

Variable:	N_1	N_2	$\Pr(life \leq 2)$
Value:	12.32	1.1	0.759 18

The simulation results, which are presented in Figures 4-6, illustrate the model's ability to account for the empirical regularities (facts A and B) associated with industry shakeouts. Figure 4 shows the simulated time paths of entry, exit and the number of firms (in the thin lines), together with the actual time series (in the thick lines). The simulated patterns capture some of key empirical features displayed in the actual time series. The simulated time path of entry, which reflects the opening of submarkets, has a Bell-shape. This shows that the model provides a natural explanation for fact A. A positive correlation exists between the simulated time path of exit and that of lagged entry, which is driven by the pattern of shakeout within each submarket. The link between the two time paths is clearly featured by 'mass exit follows mass entry'. The simulated number of firms rises when entry exceeds exit; it falls when exit exceeds entry. Since the peak of the simulated exit closely follows that of the simulated entry, exit surpasses entry shortly after the peak of the latter. That is why we see entry peaks shortly prior to the peak of the number of firms. The simulated time paths fit the actual time series quantitatively well except that the simulated post-shakeout number of firms is significantly higher than the data.

Figure 5 shows the simulated time paths of entry cohort populations for successive entry cohorts. Each cohort displaces a similar shakeout pattern; however, the initial populations of the entry cohorts vary markedly. The figure suggests that the dramatic rise and fall of the number of firms around the peak are generated by the impacts of a very few big entry cohorts (5, 6, 7 and 8), featuring 'mass entry followed by mass exit'.

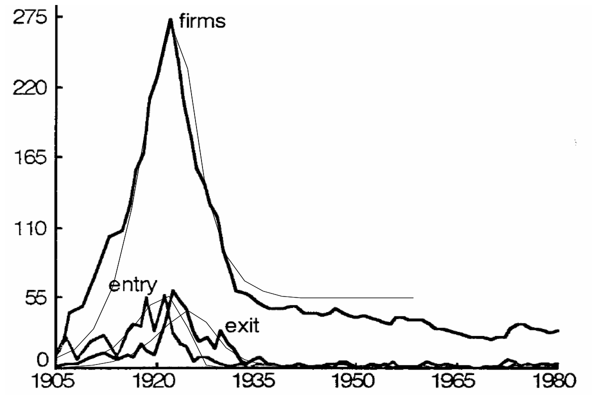

Figure 4: Actual (thick-line) and simulated (thin-line, calibration I) time-paths of entry, exit and number of firms

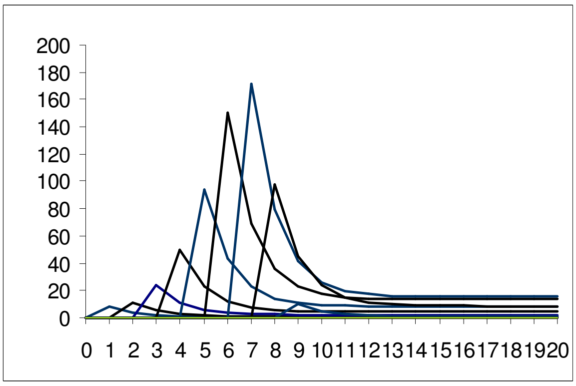

Figure 5: Simulated numbers of firms in different entry cohorts (calibration I)

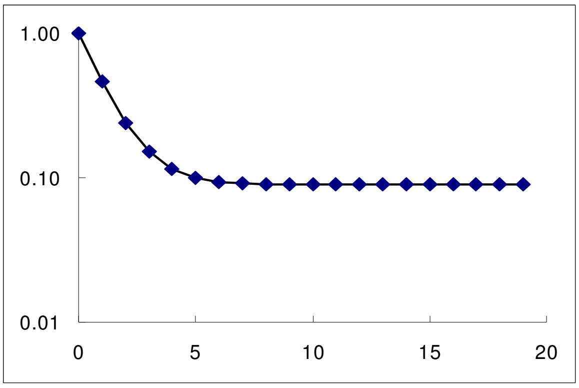

Figure 6: Simulated firm survival rate over age (calibration I)

The pattern of shakeout in each entry cohort implies a high rate of 'infant mortality', i.e., a high rate of early-age exit hazard (fact B). This stands out clearly in Figure 6, which shows the simulated firm survival rates over age.

5.3 Calibration II: introducing economies of scope and multiple-product firms

The model originally assumes that the submarkets are perfectly independent, therefore there are no economies of scope, and multiple-product firms do not possess any advantage *vis-a-vis* single-product firms. As a result, in equilibrium each firm is a single-product firm, i.e., operates in only one submarket.[21] In this section we relax the assumption of perfect independence between submarkets slightly by introducing economies of scope in the following way. Lets define established firms as those that have already had at least one high-quality product. Suppose that each established firm has certain spare capacity in some fixed input(s) (e.g., certain technological knowledge and know-how) that it has acquired for producing existing product(s). The spare capacity can be used in producing a new product and therefore reduces the fixed cost of producing the new product. This gives an established firm an advantage *vis-a-vis* a new firm in 'competing for' (i.e., entering) the new submarket. We assume that the entry cost for an established firm is $\sigma - \varepsilon$ ($\varepsilon > 0$) as opposed to σ for a new firm.

At time $t_j = 0$ for submarket j, among potential entrants, there is a finite number of established firms, and an infinite number of new firms. Since a zero-profit condition applies to the marginal entrant, by ignoring the integer effect, it can be shown that in equilibrium (i) all established firms enter if their number

[21] Although in principle it is possible to have an equilibrium outcome where some firms operate in multiple submarkets, the probability of such an outcome is zero if we assume the number of potential entrants is infinity.

does not exceed the number of entry, and the rest entering firms are all new single-product firms; otherwise, all entering firms are established firms.

To simplify the analysis, we take the limit: $\varepsilon \to 0$, therefore the established firms' advantage of scope economies does not affect the number of entry and exit of product units, but to affect the proportion of entrants that are new single-product firms. As a result, it also affects the proportion of existing product units that are single-product firms (as opposed to subsidiaries of multiple-product firms.)[22][23] The following additional aggregate variables describe the entry, exit and number of (independent) firms (as opposed to subsidiaries of multiple-product firms).[24]

Gross entry of new firms[25] G_t^N:

$$G_t^N \triangleq \max\left\{N_1 - H_t^I, 0\right\}(k_t - k_{t-1}), \tag{35}$$

where H_t^I is the number of established firms (see definition below);

Gross exit of new firms X_t^N:

$$X_t^N \triangleq \frac{N_1 - N_2}{N_1} \sum_{i=1}^{t-1} \left(G_i^N (1-\rho)^{(t-1-i)} \rho \right); \tag{36}$$

Net entry of new firms E_t^N:

$$E_t^N \triangleq G_t^N - X_t^N; \tag{37}$$

Number of (independent) firms C_t^I:

$$C_t^I \triangleq \sum_{i=1}^{t} E_i^N; \tag{38}$$

Number of established firms H_t^I:

$$H_t^I \triangleq \frac{N_2}{N_1} \sum_{i=1}^{t} \left(G_i^N \left(1 - (1-\rho)^{t-i}\right)\right), \tag{39}$$

and

$$H_1^I = 0; \tag{40}$$

Number of low-quality single-product firms L_t^I:

$$L_t^I \triangleq \sum_{i=1}^{t} \left(G_i^N (1-\rho)^{t-i} \right). \tag{41}$$

[22] Note: all new firms are low-quality single-product firms; all multiple-product firms are established firms; but not all established firms are multiple-product firms.

[23] Since the fixed cost function (8) is not affected by the economies of scope, the hazard rate of exit is the same for a new (entering) firm and a new (entering) subsidiary of an established firm.

[24] Again, all the following aggregate variable names should start with the word "expected" and end with the phrase "at time t", which are omitted for simplicity.

[25] A new firm (at time t) is one which has not competed in any submarket of the industry previously.

In what follows we re-calibrate the model. We preserve the values of parameters θ, σ, r and ρ as in the previous section. We double the number of submarkets to $b = 100$.[26] We set $S = 751.01$, $\beta = 4.2$, $a = 0.0105$ and $k_1 = 0.48$ such that the simulated peak number of (independent) firms matches the actual magnitude and timing (275 respectively between $1922 - 1924$), the simulated peak number of entry (of new firms) matches the observed number (57 between $1922 - 1924$), and the simulated number of (independent) firms for $1958 - 1960$ matches the data (about 40). Under the new calibration, the diffusion process follows

$$\begin{aligned} k_{t+1} - k_t &= 0.0105 k_t^{0.96} (100 - k_t) \\ k_1 &= 0.48. \end{aligned}$$

It can be shown that the market size parameter S satisfies Assumptions 1 and 2. The new calibration values are summarized in Tables 3 and 4, and the simulation results are presented in Figures 7 and 8.

Table 3: Calibration II

Parameter:	a	b	θ	k_1	σ	r	β	ρ	S
Value:	0.0105	100	0.96	0.48	1	0.05	4.2	0.59209	751.01

Table 4: The implied values of variables (II)

Variable:	N_1	N_2	$\Pr(life \leq 2)$
Value:	33.2	2.1	0.780 88

When we allow firms to own multiple products there is usually a gap between the number of products and the number of firms. The product-firm discrepancy as shown in Figure 7 is mainly driven by the fact that the number of firms ceases growth while the number of products is still growing. The further increase in the number of products after the peak of the number of firms is primarily attributed to the expansion into new submarkets by established firms (that enjoy an advantage of economies of scope *vis-a-vis* new firms). New firms find it harder and harder to enter new submarkets thereafter.

The simulated time paths of entry, exit and number of firms are presented in Figure 8, which illustrates that the model (under calibration II) can fit the

[26] Since in this section we re-interpret the model to allow economies of scope and multiple-product firms, it is now reasonable to admit submarkets which are "less independent" (i.e., affected by the economies of scope), and therefore to adopt a larger estimation of the number of submarkets than in section 5.2.

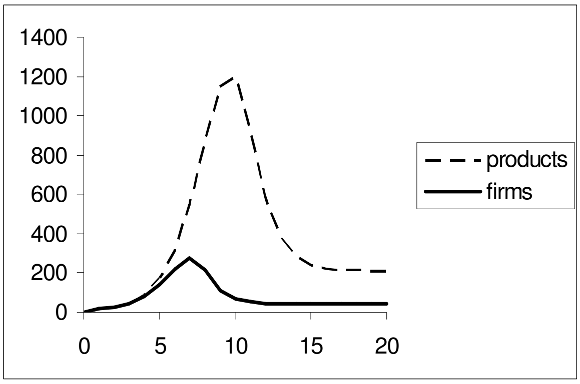

Figure 7: Simulated numbers of firms and products (calibration II)

actual data with remarkable accuracy[27]. Compared with calibration I, this re-calibration significantly improves the fit for the post-shakeout number of firms. Under the new interpretation and re-calibration of the model, the post-shakeout number of firms is less tied to the number of submarkets because the surviving firms can be multiple-product firms. This increased flexibility of the model allows admission of a larger number of submarkets and also larger numbers of competitors in each submarket (both before and after submarket shakeout).

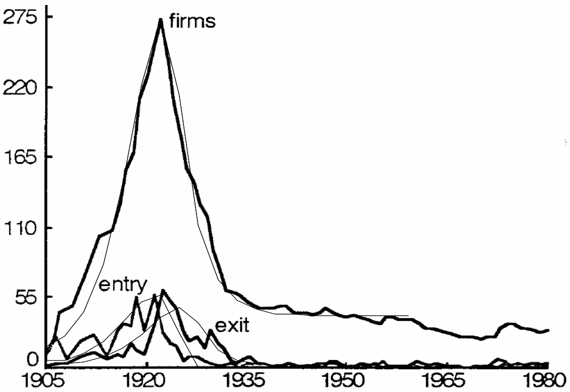

Figure 8: Actual (thick-line) and simulated (thin-line, calibration II) time-paths of entry, exit and number of firms

By taking into account the fact that some firms own multiple products which are sold in (approximately independent) submarkets, our model of successive

[27] Naturally the simulated series are smoother than the actual data because the simulation is based on behavior of the mean which reduces the randomness.

submarkets confirms Sutton (1998)'s important finding that the typical dispersed firms size distribution in an industry can be partially attributed to the fact that firms do vary widely in the number of submarkets in which one operates. This is clearly the case in the illustrative example of the U.S. automotive tire industry.[28]

Our simulation of the time paths of entry, exit and number of firms (based on calibration II) is robust to potential measurement errors in estimating the total number of submarkets b. Suppose b' (in stead of the true value b) is the estimation, hence introducing measurement error $(b' - b)$. Since a and N_1 are unobservable parameter and variable, they are determined by calibrating the model (including b' and other pre-determined variable values) to the moments of the data. The resulting estimates of a and N_1, denoted by a' and N_1', therefore deviate from their true values, and introduce additional errors $(a' - a)$ and $(N_1' - N_1)$. Our robustness check exercise shows that $(a' - a)$ and $(N_1' - N_1)$ have the oppose sign to $(b' - b)$. In simulating the time paths of entry, exit and number of firms, these measurement error and induced deviations approximately offset each other and leave no significant effect on the simulation. This result is reassuring about the model's ability to replicate a dramatic shakeout process. It shows that the explanatory power of the model does not rely on any arbitrary element in the empirical submarket definition exercise.

6 Conclusion

This paper explains contemporaneous exit and entry in a new industry with a diffusion process across submarkets. It allows a re-interpretation of the shakeout process in some industries in a novel way. A new industry is a collection of initially inactive submarkets; the timing of their activation is determined by an exogenous aggregate diffusion process. Submarket-specific sunk costs facilitate approximate independence of submarkets, which allows new firms the opportunities to enter new submarkets. However, the post-entry endogenous sunk investment requirement induced by innovations also forces exit to follow entry closely and to a large extent. When the aggregate diffusion process is described by a plausible generalized logistic equation, and calibrated to the data of the U.S. tire industry, in particular to the delineation of the submarkets, the model can match the aggregate market data remarkably well, and replicate the dramatic shakeout of the U.S. tire industry.

The existing literature on industry shakeouts has emphasized the effect on firm survival (and or growth) of exogenous heterogeneity in firm efficiency (see Klepper and Graddy (1990), Klepper (1996), and Hopenhayn (1993)). The models in this literature shed light on the ways that the exogenous difference in firm efficiency translates (through market competition) into firm size difference. The current paper abstracts from this 'selection' aspect of industry shakeouts by

[28] According to French (1991), the dominant firms offered full range of tires, and they had large numbers of branch stores all over the U.S., while the small tire manufaturers only operated in niche (product) markets, and lacked national distribution.

assuming away any exogenous heterogeneity among firms. In doing so the model acquires the ability to highlight the economic forces that tend to create asymmetry among firms purely as an equilibrium outcome. Such forces identified here are the endogenous sunk investment requirement induced by innovations, and economies of scope shared across submarkets. This finding complements the existing literature in explaining the uneven firm size distribution, which is commonly observed among industries. Having demonstrated the effects of these economic forces in the current setting, it is conceivable that these factors could possibly reenforce and amplify the exogenous asymmetry in firm efficiency and facilitate greater degree of asymmetry in firm size distribution. Among other things, it would be interesting (for future research) to investigate whether firms differ exogenously in their capabilities of investing in innovations, and exploring economies of scope across submarkets; and, if yes, what the implications are on firms' entry, exit and expansion decisions, and on firm size distribution and the evolution of market structure.

References

Agarwal, Rajshree and Michael Gort, "The Evolution of Markets and Entry, Exit and Survival of Firms," *Review of Economics and Statistics*, 1996, *78* (3), 489–498.

Caves, Richard E., "Industrial Organization and New Findings on the Turnover and Mobility of Firms," *Journal of Economic Literature*, 1998, *36*, 1947–1982.

Church, Jeffrey and Roger Ware, *Industrial Organization: A Strategic Approach*, McGraw-Hill, 2000.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *Rand Journal of Economics*, 1988, *19* (4), 495–515.

French, M., *The U.S. Tire Industry: A History*, Twayne Publishers, 1991.

Geroski, P. A., "What Do We Know about Entry?," *International Journal of Industrial Organization*, 1995, *13*, 421–440.

Gort, Michael and Steven Klepper, "Time Paths in the Diffusion of Product Innovations," *Economic Journal*, 1982, *92*, 630–653.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 1957, *25* (4), 501–522.

Hopenhayn, Hugo A., "The Shakeout," 1993. Economics Working Paper 33, Universitat Pompeu Fabra.

Horvath, Michael, Fabiano Schivardi, and Michael Woywode, "On Industry Life-Cycle: Delay, Entry, and Shakeout in Beer Brewing," *International Journal of Industrial Organization*, 2001, *19*, 1023–1053.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 327–338.

___ **and Glenn M. MacDonald**, "The Life Cycle of a Competitive Industry," *Journal of Political Economy*, 1994, *102* (2), 322–347.

Klepper, Steven, "Entry, Exit, Growth, and Innovation Over the Product Life Cycle," *American Economic Review*, 1996, *86* (3), 562–583.

___ , "Firm Survival and the Evolution of Oligopoly," *Rand Journal of Economics*, 2002, *33* (1), 37–61.

___ **and Elizabeth Graddy**, "The Evolution of New Industries and the Determinants of Market Structure," *Rand Journal of Economics*, 1990, *21* (1), 27–44.

___ **and John H. Miller**, "Entry, Exit, and Shakeout in the United States in New Manufactured Products," *International Journal of Industrial Organization*, 1995, *13*, 567–591.

___ **and Kennith L. Simons**, "The Making of an Oligopoly: Firm Survival and Technological Change in the Evolution of U.S. Tire Industry," *Journal of Political Economy*, 2000, *108* (4), 728–760.

___ **and** ___ , "Industry Shakeouts and Technological Change," 2001. Memeo, Carnegie Mellon University.

Panzar, C. and R. Willig, "Economies of Scope," *American Economic Review*, May 1981, *71* (2), 268–272. Papers and Proceedings of the Ninety-Third Annual Meeting of the American Economic Association.

Robinson, J., *The Economics of Imperfect Competition*, London: MacMillan, 1954. 1933.

Rogers, Everett, *Diffusion of Innovations*, New York: The Free Press, 1962.

Schmalensee, R., "Another Look at Market Power," *Harvard Law Review*, 1982, *95*, 1789–1816.

Sutton, John, *Sunk Cost and Market Structure*, Cambridge, Massachusetts: The MIT Press, 1991.

___ , "Gibrat's Legacy," *Journal of Economic Literature*, 1997, *35*, 40–59.

___ , *Technology and Market Structure*, Cambridge, Mass: The MIT Press, 1998.

Verhulst, P. F., "Notice sur la loi que la population pursuit dans son accroissement," *Correspondance mathématique et physique*, 1838, *10*, 113–121.