Strategic Management Journal
Strat. Mgmt. J., **35**: 1569–1584 (2014)
Published online EarlyView 26 November 2013 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2197
Received 12 June 2011; Final revision received 14 August 2013

HIDDEN BUT IN PLAIN SIGHT: THE ROLE OF SCALE ADJUSTMENT IN INDUSTRY DYNAMICS

THORBJØRN KNUDSEN,[1] DANIEL A. LEVINTHAL,[2*] and SIDNEY G. WINTER[2]
[1] *Strategic Organization Design Unit (SOD), Department of Marketing and Management, University of Southern Denmark, Odense M, Denmark*
[2] *Department of Management, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.*

While much is understood about the general pattern of industry dynamics, a critical element underlying these dynamics, the rate of the expansion of individual firms, has been largely overlooked. We argue that the rate at which firms can reliably increase their scale of operations is a critical factor in understanding the structure of industries. Further, success at scaling-up the firm's operations provides a dynamic-isolating mechanism that insulates established firms from new competition. We show that the bases of profitability in the industry (monopoly-like profits stemming from the restriction of output, efficiency rents based on firm-specific productivity differences, or transitory Schumpeterian profits) can be traced to the scale adjustment process. We explore these issues in a computational model of industry dynamics.

INTRODUCTION

Strategy scholars have, to a large extent, defined their mission as unpacking the sources of heterogeneity in performance differences among firms (Rumelt, Schendel, and Teece, 1994). While early examples of economic approaches were focused on the industry level of analysis (Porter, 1980), subsequent research has centered on heterogeneity at the firm level (Rumelt, 1984; Wernerfelt, 1984). Of course, there are important linkages between accounts of firm-level heterogeneity and industry performance (Demsetz, 1973; Nelson and Winter, 1982). In this work, we wish to focus on a particular mechanism of linkage: the rate of adjustment of firm-level scale. This linkage comes to the fore once we recognize that the expansion of a firm is restrained in ways that interact with the underlying nature of its capabilities.

Firm capabilities are not simply a scale-free attribute (Levinthal and Wu, 2010); but as Penrose (1959) long ago argued, there is a limit to the rate of scale adjustment that a firm can realize in a given span of time. As the firm pushes toward that limit and compresses capability development, it is increasingly likely that such efforts will prove disruptive to its operations (Dierickx and Cool, 1989). Both of these properties not only determine the fate of the individual firm, but also influence industry structure. Prior work has shown that the rate of scale adjustment influences the evolution of industry structure (Nelson and Winter, 1982), but the possibility that firm capabilities may be altered in that process poses additional issues. In this work, we wish to examine how the rates of scale adjustment and firm-level capabilities jointly determine the evolution of industry structure and the implications of these processes for firm-level profitability.

Keywords: industry evolution; firm growth; competitive advantage; adaptation; NK model
*Correspondence to: Daniel A. Levinthal, Department of Management, The Wharton School, University of Pennsylvania, 2000 Steinberg-Dietrich Hall, PA 19104–6370, U.S.A. E-mail: dlev@wharton.upenn.edu

Prior models of industry evolution have highlighted the fact that entry barriers may rise as a result of the increasing level of the average quality of firms, typically characterized in terms of the incumbents' production costs (Lippman and Rumelt, 1982). As a consequence, the threshold of what constitutes a viable entrant continues to rise; i.e., the required level of realized production costs is driven down. However, even a "successful" entrant, successful in the sense of obtaining some initial toehold of economic activity, faces a subsequent challenge in terms of scaling up its operations to achieve its full profit potential. We suggest that this challenge of reliably scaling up one's operations supplements Rumelt's (1984) notion of isolating mechanisms. This additional isolating barrier is inherently dynamic, as it is associated with what can be viewed as a liability of growth. Reliably replicating across time a firm's production processes is not a trivial matter (Nelson and Winter, 1982; Rivkin, 2001), a challenge considerably heightened as firms attempt to increase their scale of operations (Winter and Szulanski, 2001). During the process of upward scale adjustment, firms are exposed to the liability of growth as the adjustment process may give rise to spells with heightened production costs and loss of efficiency. Of course, in a benign environment where the underlying technology and capabilities can easily support one's expansion of operations, most firms will survive these spells of possible disruption. In that case, the dynamic-isolating mechanism is very weak. By contrast, in cases where growth is notably error-prone, the dynamic-isolating mechanism can play an important role as a barrier that protects incumbents from being challenged by new entrants.

We explore these issues in a computational model of industry dynamics. We show that the rate of scale adjustment, in conjunction with adjustment error, is central to a mechanism that tends to shield off a small set of established firms from the competitive force of continuing entry—even when new firms enter with lower cost values. Our analyses establish how and why this isolating mechanism operates. We are also able to establish a relation between the rate of scale adjustment and well-known forms of rent (i.e., Ricardian rents, Schumpeterian rents, and monopoly rents). Our analyses show how the basis of rent changes as a function of the level of adjustment error, the rate of scale adjustment, and the degree of interdependency in firms' activity systems.

OVERVIEW: FIRM GROWTH, SCALE, AND INDUSTRY STRUCTURE

We focus in this paper on the rate at which firms can reliably increase the scale of their activities, arguing that this is a neglected but critical factor in understanding the evolution and structure of industries. Substantial literatures exist that study closely related issues from a variety of perspectives, so the claim that our specific question has been "neglected" may seem surprising. Before proceeding to the description of our model, we seek here to substantiate that claim by referencing some of the adjacent literatures and identifying their specific concerns, which contrast with our own.

In the literature of industrial organization economics (IO) and subsequently in the IO-based strategy literature (e.g., Caves and Porter, 1977; Porter, 1980), static scale economies were proposed as one source of entry barriers and an important determinant of industry structure (Bain, 1956; Porter, 1980; Sylos-Labini, 1993). Indeed, scale economies constituted the leading candidate for a technological determinant of "structure" under the "structure, conduct and performance" paradigm of the "old IO" (Scherer and Ross, 1990). Among sources of such economies, the dominant focus was on the indivisibility of plant and equipment, which implied an emphasis on economies at the establishment (plant) level, but firm-level economies in functions like research and development (R&D) and advertising also received attention. For present purposes, we simply note that this line of inquiry involved only perfunctory reference to dynamic issues in general and none to the adjustment costs that firms might face in changing scale or the ways that firm-scale considerations connect to the historical patterns of industry evolution. For analytical purposes, this line of argument takes both the purported causes and the observed reality of large firms as essentially "given" timeless phenomena.

In the present work, we explicitly address the dynamics of firm growth, believing that it is inappropriate to disregard the obvious fact that large firms become large by a process of protracted growth. Indeed, there is a substantial literature devoted to understanding how characteristics of firm growth processes underpin the related phenomena of the existence of large firms, the concentrated structures of many industries, and the characteristic shapes of firm size distributions. The longitudinal dimensions of firm growth phenomena have been explored largely under the headings of "industry life cycle" or "industry evolution", or simply "business history". A particularly significant phenomenon underscored by this line of inquiry is the phenomenon of the "shakeout"—the dramatic reduction in the number of surviving firms that typically occurs a few decades into the history of a new industry. In addition, the elevated mortality rates of small and young firms are an aspect of the economy that analyses of firm growth have noted (Dunne, Roberts, and Samuelson, 1988; Hannan and Freeman, 1989).

As to the sources of adjustment error that constrain firms' growth rates, a number of authors have argued that the primary source of adjustment costs do not lie in the external market (where a surge of demand for inputs might bid up prices), but in considerations internal to the growing firm. A classic in this tradition is Penrose (1959), who argued that firm growth is constrained by the firm's ability to manage it. Lucas (1967) invoked "internal costs of investment in the form of output foregone" in the construction of a theory of firm investment. In another contribution, Rubin (1973) extended the Penrose logic deeper into firm operations and sketched an associated formalization. More recently, Winter and Szulanski's (2001) discussion of replicators suggests that firms and business formats may vary in the ease with which they may scale up without damaging the firm's operating effectiveness. Central to their argument is the role of knowledge and, in particular, the fact that instilling the relevant knowledge in new capacity is a time and resource-consuming process, as Rubin (1973) also argued. Closely related to that logic is the finding that a strong organizational culture can be threatened by a rapid influx of newcomers (Harrison and Carroll, 2006). These arguments all indicate that the underlying sources of adjustment error can be attributed to error-prone transmission of firm-specific knowledge when the adjustment process occurs too quickly.

Organizational interdependencies are another important source of adjustment error noted in the management literature. Rivkin (2001) highlights the role that interdependencies among the firm's policy choices may play in the reliable replication of a given competitive position. If organizational practices are only loosely coupled in the organizational system, then changes in any given practice will not have dramatic consequences for overall organizational performance (Levinthal, 1997); however, in settings where these practices are more interdependent, then perturbations in a single practice may have broader and more significant consequences. Thus, in the latter setting, even modest rates of unreliable replication of particular policies and processes across time may be quite damaging to the prospect of a firm sustaining an enduring competitive advantage (Rivkin, 2000, 2001).

Work on the diseconomies of organizational scale (McAfee and McMillan, 1995) is also suggestive of the possible bases of adjustment error. As the firm grows, information about the relevant capabilities gets more dispersed among the individuals in the firm. To the extent that growth precedes design and implementation of proper coordination and control mechanisms, a likely scenario with rapid expansion is an increasing level of adjustment error. The sources of adjustment error highlighted in prior literature thus include error-prone transmission, or replication, of firm-specific knowledge, amplified by organizational interdependencies and organizational diseconomies of scale adjustment. Illustrative of these conceptual arguments are Toyota's recent difficulties that stemmed from their aggressive efforts to increase their scale so as to become the largest global automobile firm. This led to a loss of control of their product development and manufacturing systems, resulting in major product defects and subsequent recalls (Linebaugh and Shirouzu, 2010). Similarly, Senge's (1990) account of the rapid rise and subsequent collapse of People's Express reveals the challenges of maintaining adequate systems and skilled personnel in the context of firms' rapid expansion. The present contribution draws on these various sources of insight into "adjustment costs," broadly construed, but pursues the implications that appear when the problem of scale adjustment is explicitly introduced to the dynamics of firm and industry growth—and hence into the explanation of industry structure. This juxtaposition, and the interactions discovered in it, is the focus of the present paper. Consistent with previous work in the industry dynamics tradition, we view structure as an emergent property from the process of industry evolution, and we relate our analysis to the familiar patterns of change in industries.

DOI: 10.1002/smj

MODEL STRUCTURE

The model comprises a characterization of competitive dynamics at the firm level and of the associated market processes. At the firm level, the essential elements are the characterization of firm heterogeneity and a specification of how firm growth dynamics impact firms' capabilities. At the market level, the key elements are the characterization of market competition, for which the Cournot model is adapted, and a competitive process of entry and exit. We use the Cournot model to capture period-to-period firm-level competitive interactions because it is commonly used for this purpose, both in the literature on industry organization and in models of industry evolution.

Following on the work of Lenox, Rockart, and Lewin (2006, 2007), we use the structure of NK fitness landscapes to link a firm's production choices to a measure of productivity. Placement on the NK landscape determines the firm's unit cost of production (Lenox *et al.*, 2006, 2007). Further, the NK structure provides a mechanism by which we can consider both how unintended changes in the firm's capabilities appear as a result of changing scale and the possibility of firms adapting and "repairing" such errors—and in the process possibly even improving their performance by search on the fitness landscape.

More formally, the fitness landscape has two basic parameters, N, the total number of policy choices, and K ($<$ N), the number of policy choices that each choice depends upon. Each of the choices is assumed to be binary. The fitness contribution for each of the 2^{K+1} distinct payoff-relevant combinations (two values for the focal choice value and 2^K possible combinations of the K policies on which the focal policy depends) is drawn randomly from a uniform distribution over [0,1]. Note that with K equal to its minimum value of 0, the fitness landscape is smooth and single-peaked: Changes in the setting of one choice variable do not affect the fitness contributions of the remaining N-1 choice variables. At the other extreme, with K equal to N-1, a change in a single attribute of the organization changes the fitness contribution of all its attributes, which can be shown to produce many local peaks rather than just one, with each peak associated with a set of policy choices that are internally consistent. No local peak can be improved on by perturbing a single policy choice, but local peaks may vary considerably in their associated fitness levels. In our analysis, we vary the value of K to capture the effect of greater or less interdependencies but keep the value of N equal to 15.

Firms enter the industry at a specified minimum scale of production, q_0. This assumption is broadly consistent with the available evidence that small-scale de novo entry is common in most industries (Geroski, 1995). The value used for minimum scale of production, $q_0 = 0.05$, is calibrated with respect to the demand environment we investigate. A (much) higher value of q_0, relative to the carrying capacity of the market (as defined by the demand function), would act like a traditional entry barrier of large minimum efficient scale relative to market demand. Thus, higher values of q_0 would restrict the number of firms in the industry. In contrast, much lower values of q_0 would introduce firms with vanishing small scale and therefore tend to delay the point at which a shakeout occurs. However, numerous robustness checks with alternative positive values of q_0 do not reveal any qualitative changes in results.

To model competition in the product market, we focus on the equilibrium outcome of Cournot competition under constant returns with heterogeneous firms. This makes the desired scale of a firm's operations a function of overall market demand and the distribution of cost values among competitors. By cost values, we refer to unit costs of individual firms. A firm calculates its optimum Cournot quantity and hence it's desired scale of production, q*, based on the unit costs of firms that are currently operating. Hence the desired scale of production q* is a moving target that the firm aims to achieve by expanding (or reducing) its current production capacity. The scale of production at period $t+1$ is

$$q_{t+1} = \delta\, q^*_{t+1} + (1-\delta)\, q_{t,} \qquad (1)$$

where δ represents the rate of scale adjustment used by the firm to approach the period $t+1$ optimal scale q^*_{t+1}. We can allow for immediate adjustment by the individual firm to the desired target by setting $\delta = 1$ but recognize that this is just one extreme setting for a more general process. At the opposite extreme, $\delta = 0$, is the situation where a firm cannot adjust its scale of production.

A critical feature of our argument is that rapid adjustment in scale (higher δ) may cause the firm to inadvertently disrupt some of its

DOI: 10.1002/smj

operating practices and thereby alter its cost value. We represent this risk by postulating that the probability that one of the N policy choices randomly shifts from one period to the next is an increasing function of the rate of change in the scale of operations. Thus, the probability that any one of the N elements changes is

$$p_e = \left(1 - 1/\exp\left(\gamma \text{ abs}\left[q_{t+1} - q_t\right]/q_t\right)\right), \qquad (2)$$

where the error rate (p_e) is a probability that a given element of the firm's production practices changes as a function of the adjustment of the firm's scale $[q_{t+1} - q_t]/q_t$ and the fragility (γ) of the firm's practices to changes in the scale of operations. Note that the rate of scale adjustment δ determines q_{t+1} (Equation 1) and thereby the adjustment of the firm's scale that can be realized in a given period, $[q_{t+1} - q_t]/q_t$. With higher δ, the increment $[q_{t+1} - q_t]/q_t$ is increased, and so is the error rate (p_e), for any given level of fragility (γ). We use the term "fragility" (γ) to refer to the likelihood of unintended disruption of production practices when a firm adjusts its scale. In our model, the parameter γ varies between zero and one. When γ is set to zero, then the error rate $p_e = 0$ and no unintended change occurs in a firm's cost value no matter how much it adjusts the scale of production. As γ increases, so does the error rate (p_e) for a given change in scale.

While our focus is on the impact of the constraints on growth, or more precisely the potential adverse effects rapid growth may have on firms' unit costs, it is also worth considering how a firm's reduction in its scale of operations might impact its operating capabilities (Nelson and Winter, 1982: 121). We explore both a symmetric set-up where the adverse impact of growth and decline are held to be the same and an asymmetric set-up in which only increases in scale pose the risk of negative impacts on operating capabilities. As results only differed marginally between these treatments, we confine the reported results to the asymmetric set-up where adjustment error is only present when firms increase scale.

Firm capabilities may also change through processes of intentional firm-level adaptation. We consider the most basic of such adaptive processes, that of local search. Adding broader search to our model would impact the observed industry dynamics, and in particular reduce the level of industry concentration, but not qualitatively alter the basic tensions and effects identified in the analysis presented here.

In each period, firms may consider an alternative configuration one step removed from their current activities. This capacity for local search may, particularly in low-K landscapes, ameliorate the negative effects on firm's operating capabilities of unintentional changes in policy values as a result of changing the scale of operations. However, as is typically assumed (Lenox *et al.*, 2006; Levinthal, 1997; Rivkin, 2000), firms are unable to detect the root cause of unintended changes and examine an individual policy choice identified at random. In a high-K setting, even the capacity of engaging in local search therefore may not allow a firm to reestablish its former operating capabilities. Furthermore, the level of heterogeneity that we observe among firms is driven by K, as more or less rugged landscapes impact the capacity of firms to identify alternative positions in the competitive landscape (Levinthal, 1997).

Each period, a cohort of N_C firms considers entering the industry. Each member of this cohort is randomly assigned a set of N policy attributes, which in turn determines the firm's unit cost value. A firm enters the industry if its cost value is less than the current market price of a unit of output. Thus, while the cohort size of potential entrants is held fixed over time, the number of actual entrants varies a great deal, with the typical pattern being that the set of entrants is equal or nearly equal to the number of potential entrants early on and declining over time. Further, while the distribution of policies from which the potential entrants draw is held fixed, the quality of firms that enter increases over time as the competitively determined threshold of the minimally viable cost position increases (i.e., the cost value to satisfy the entry requirement declines). Furthermore, we have engaged in supplementary analyses in which we allow for the possibility that entering cohorts of firms may be able to imitate the practices of established enterprises, including the possibility that entrants may be able to identify and imitate, with some probability, the practices of the leading firms in the industry. As further explained in the concluding section, the effect of this modification is to speed up the process of diluting industry concentration, and thereby reduce industry profitability.

In addition to a process of entry, it is necessary to characterize a process by which firms withdraw

DOI: 10.1002/smj

from competition in the industry. Symmetric to the entry choice, firms are assumed to exit when their cost of production exceeds the market price. An additional exit criterion is whether the actual scale q_t at any point in time becomes less than the minimum scale of production q_0. We have tracked whether the q_0 constraint or the constraint on the firm's unit cost being below the current market price formed the basis of exit and found that, in all instances, unit costs were the binding constraint.

As in Klepper (1996), the entry and exit decisions are made simpler by the absence of fixed costs. The presence of such costs would require firms to engage in a multi-period calculation as to their likely stream of profits and whether they will be able to recover any fixed investments. Other forces may also require multi-period calculations, including a shifting demand function and, more subtly, expectations regarding the possible impact of a firm's adjustment path on its cost value. We abstract from these possible multi-period calculations and treat firms as making rational calculations from a one-period, myopic perspective.

Finally, there is the question of the demand environment in which firms operate. We focus on the case of an exponential demand curve, where

$$P = P_0 \exp\left(-Q/Q_1\right), \tag{3}$$

and P_0 and Q_1 are fixed parameters, with P being the industry-level equilibrium price at time t and Q the aggregate output value, summing over all incumbents at time t. The exponential demand function shares some analytical properties with linear demand, but has the additional virtue that price is strictly decreasing over its unbounded domain of quantity values. This is a valuable property when global comparisons are made across a wide range of cost and capacity conditions. For the exponential case, the elasticity of demand is Q_1/Q, and the Cournot equilibrium output for firm i is:

$$q^* = Q_1 \left(1 - W\left(\sum c_i / P_0 \exp(n)\right) c_i / \sum c_i\right), \tag{4}$$

where n is the number of incumbent firms in the industry; c_i is the unit cost value for firm i; and W is the Lambert W function (see Appendix S1 for derivation).

The link between the relevant input variables of the model and the output variables is derived as follows. First, the distribution of the incumbent firms' unit costs c_i is determined. Each firm randomly draws a configuration of $N = 15$ policies. Placement of the resultant configuration of policy attributes on the NK landscape determines the firm's unit cost of production, c_i. Given the distribution of unit costs, the desired (optimum) scale of production q^* is derived from the Cournot model (Equation 4); the industry-level price P is computed (Equation 3); and profits are determined, which equal the difference between P and individual unit costs (c_i) times the firm's quantity of production (q_t). Second, the rate of scale adjustment (δ) determines how fast the firm can move toward its desired optimal scale (q^*). The realized adjustment of scale is an increment ($q_{t+1} - q_t$) with which the firm moves toward optimum scale q^* – a higher δ, the larger q_{t+1}, and the larger $[q_{t+1} - q_t]/q_t$ (Equation 1). Third, the level of the fragility of production practices to changes in scale (γ) and the chosen adjustment of scale ($[q_{t+1} - q_t]/q_t$) then jointly determine the error rate p_e for each firm (Equation 2). Each policy attribute is perturbed with probability p_e, and a new cost value c_i is drawn from the NK landscape.

RESULTS

Each run of the model produces a particular history of an industry. Unless indicated, the results reported here are based on averages of 100 such histories with each "history" based on an independently specified NK landscape with a common value of K and spanning 1,000 discrete time-steps. We set the default cohort size to be 25 potential entrants. We choose this specification for simplicity. Arguably, a more realistic picture of industry evolution might involve a gradual increase in the number of candidate firms that consider entry. Additional simulations show that our main findings are robust to this respecification.

The baseline setting of the demand environment is to set the parameters of exponential demand to be $P_0 = 1$ and $Q1 = 100$. The default cohort size and demand function parameters are calibrated so that results are fairly congruent with common patterns that have been observed in well-known empirical examples of firms operating in product markets such as tires, television, and pharmaceuticals (Klepper, 1997). Further, our

DOI: 10.1002/smj

results are qualitatively robust to changes within a broad range of these parameters and are not a "knife edge" property of the model.

Effect of scale adjustment and industry shakeout

To provide the simplest baseline situation, consider a setting in which there are no inadvertent policy changes as the scale of the firm changes ($\gamma = 0$) and there is no firm-level adaptation (i.e., local search). Figure 1 shows the number of firms in the industry over time.

We see the classic pattern of "shakeouts" that has been well documented in the empirical literature (Klepper, 1997; Klepper and Graddy, 1990). By shakeout, we mean an elimination of incumbents measured by a decline in the number of firms operating in the industry. The number of firms in the industry rapidly rises to some 90 participants and then there is a marked decline over subsequent time periods until a relative stable number of some 50 firms is established. Figure 1 uses 1,000 time periods as a reference point in order to illustrate the long-run steady state behavior of the model. However, this longer timescale makes the exposition of the industry shakeout rather compressed, and its signature thereby appears a bit "spiky." In the example shown in Figure 1, the number of firms in the industry has largely stabilized by period 175, which is about 150 periods subsequent to the onset of the shakeout. We have therefore inserted the portion of Figure 1 that illustrates the industry shakeout on a time-scale that is more commonly used for this purpose (i.e., from the birth of the industry to the establishment of a relatively stable number of firms in the industry). This smaller portion of Figure 1 has the familiar signature of industry shakeout that we see from empirical illustrations.

Competitive pressure intensifies as industry participants scale up their operations toward the level implied by their position in a prospective Cournot equilibrium determined by the unit cost levels of the current industry participants. However, it is important to note that this desired capacity itself is a "moving target" as, each period, firms recalculate their desired capacity (q*) based on the changing competitive conditions that they face. It is also important to recognize that the relative stability in the number of firms from period 175 onwards

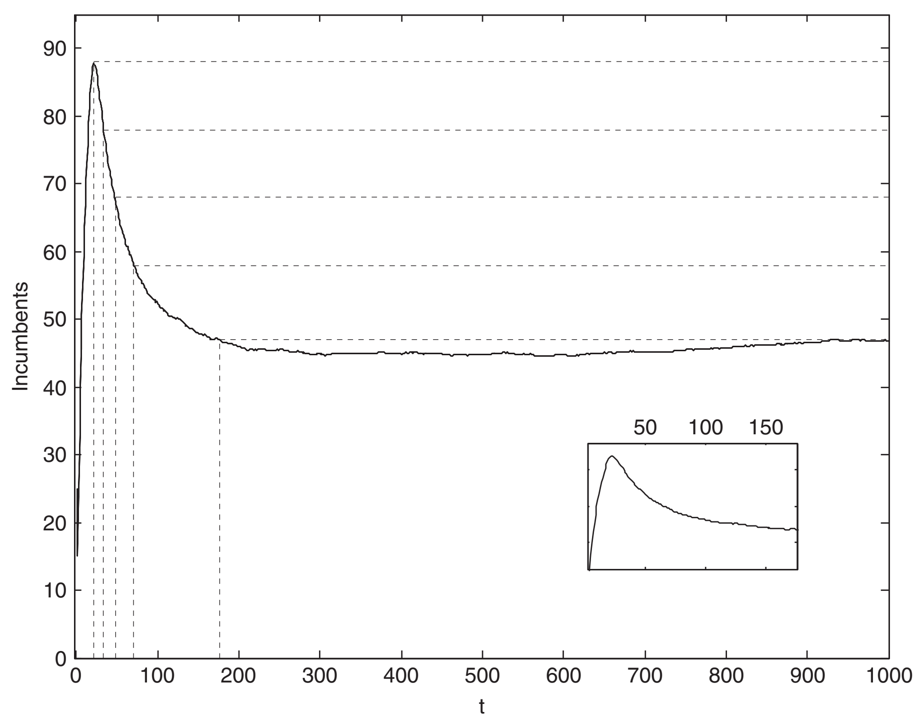

Figure 1. Number of firms over time. Fragility, $\gamma = 0$, rate of scale adjustment, $\delta = 0.01$. Average over 100 samples. No local search, $N = 15$, $K = 0$.

DOI: 10.1002/smj

does not imply that it is the same firms competing throughout. There is still significant churn as occasionally a successful new entrant may exert competitive pressure and drive out a formerly viable firm from the industry. Indeed, the average entry time of those firms in operation at the end of the simulation in period 1,000 in this setting is roughly period 500, suggesting that much of the entry occurred after the initial period. This image of continuing entry, largely ineffectual churn at the bottom of the size distribution, and occasional entrant success is broadly consistent with empirical findings (Audretsch and Mahmood, 1995; Dunne *et al.*, 1988).

Our analyses show that growth in firm size is a necessary condition for a shakeout under the demand context studied here. In the absence of growth in the size of individual firms, the competitive dynamic would consist solely of one equal sized firm displacing another equal size firm, but one which has an inferior cost value. The industry would simply fill up with incumbents until the number of firms asymptotes at its equilibrium value, as determined by the carrying capacity given by the demand context—a bit like pouring water in a bathtub. When the tub is full, a trickle of overflow roughly matches the small inflow, since exit is in effect and operating in an incremental manner. There is a continual entry of firms as it is possible that a prospective entrant will discover an ever lower cost value. The result is an unconcentrated atomistic industry structure with many, very small, equally sized firms. As the dynamics play out for arbitrarily long periods, these firms would all have the exact same minimal cost value.

While the fact that the growth of individual firms is a necessary condition for a shakeout may be fairly intuitive, the impact of varying the rate of scale adjustment (δ) is considerably less so. We measure the magnitude of the industry shakeout as the difference between the maximal number of firms present at any time period and the number of firms present at the end of the simulation in period $T = 1000$, with the number of firms present at the end of the simulation ($T = 1000$) being representative of the steady state, which typically occurs much earlier (generally prior to period 500). Using this measure, we find that the magnitude of the shakeout is actually negatively correlated with the rate of scale adjustment (for instance, the overall correlation coefficient computed for all treatments depicted in Table 1 is, $r = -0.52$, $p < 0.001$).

Competitive pressure in the industry grows with the arrival of successful entrants and the growth in size of industry incumbents. Unsuccessful potential entrants are those firms whose initial cost value exceeds the current industry price P. These new arrivals never enter the industry. With the possible exception of the earliest few periods, the number of incumbents considerably exceeds that of the number of successful entrants in a given period. Therefore, with a low rate of growth of individual firms (i.e., low δ), the cumulative growth in competitive pressure is attenuated. As a consequence, the timing of the shakeout is later, the lower the rate of scale adjustment (see Table 1). But while the shakeout is less aggressive in the sense of its timing being delayed, the severity of the shakeout itself actually increases in settings in which the rate of scale adjustment is reduced (see the results labeled "Magnitude of Industry Shakeout" in Table 1). A low rate of scale adjustment (low δ) allows a large number of firms to enter the industry successfully as even early successful entrants continue to operate at a modest scale. Therefore, the cumulative growth in pricing pressure proceeds at a modest pace, and shakeout effects arise in later time periods.

Thus, there are two elements underlying the large shakeout with a low rate of adjustment. Consider the distribution of latent cost values. The competitive conditions impose a constraint on what portion of this distribution of cost values is competitively viable. Naturally, in a more densely populated industry, one would expect a larger number of firms to lose competitive viability for any incremental movement downward in market price. However, even more critical to the dynamics that we observe for low adjustment rates is that restoring a new industry-level balance requires a larger population of incumbent firms to be displaced with new entrants.

Industry concentration and dynamic-isolating mechanisms

Across all treatments in Table 1, we see the following pattern: For almost any rate of scale adjustment (δ), higher levels of fragility (γ) are associated with notable increases in the magnitude of the industry shakeout as well as higher industry concentration and larger profits (we measure concentration as the number of incumbents at T, and omit other concentration measures such as

Table 1. The effects of rate of scale adjustment (δ) and fragility (γ)

	Magnitude of industry shakeout			Timing of shakeout			Number of incumbents at T			Total number of entrants at T			Total number of exits at T		
$\delta\backslash\gamma$	0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05	0.50
0.001	100	107	112	84	89	97	47	45	51	413	541	1142	366	496	1091
0.005	66	71	83	28	30	32	29	28	27	269	393	1405	240	365	1378
0.01	49	54	69	18	18	20	27	27	22	228	398	1751	201	371	1729
0.02	36	41	55	12	12	12	25	26	16	199	414	2062	174	388	2046
0.03	29	34	45	11	10	10	25	25	14	186	456	2261	160	431	2247
0.04	23	28	40	8	8	9	26	24	13	179	487	2389	153	463	2376
0.05	19	24	36	7	7	7	27	24	12	175	503	2406	149	480	2394
0.10	11	17	24	1	1	1	25	17	9	158	625	2673	133	607	2664
0.50	4	18	15	1	1	1	26	7	10	130	1586	4417	105	1579	4407
1.00	5	19	18	1	1	1	25	6	7	130	2360	4176	105	2354	4169

	Average unit costs at T			Std. dev. of unit costs at T			Average profit per firm at T			Cumulative profit per firm at T		
$\delta\backslash\gamma$	0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05	0.50
0.001	0.28	0.29	0.33	0.02	0.02	0.02	0.15	0.17	0.11	129	137	105
0.005	0.28	0.29	0.35	0.01	0.01	0.02	0.12	0.18	0.20	140	166	185
0.01	0.27	0.29	0.35	0.01	0.01	0.02	0.12	0.16	0.28	124	167	269
0.02	0.27	0.30	0.36	0.01	0.02	0.02	0.13	0.16	0.42	136	164	398
0.03	0.27	0.30	0.36	0.01	0.02	0.03	0.12	0.17	0.51	130	169	452
0.04	0.27	0.31	0.36	0.01	0.02	0.03	0.12	0.17	0.57	128	173	496
0.05	0.27	0.31	0.36	0.01	0.02	0.03	0.12	0.18	0.65	118	196	552
0.10	0.27	0.31	0.36	0.01	0.02	0.04	0.13	0.31	0.94	128	301	677
0.50	0.27	0.33	0.40	0.01	0.05	0.05	0.12	1.15	0.80	128	645	119
1.00	0.27	0.36	0.40	0.01	0.06	0.04	0.13	1.44	1.01	126	532	8

All results are based on averages over 100 samples from simulations of our baseline model (K = 3, no local search) with T = 1000 time periods. The magnitude of the industry shakeout is measured as the difference between the maximal number of firms present at any time period, and the number of firms present at T = 1000.

C4 and HHI because they act in a similar manner as the measure provided). The exception is that a combination of very high scale adjustment rates ($\delta \geq 0.50$) and fragility ($\gamma = 0.50$) greatly reduces the magnitude of industry shakeout, industry concentration, and profits (cumulative as well as average profits per firm at T). This latter set of conditions lead to an ongoing churn where, in each time step, failing incumbents are replaced by new entrants that soon become the next period's failures. At the opposite extreme, very low adjustment rates ($\delta = 0.001$ or less) and high rates of fragility increase the magnitude of the shakeout but at the same time dilute industry concentration and profits. This is because the failing incumbents are of very small size at the time when they are replaced by new entrants. The results reported in Table 1 were obtained with a value of K = 3. Numerous robustness checks indicate that this qualitative pattern is largely independent of the value of K.

Our results further show that higher levels of fragility are associated with increasing means and standard deviations of firms' unit costs. This observation indicates that higher levels of fragility both decrease competitive pressure (firms with fairly high costs are viable over the longer term) and also isolates large-scale incumbents from the threat of low-cost entrants that might otherwise garner a share of the market. How can it be that more inefficient firms with higher costs are able to resist the ongoing pressure from low-cost firms that continue to enter the industry? What is the isolating mechanism?

A detailed analysis of the underlying dynamics provides an answer. Consider Figure 2, which provides a sense of the ongoing dynamics at a late stage (final 30 periods) of the industry evolution.

DOI: 10.1002/smj

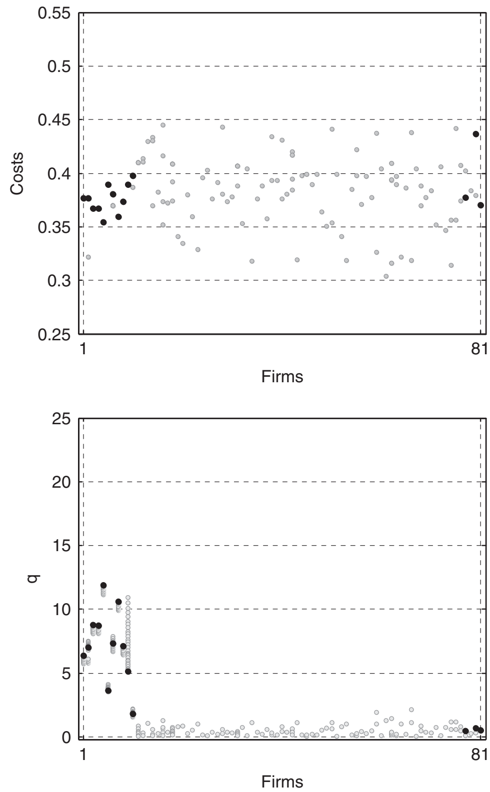

Figure 2. Panel 1: Firm unit cost values (over the last 30 periods). Firm unit cost values are ordered by entry time with early entrants on the left-hand side and later entrants toward the right of the x-axis. Unit costs for each surviving incumbent (at t = 1000) are shown in black. For firms that had more than one distinct cost value over the final 30 periods, unit costs during the prior periods are shown as a gray dot above or below this final value with each dot representing each of these firm-specific firm cost values. Firms that exited the industry during last 30 periods appear only as a gray dot(s); if these firms experienced more than one distinct cost value within the final 30 periods of the simulation, all such values are indicated by a gray dot in a single column associated with that firm. Evidence from a single history, fragility, $\gamma = 0.5$, $\delta = 0.05$. No local search, K = 3. Panel 2: Firm-specific quantity levels over time (over the last 30 periods). Firm quantity values are ordered in the same way as cost values.

Each point on the x-axis corresponds to the firm number for one of the 81 firms that was present at some time during the last 30 periods, with the numbers assigned according to the order of entry. Earlier entrants appear on the *left* of the x-axis, while later entrants appear on the *right*. In panel 1 of Figure 2, each "dot" in the figure corresponds to the cost value of a specific firm at a particular time. Note that each firm can have multiple cost values. This is because adjustment error may perturb a firm's cost value. In panel 2 of Figure 2, we show output values in a similar way. Since a firm can be in the process of scaling up or down toward its target output level and indeed this target itself can change over time, it can have multiple output values.

A black dot in these figures indicates the cost (output) value at the end of the simulation (t = 1000) of those firms that have managed to survive until the end of the simulation. In gray, we indicate prior cost (output) values during last 30 periods of the simulation. Thus, firms that exited the industry during the last 30 periods appear as a column (if more than one cost or output value) of gray dots, while surviving firms can be identified by columns (if more than one cost or output value during the last 30 periods) that contain a black dot.

Note the strong association of entry timing and survival, despite the fact that many of the younger firms achieved cost values that were lower and therefore superior to many incumbent firms. The set of firms that survive until the end of the simulation are either (1) very early entrants who both have fairly low unit costs and, equally importantly, have done so while achieving significant scale or (2) some very recent entrants that have cost values that make them viable in the industry but are operating at a scale of a tenth or less compared to the established firms (for illustration, see panel 2 of Figure 2). As these later entrants attempt to scale up to achieve their desired target output levels, they tend to disrupt their strong cost position. In contrast, the set of more mature firms operating at a scale proximate to their targeted output level have already surmounted the treacherous road to full scale.

This "road" to successful achievement of desired scale can be observed in Figure 2. Note that in panel 1 of Figure 2 that there are 11 firms marked with black dots to the left part of the x-axis. These firms have survived the process of scaling up toward the output levels that are shown in panel 2 of Figure 2. Even so, they engage in ongoing small adjustments in their scale because of continuing new entry and exit, which in turn changes their desired output level. These small adjustments may, with a small probability, cause a perturbation in the firm's cost value. However, mature firms will, on average, have stable cost values as any change in output level is likely to be modest and, therefore, the associated risk of

DOI: 10.1002/smj

unintended change in their cost level is rather low. Consistent with this general characterization, of the 11 "mature" firms, firms to the far left part of the x-axis, only three experienced a change in their cost value during the last 30 periods. These three firms appear as the 2nd, 7th, and 11th firm on the left on the x-axis (see panel 1 of Figure 2). The 10th most leftward firm is also scaling down as the result of a shift in its cost value prior to the 30th period.

Further, in panel 2 of Figure 2, note the long gray scatter of low values near the x-axis. These are output values of firms that started out with unit costs sufficiently low to make them viable at the point of entry and then, during the process of scaling up, suffered from adjustment error that led to a cost value that was higher than the industry price. These firms therefore had to exit and the shadow of their prior existence is marked by gray columns. The corresponding gray shadow of cost values is shown in panel 1 of Figure 2.

Finally, to the right side of the x-axis (in both panels of Figure 2), we see three black dots, indicating three new firms that entered toward the later part of the simulated industry history. These fairly young firms have not yet scaled up to their desired output level. Therefore, there is a substantial likelihood that adjustment error, at some point in time during the process of scaling up, will lead to an adverse cost value that is higher than the industry price. While the "noise" generated by the adjustment process may, by chance, actually enhance a firm's cost position, a random perturbation of policy choices, on average, will be a negative event for a firm that has a superior cost position. As a result, very few firms manage to traverse the entire path of scaling up in the presence of adjustment error. For the setting shown in Figure 2, a handful of the firms marked with black dots at the far left of the x-axis have entered the industry during a relatively early period of the industry evolution. The advantage of the dynamic-isolating mechanism, which is even more pronounced at high levels of adjustment error, is very long lasting.

This analysis reveals two conditions that jointly serve as a dynamic-isolating mechanism. First, adjustment error makes the passage to full scale very hazardous. Very few firms are able to make this journey. Further, note that the lower the cost value, the higher is the target output level. Low-cost firms therefore must survive a potentially longer journey to desired scale, with the associated risk of unintended change in policies and, in turn, a tendency to experience an increase in unit costs. It is therefore very unlikely that a firm can both hold on to a low-cost value and also be able to scale up to its target output. In panel 2 of Figure 2, this property is indicated by the fairly low output values of all the firms that have exited (all the gray columns). It will only be possible for new entrants to maintain a low-cost value if they scale up very slowly. At the same time, a slow process of scaling up subjects firms to multiple adjustments, which at some time may increase unit costs above industry price. Both conditions imply that a small group of firms that happen to reach a desired scale of operations, and at the same time maintain fairly low unit costs, are somewhat isolated from competitive forces. Most other firms will fail to achieve both outcomes (low cost and substantial scale) subsequent to the industry shakeout.

This insight explains why we see an increase in the mean and variance of cost values with higher levels of fragility (γ). The few mature firms in the industry have moderately low, but not extreme, unit costs; they are nevertheless protected by the dynamic-isolating mechanism that prevents lower cost firms from maintaining their cost values as they scale up. However, even the mature firms face some risk as they engage in minor adjustments of output and face a small probability of further increasing their costs. Should this happen, the effect can be a dramatic decline and trigger a subsequent process whereby a new industry configuration is established. Per this logic, the life expectancy of firms declines with higher levels of fragility to changes in scale (γ).

This general pattern regarding the impact of adjustment error holds for the whole range of levels of interdependence among policy choices, from a setting of no interdependence ($K = 0$) to a maximally rugged landscape with $K = 14$ (i.e., $K = N-1$). The level of interdependence exerts a "main" effect on the results. A greater value of K tends to increase the variation in cost values of firms entering the industry. As a consequence of this greater variability, we see higher levels of concentration for higher levels of K, consistent with Lenox *et al.* (2007).

Selection over adjustment rates

In the prior analysis, we considered the impact of varying the rate of scale adjustment across

DOI: 10.1002/smj

populations of firms that themselves share a common adjustment rate. We extend this analysis to include an examination of a setting in which firms are endowed with heterogeneous, but fixed, scale adjustment rates. We wish to examine whether more (or less) rapid rates of scale adjustment would be positively selected for. We examine this possible differential selection over adjustment rates by seeding the population of firms with an adjustment rate that is drawn from a uniform distribution between zero and one. Selection favors firms that do not rapidly expand to meet a desired scale of production as defined by the Cournot equilibrium. For example, with $K = 0$ and a level of fragility of $\gamma = 0.05$, there is a fairly rapid decline from the mean value of $\delta = 0.5$, which corresponds to the mean of the underlying distribution of adjustment rates, to a value of roughly $\delta = 0.2$. As a further extension, we have also examined a setting in which the selection criterion is not based on the relationship between the firm's current cost and market price, but cumulative profitability. This selection criterion results in a positive selection for more rapid rates of scale adjustment when there is no error in the adjustment process. Getting big fast may allow the firm to accumulate substantial profits. As adjustment error is introduced, this effect is mitigated and for larger adjustment errors the positive selection for rapid rates of adjustment is eliminated.

Local search

The prior results regarding the impact of adjustment errors raise the question of the possibility of firms recovering from such errors. In particular, we consider a simple intentional process of incremental change that parallels the mechanism of unintentional change associated with adjustment error. We consider the implications of allowing the firm each period to identify a change in one of the N elements that constitute its set of policy choices, with the firm adopting the change if it would enhance its performance (i.e., reduce its cost value).

In detailed analyses of the effects of local search, we examined the degree of interrelationship among the policy choices (K), the rate of adjustment, and the possibility of local search. In absence of substantial interdependencies, local search substantially reduces the heterogeneity among firms. In the limit with $K = 0$, given sufficient time, all firms would find the optimal set of policy choices yielding the lowest cost value. Indeed, we found that with $K = 0$, with the exception of high levels of fragility and adjustment rate, the standard deviation in cost values among firms is less than 0.005. Further, for low values of scale adjustment and fragility, we see a very mild industry shakeout.

However, interestingly, for moderate and higher rates of scale adjustment and fragility, the results under local search tend to converge to those without search. Even with local search and a modest degree of interdependencies among policy choices, rapid rates of scale adjustment and substantial adjustment errors can create sufficient perturbations in cost values so as to make the behavior in the presence and the absence of local search relatively similar. This same pattern is present at higher levels of K. For instance with $K = 3$, at very low levels of scale adjustment the possibility of local search leads to lower variation in cost values among firms and a more moderate shakeout; however, above such low rates of scale adjustment, the results with and without local search tend to be similar.

In contrast, the presence of local search has a stronger and more robust effect on the average age of surviving firms. Local search both allows early entrants an opportunity to refine their set of choices so as to improve upon their cost value, as well as the possibility to recover from perturbations in their cost value as they scale up the magnitude of their operations. Thus, while we still observe a shakeout with $K = 0$ and local search, its magnitude is greatly reduced relative to what we observe in settings with a greater degree of interdependence among policy choices, as these interdependencies result in more persistent heterogeneity among the surviving firms. Again, with moderate to high values of K, local search becomes constrained by the complexity of the production process, and the results with local search are not significantly different from the impact of adjustment errors in the absence of local search.

In a similar vein, we find that with local search, there is a lower level of flux of entry and exit (i.e., fewer gray or black dots associated with late entrants). Further, with local search, all the firms that survive until the end of the simulation are firms that entered relatively early, while in the absence of local search the ongoing turbulence endured by incumbents creates the opportunity for a new fringe of small entrants even in the late

DOI: 10.1002/smj

stages of the simulation. These results are valid for modest levels of K; however with a very high value of K ($K = 14$), such local adaptation is not feasible.

Profitability

As a final element in our analysis, we consider the implications for profitability. The model generates bases of profitability that are associated with firm-level performance differences, which can be viewed as Ricardian rents, as well as more fleeting firm-level differences in performance that have, in that sense, more of the characteristic of Schumpeterian rents (Peteraf, 1993; Winter, 1995). Finally, there are settings in which the basis of profitability lies at the industry level and is driven by a classic Cournot restriction of output and therefore can be interpreted as monopoly rents.

In the limit, with $K = 0$, the only bases of possible rent are those associated with the Monopolistic restriction of output. $K > 0$ is a necessary condition for non-monopoly rents. This is because all firms will ultimately be able to identify the "best practice" in the presence of local search when $K = 0$. Stable firm-level differences, i.e., Ricardian rents, require that firms are able to grow and achieve significant scale while still holding on to their distinctive capabilities, which in turn implies that they have some combination of a modest rate of scale adjustment and a low rate of error associated with increases in scale. In settings of more rapid rate of adjustment or high rates of error in the adjustment process, firm-level advantages will tend to be transitory and therefore will yield a form of Schumpeterian rents.

CONCLUSION

Central to any evolutionary account are the dynamics of the system under consideration. Yet, despite an extensive prior literature on industry evolution, these industry-level dynamics have not been linked to the dynamics of scale adjustment at the firm level. The current work provides an initial effort at addressing that gap as we show that the feasibility and reliability of increases in scale at the firm level have important implications for both the short- and long-term composition of industries.

Scale adjustment of individual firms is a necessary and sufficient condition for the existence of a "shakeout" in the population of competitors. Furthermore, the rate and reliability of the scaling process conditions both the timing and degree of the shakeout. Our model offers two broad sets of empirical implications. One set of results revolves around the relationship between the rate of scale adjustment and the timing and magnitude of industry shakeouts. While it is reasonably intuitive that high levels of scale adjustment lead to a more rapid onset of industry shakeouts, the other basic result in this regard—that the severity of a shakeout diminishes with the rate of scale adjustment—is far less intuitive a priori. Yet, examining the patterns across industries in Klepper and Graddy (1990) provides suggestive support for this claim. Furthermore, Dunne *et al.* (1988) and Geroski (1995) provide evidence that the typical rates of scale adjustment are rather modest.

Further, we expect that industries in which firms have higher adjustment rates will not only have smaller industry shakeouts but also be more concentrated and more profitable. Similarly, industries with higher levels of adjustment error (as a function of the growth rate) will not only have larger industry shakeouts but also be more concentrated and more profitable. When analyzing the causes of observed industry configurations, empirical work needs to control for the rate of scale adjustment as well as the level of adjustment error that is present in the focal industries. Absent these measures, observed differences in concentration measures, profitability, and patterns of industry dynamics may either be misattributed to other causes or, less damaging, to unexplained variance.

Our results add to prior work on firm-level competitive advantage and rent generation as we identify a new dynamic-isolating mechanism. Surprisingly, the rate of scale adjustment in conjunction with adjustment error turns out to shield off a small set of firms from the competitive force of continuing entry, even when new firms enter with lower cost values. Our analyses establish how and why this isolating mechanism operates. While Lippman and Rumelt (1982) point to the importance of recognizing the persistent effect of firm heterogeneity on industry dynamics, a persistent heterogeneity due to "isolating mechanisms" (Rumelt, 1984) that prevent capabilities from diffusing among competitors, the challenge of scaling up poses an additional isolating mechanism. Entering an industry with a potentially superior cost value, or business model, is not sufficient. To

DOI: 10.1002/smj

achieve significant profitability and to establish a major position in the industry, a firm must effectively scale up its operations. While the notion of static scale efficiencies are well appreciated in the literature, the transformation from modest to substantial scale has been less appreciated. Even in the absence of economies of scale, an established firm operating at significant scale benefits from an advantage over potential and actual entrants.

The challenge of scaling up is importantly connected to the level of interdependencies among policy choices, which in turn links to the nature of rents obtained by competitors. Rapid scaling up of firms operating in relatively simple, or smooth, landscapes with low levels of interdependence can lead to the monopoly-like rents associated with market power in Cournot competition. More complex sets of interdependences lend themselves to the realization of more enduring and more pronounced levels of firm heterogeneity and, as a result, the presence of relatively large Ricardian rents.

Finally, while strategy scholars have often aimed at identifying sustainable firm-level profits, much of the accrual of profits have a more temporary or, Schumpeterian, nature. Even in the absence of technological change, or changes in the bases of competitive advantage, the fact that industries do not instantaneously reach some equilibrium state provides an opportunity for such transitory rents. The speed and reliability with which firms can achieve greater scale of their operations is central to their ability to realize and appropriate such transitory rents. Thus, while the literature has focused in recent years on dynamic capabilities that may facilitate the transition from one basis of competitive advantage to another, the arguably simpler dynamic capability to scale up in the context of a stable basis of competitive advantage is itself quite critical.

The industry dynamics we describe certainly depend to some extent on the assumption that the industry's product is substantially homogeneous, and thus it applies to only a subset of empirical settings. As firms grow, they typically elaborate their product offerings and, at larger scales, they engage in related diversification (Bottazzi and Secchi, 2006; Klepper and Thompson, 2006). This raises questions about the transfer of knowledge and capabilities that reach beyond the narrow "replication" paradigm that our formal modeling invokes. This is clearly an important area for future inquiry.

With respect to narrower changes in the characterization of the demand environment, our results seem to be quite robust. For example, we have conducted additional simulations showing that a linear demand function gives qualitatively similar results as the exponential demand function we use. We have also conducted additional experiments shifting the latent cost distribution upwards so that firms operate further into the elastic part of the demand curve. As expected, this has the effect of ameliorating the force of industry shakeouts but does not qualitatively change our results.

Another possible objection to our analysis relates to our assumptions about the distribution of latent cost values. The shape of the latent cost distribution used here is generated by the NK formalism, where the underlying fitness distribution is symmetric and bell-shaped with finite support. As N goes to infinity, it approximates a Normal distribution. On its face, this might be considered empirically implausible. Note, however, that the *latent* cost distribution of the model should not be regarded as a prediction of the realized distribution of cost values generated by the model, and it is the realized rather than the latent cost values that are available for observation in empirical studies. In our model, realized cost values appear as a function of entry conditions and scale adjustment in the course of the competitive process. Both the profitability tests at the point of entry and the subsequent selection over firms produce drastic discrepancies between the distribution of latent cost values and the distribution of realized cost values. Empirically, realized cost values tend to exhibit a skewed, long-tailed distribution with a few small values and many somewhat larger values. This realistic signature is in fact generated by our model. Thus, the particular mechanism by which we generate a distribution of cost values does not appear to be a significant limitation.

We have explored the implications of the possibility that the distribution of cost values from which potential entrants draw is not fixed but rather improves along with the improving distribution of production efficiencies of incumbent firms. In particular, we consider a variant of the model in which potential entrants have a probability r of imitating the policy choices of the best performing incumbent firm and a probability (1-r) of drawing a set of policy values at random from a fixed distribution. As r increases, industry concentration declines as the number of viable entrants increases. While long-established incumbents may be of larger scale, the large number of successful entrants imposes increasing competitive pressures on price–cost margins. In the limit as r increases, the industry approaches a setting of perfect competition and the absence of industry profits.

While our focus has been on identifying broad empirical patterns regarding the timing and severity of industry shake-outs and the associated implications for industry profits, an appreciation of these sorts of relationships, as Porter (1980) long ago noted, can inform managers' strategic actions. Firms face a trade-off regarding the degree to which they rapidly scale up and potentially capture transitory rents and possibly contribute to the reduction in the rate of entry with the risk to their own competitive position. Further, the fact that an industry may be dominated by a few large, and not necessarily the most efficient firms, need not imply a rapid shift in market shares when a more efficient challenger appears. The difficulty of achieving efficiency with scale is another barrier protecting the established firms' competitive position.

While a firm may not be able to influence the overall rate of scale adjustment in an industry, the firm can make its own distinct choices in response to that of other firms and the reliability, or conversely fragility, of its own production process in response to changes in scale. First, the analysis points to the importance of not only higher or lower unit cost of production (the usual consideration in models of industry competition), but also to the scalability of a given production process. A firm may be better off with a slightly higher unit cost if this alternative production process is more reliably scalable. A separate issue is how other firms' choice of scale adjustment should affect a focal firm's choice. A fairly direct implication is that a firm should not be lured into a competitive matching of increased scale in response to the possibly rapid growth by its competitors (an important caveat here is that our analysis does not deal with issues of scale economy). These rapidly growing firms run the considerable risk of losing tight control of their business processes and seeing their cost structure rise as a result. However, this argument is not symmetrical. Consider an extreme case in which one's competitors were inert and did not change their quantity level. The focal firm should still not choose to scale up rapidly as such an effort would still risk damage to its own cost position. Thus, in contrast to "arms race" like dynamics that occur in the context of learning/experience curve effects, the preferred firm choice of scale adjustment is not highly dependent on others' choice of scale adjustment.

From a managerial perspective, our work highlights two critical factors that jointly determine the realized adjustment error. While the base-rate of adjustment error may lie outside managerial control and reflect features of the industry and production technology, other factors are clearly strategic parameters. In particular, our work points to the importance of firm-level growth rate as a strategic parameter that may influence a firm's long-run competitive advantage. Further, a firm's capacity to transmit or replicate firm-specific knowledge reliably may allow the firm to scale at a more rapid rate while preserving its competitive advantage. More generally, a firm's choice of organizational form and structure and a firm's investment in systems and procedures may importantly impact the profitability of rapid scale adjustment. To conclude, and borrowing an image from work on the evolution of technologies, we have tried to introduce a new "dimension" to the analysis of industry evolution. While in any model of industry evolution assumptions, explicitly or implicitly, are made about the rate of scale adjustment in firms' size, it has not been in the foreground in these analyses. In addition to the particular substantive results we provide, more generally we hope we have provided an argument for bringing the questions of scale adjustment and its reliability to the foreground of our considerations of industry evolution.

ACKNOWLEDGEMENTS

The authors benefited from comments on previous drafts from Stephan Billinger, Tobias Kretschmer, Ulrik W. Nash, and Nils Stieglitz and two anonymous reviewers, as well as from participants at the 2009 DRUID conference and the 2010 conference of the International Schumpeter Society. This work was supported by the Danish Council for Independent Research.

REFERENCES

Audretsch DB, Mahmood T. 1995. New-firm survival: new results using a hazard function. *Review of Economics and Statistics* **77**: 97–103.

Bain JS. 1956. *Barriers to New Competition*. Harvard University Press: Cambridge, MA.

Bottazzi G, Secchi A. 2006. Explaining the distribution of firm growth rates. *RAND Journal of Economics* **37**(2): 235–256.

Caves RE, Porter ME. 1977. From entry barriers to mobility barriers: conjectural decisions and contrived deterrence to new competition. *Quarterly Journal of Economics* **91**: 241–262.

Demsetz H. 1973. Industry structure, market rivalry, and public policy. *Journal of Law & Economics* **16**: 1–9.

Dierickx I, Cool K. 1989. Asset stock accumulation and sustainability of competitive advantage. *Management Science* **35**: 1504–1511.

Dunne T, Roberts MJ, Samuelson L. 1988. Patterns of firm entry and exit in U.S. manufacturing industries. *RAND Journal of Economics* **19**: 495–515.

Geroski PA. 1995. What do we know about entry? *International Journal of Industrial Organization* **13**: 421–440.

Hannan MT, Freeman J. 1989. *Organizational Ecology*. Harvard University Press: Cambridge, MA.

Harrison JR, Carroll JR. 2006. *Culture and Demography in Organizations*. Princeton University Press: Princeton, NJ.

Klepper S. 1996. Entry, exit, growth and innovation over the product life cycle. *American Economic Review* **86**(3): 562–583.

Klepper S. 1997. Industry life cycles. *Industrial and Corporate Change* **6**(1): 144–181.

Klepper S, Graddy E. 1990. The evolution of new industries and the determinants of market structure. *RAND Journal of Economics* **21**(1): 27–44.

Klepper S, Thompson P. 2006. Submarkets and the evolution of market structure. *RAND Journal of Economics* **37**(4): 861–886.

Lenox M, Rockart S, Lewin A. 2006. Interdependency, competition and the distribution of firm and industry profits. *Management Science* **52**: 757–772.

Lenox M, Rockart S, Lewin A. 2007. Interdependency, competition, and industry dynamics. *Management Science* **53**(4): 599–615.

Levinthal DA. 1997. Adaptation on rugged landscapes. *Management Science* **43**: 934–950.

Levinthal DA, Wu B. 2010. Opportunity costs and non-scale free capabilities: profit maximization, corporate scope, and profit margins. *Strategic Management Journal* **31**(7): 780–801.

Linebaugh K, Shirouzu N. 2010. Toyota heir faces crisis at the wheel. *Wall Street Journal* 27 January. Available at: http://online.wsj.com/article/SB10001424052748704094304575029493222357402.html (accessed 25 September 2013).

Lippman S, Rumelt R. 1982. Uncertain imitability: an analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* **13**: 418–438.

Lucas, R. 1967. Adjustment costs and the theory of supply. *Journal of Political Economy*, **75** (Part I): 321–334.

McAfee RP, McMillan PJ. 1995. Organizational diseconomies of scale. *Journal of Economics & Management Strategy* **4**(3): 399–426.

Nelson RR, Winter SG. 1982. *An Evolutionary Theory of Economic Change*. Belknap Press: Cambridge, MA.

Penrose E. 1959. *The Theory of the Growth of the Firm*. Blackwell Publishers: Oxford, UK.

Peteraf M. 1993. The cornerstones of competitive advantage: a resource-based view. *Strategic Management Review* **14**: 179–191.

Porter ME. 1980. *Competitive Strategy*. Free Press: New York.

Rivkin J. 2000. Imitation of complex strategies. *Management Science* **46**: 824–844.

Rivkin J. 2001. Reproducing knowledge: replication without imitation at moderate complexity. *Organization Science* **12**(3): 274–293.

Rubin P. 1973. The expansion of firms. *Journal of Political Economy* **81**: 936–949.

Rumelt RP. 1984. Towards a strategic theory of the firm. In *Competitive Strategic Management*, Lamb R (ed). Prentice-Hall: Englewood Cliffs, NJ; 556–570.

Rumelt RP, Schendel D, Teece D (eds). 1994. *Fundamental Issues in Strategy: A Research Agenda*. Harvard Business School Press: Boston, MA.

Scherer F, Ross D. 1990. *Industrial Market Structure and Economic Performance*. Houghton Mifflin: Boston, MA.

Senge PM. 1990. *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday: New York.

Sylos-Labini P. 1993 (orig. pub. 1957). *Oligopoly and Technical Progress*. A.M. Kelly: Fairfield, NJ.

Wernerfelt B. 1984. A resource-based view of the firm. *Strategic Management Journal* **5**: 171–180.

Winter SG. 1995. Four R's of profitability: rents, resources, routines and replication. In *Resource-Based and Evolutionary Theories of the Firm: Towards a Synthesis*, Montgomery C (ed). Kluwer Academic Publishers: Boston, MA; 147–158.

Winter SG, Szulanski G. 2001. Replication as strategy. *Organization Science* **12**: 730–743.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix S1. Derivation of the optimal output decision for a firm given its unit cost and the unit cost of all other firms.