Creative Destruction in Industries

Boyan Jovanovic and Chung-Yi Tse*

February 28, 2006

Abstract

Most industries go through a "shakeout" phase during which the number of producers in the industry declines. Industry output generally continues to rise, however, which implies a reallocation of capacity from exiting firms to incumbents and new entrants. Thus shakeout seem to be classic creative destruction episodes.

Shakeouts of firms tend to occur sooner in industries where technological progress is more rapid. Existing models do not explain this. In fact, the relation emerges in a vintage-capital model which shakeouts of firms accompany the replacement of capital, and in which a shakeout is the first replacement echo of the capital created when the industry is born. We fit the model, with some success, to the Gort-Klepper data.

1 Introduction

Replacement of old capital and labor by new capital and labor is sometimes referred to as "creative destruction". It is also thought that such replacement is often accompanied by changes in the numbers of firms. Creative destruction may occur in many sectors at once, or it may involve a single sector, a single industry. At the industry level, the "industry shakeout" episodes that Gort and Klepper (1982) documented are classic creative destruction episodes: During a shakeout, a substantial fraction of firms exits, and yet industry output on average continues to rise implying that aggregate capacity does not fall. The capacity withdrawn by the exiting firms is replaced by incumbents or by new entrants, presumably using new capital.

Several explanations for these creative-destruction episodes have been advanced. When the entire economy is involved, the shock that prompts the wave may be a system-wide reform or deregulation (Atkeson and Kehoe 1993), or the arrival of a new general-purpose technology, a topic surveyed by Jovanovic and Rousseau (2006).

*We thank S. Kortum for comments, and the NSF for support.

At the level of an individual industry, four broad sets of explanations for such creative destruction waves have been put forward:

Temporary demand declines.—Caballero and Hammour (1994) find that if replacement of capital causes interruptions in production, it is optimal to replace capital when demand temporarily drops, for then the foregone sales are at their lowest. Klenow (1998) finds that when, in addition, the productivity of new capital rises with cumulative output, a firm will replace capital when the recovery is about to start. This raises the productivity growth of the new capital.

Technological advances.—A new technology like the assembly line can raise the efficient scale of firms in an industry, thereby crowding some of them out. This is argued by Hopenhayn (1993) and Jovanovic and MacDonald (1994). There is little doubt, for instance, that the arrival of the assembly-line technology raised the optimal scale of auto-manufacturing plants and caused a large reduction in the number of auto producers. Related to this, Utterback and Suarez (1993) argue similarly that a dominant design emerges, the winning model, which forces other models out and with them some other producers as well. In a similar vein, Klepper (1996) argues that larger firms will find it more profitable to do R&D because they can spread the result over a larger number of units, and eventually drive smaller firms out through a series of cost-reducing innovations, which amounts to there being dynamic increasing returns at the firm level during an industry's early stage.

Exit after learning through experience.—The shakeout usually comes after a wave of entry. That was clearly the case in the automobile and the automobile-tire industries, for instance. Most of the entrants in the entry wave would, some time later, find out that their ideas and methods, their products and technologies they used, were ill-suited for the industry at hand, and would consequently exit. Jovanovic (1982) argued that this mechanism could explain some other phenomena, but Horvath, Schivardi and Woywode (2003) point out that it can also help explain a shakeout conditional on there being a run-up in entry at some point in the industry's life.

Consolidations for other reasons.—Firms that exit during a shakeout are often acquired by other firms in the same industry. Consolidations and merger waves can occur for reasons unrelated to drops in demand and to advances in technology. Deregulation, for instance, has led to merger waves in the airlines and banking industries (Andrade *et al.* 2001), and both have consequently seen sharp reductions in the numbers of producers.

There is much to be said for each of these explanations. One obvious explanation for creative destruction episodes seems, however, to have gone unnoticed: Producers exit an industry when their capital comes up for replacement. Before its optimal replacement date, the sunk costs in the capital stock will tend to keep a firm in the industry even if, on other grounds, it may prefer to exit. So, for example, when a plant reaches the age when it needs to be replaced, its owner may decide that the moment is then right to move out of the industry and let some other firm – incumbent

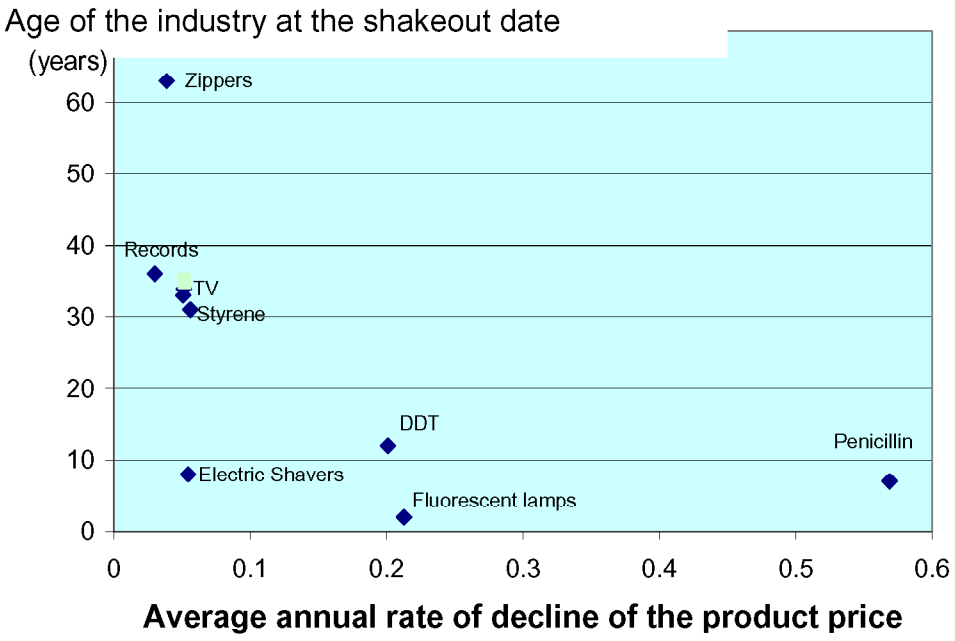


Figure 1: TECHNOLOGICAL PROGRESS AND INDUSTRY AGE AT SHAKEOUT

or a new entrant – provide the new capacity. Viewed in this way, a shakeout is but an "echo" of the burst of investment that occurs when an industry comes into being. This explanation is suggested by the pattern shown in Figure 1. The pattern is that the shakeout occurs earlier in those industries in which technological progress – as measured by the rate at which its product price declined prior to the shakeout – is faster.[1]

All technological change in our model is embodied in capital, which means that TFP should be constant when one adjusts inputs for quality. During the shakeout, a fraction of the capital stock is replaced by new capital; the number of efficiency units of capital stays the same but the productivity of capital per *physical* unit rises.

Our model is similar to those of Mitchell (2002) and Aizcorbe and Kortum (2004) who analyze industry equilibrium in steady state, i.e., the stationary case in which the effect of initial conditions has worn off, and after any possible spikes that may be caused by initial conditions have worn off. By contrast, we focus on the dynamics leading up to the first investment echo, and on whether its size and timing can explain the shakeout patterns from the Gort-Klepper data. Thus it is the same model, but used in a different way and applied to a different set of facts. Jovanovic and Lach

[1] Table 7 of Gort and Klepper (1982) reports the number of years until shakeout for 39 industries. But only for 8 of them does Table 5 report data on the rate of price decline up to the shakeout. These are the eight reported in Figures 1 and 3.

(1989) analyze transitional dynamics in a vintage-capital model, but are not able to generate a net decline in the number of producers.[2]

Plan of the paper.—Section 2 presents the model and states two propositions. The tests of those two propositions are described in Section 3. Section 4 presents some other properties of the model and Section 5 concludes the paper. The Appendix contains some proofs.

2 Model

We consider a small industry that takes the rate of interest and the price of its capital as given at unity. The product is homogeneous, and technology improves exogenously at the rate g. To use a technology of vintage t, however, a firm must buy capital of that vintage. The productivity of vintage-0 capital is normalized to 1, and so the productivity of vintage-t capital is e^{gt}.

Each firm is of measure zero, and so it takes prices as given. Let $p(t)$ be industry price, $q(t)$ industry output, $D(p,t)$ the demand curve, and r the constant interest rate. The demand curve is undefined for $t<0$ because no one is aware of the market before it has come into being at date zero.[3] To avoid having to worry about corners, we also assume that $D(0,t)=\infty$ for all t.

The price of capital is unity. Capital must be maintained at a cost of c per unit of time; c does not depend on the vintage of the capital, nor on time. Capital cannot be resold; it has a salvage value of zero.

Let $K(s,t)$ be the date-t stock of capital of vintage s or older, not adjusted for quality. There is no labor. Industry output at time t then equals the total number of efficiency units of capital:

$$q(t)=\int_{\max(0,t-T)}^{t} e^{gs}K(ds,t), \tag{1}$$

where T is the age at which a unit of capital is retired – this number turns out to be a constant in equilibrium, so we shall simply assume this in the notation henceforth. We shall then verify, in Proposition 1, that a constant-T equilibrium exists.

Evolution of the capital stock: No capital is in the industry at date zero. Capital evolves smoothly except for a countable set of dates, at which a positive fraction of the industry's capacity is replaced, and at which we observe an investment spike. Since the inter-spike waiting times are T, and since the first spike occurs at $t=0$, the date of the i'th investment spike is $(i-1)T$, for $i=1,2,....$ Let $i(t)=\max\{i \mid (i-1)T\leq t\}$

[2] Nor, therefore, a shakeout in Gort and Klepper's sense. See footnote 6 of Jovanovic and Lach's paper.

[3] Alternatively, we may assume that the demand for the product is in place before date zero and that people are aware of this, but that the production of the good becomes feasible only at date zero.

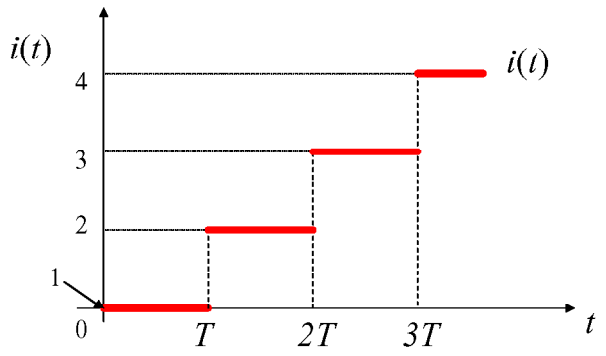


Figure 2: THE BEHAVIOR OF THE FUNCTION $i(t)$.

be the integer index of the most recent spike. We illustrate the function $i(t)$ in Figure 2.

We shall show that in equilibrium, all the capital created at one spike date is replaced at the following spike date. Therefore the capital stock never embodies capital created at more than a single spike date. For calculating the capital stock, that is, only the immediately preceding spike will matter. Let X_i denote the size of the i'th investment spike. At date t, then, the amount of capital accounted for by the last spike is $X_{i(t)}$. So referring to Figure 2, for $t \in (2T, 3T)$, say, the contribution of capital created by the last spike would be X_{2T}. In general, the date-t distribution of capital by vintage is

$$K(s,t) = I_{\{s-(i[t]-1)T \geq 0\}} X_{i(t)} + \int_{\max(0,t-T)}^{s} x(u)\, du, \tag{2}$$

where I_A is the indicator function of the set A.

Equilibrium.—The definition of equilibrium is simple if $x(t) > 0$ for all t (we shall later provide conditions – in (10) – that guarantee this outcome). Such an equilibrium consists of a product price function $p(t)$, a retirement-age of capital, T, investment flows $x(t)$, and investment spikes X_i accruing at dates $(i-1)T$ $(i = 1, 2, ...)$ that satisfy

1. *Optimal retirement of capital*: Since price declines monotonically, it is optimal to replace vintage-t capital as soon as its revenue equals its maintenance cost:

$$e^{gt} p(t+T) = c. \tag{3}$$

2. *Optimal investment*: If investment $x(t) > 0$, the present value of a new capital good must equal its cost. The present value of the net revenues derived from that (vintage-t) unit of capital was

$$1 = \int_{t}^{t+T} e^{-r(s-t)} \left(e^{gt} p(s) - c \right) ds. \tag{4}$$

If $x(t) = 0$ at some dates, (4) may become a strict inequality.

3. *Market clearing*:

$$D(p[t], t) - e^{gi(t)T} X_{i(t)} + \int_{\max(0,t-T)}^{t} e^{gs} x(s)\, ds. \tag{5}$$

Existence of a constant-T equilibrium.—Because the supply of entrants is infinitely elastic, the time path of $p(t)$ depends on the cost side alone. We now show that $p(t)$ declines at the rate g. That is, for some initial price p_0,

$$p(t) = p_0 e^{-gt}. \tag{6}$$

In that case, (3) implies

$$p_0 e^{-gT} = c, \tag{7}$$

while (4) implies

$$1 + c\frac{1 - e^{-rT}}{r} = p_0 \frac{1 - e^{-(r+g)T}}{r+g}. \tag{8}$$

Equations (7) and (8) are to be solved for the pair (p_0, T). If we eliminate p_0, we reach the implicit function for T alone:

$$\left(\frac{r+c}{c}\right)(r+g) = ge^{-rT} + re^{gT}, \tag{9}$$

for $t \geq 0$.

Proposition 1 *Eq. (9) has unique solution for $T \equiv \tilde{T}(r, c, g)$.*

Proof. The LHS of (9) is constant and exceeds $r+g$. On the other hand, the RHS of (9) is continuous and strictly increases from $r+g$ to infinity, having the derivative $rg\left(e^{gT} - e^{-rT}\right) > 0$ for $T > 0$. Therefore the two curves have exactly one strictly positive intersection. ■

Given this result it then follows from (7) that

$$p_0 = ce^{g\tilde{T}(r,c,g)}.$$

Sufficient conditions for $x(t) > 0$ for all t.—Proposition 1 relies on the assumption that investment is always positive. A necessary and sufficient condition for that to be true is that output be increasing, which is another way of saying that capacity is always expanding. That condition is

$$\frac{d}{dt} D(p, t) = \frac{\partial D}{\partial p}\frac{dp}{dt} + \frac{\partial D}{\partial t} = gp\left|\frac{\partial D}{\partial p}\right| + \frac{\partial D}{\partial t} > 0. \tag{10}$$

never decrease in t. Thus it is possible for the time-derivative of D to be negative, but not too negative to offset the positive effect on output of the fall in price. Now, condition (10) is of limited value because it involves the endogenous variable p. One can, however, reduce (10) to a condition on primitives in some special cases. Take the case where population grows at the rate γ_t and where each consumer's demand is iso-elastic, i.e., the case where $D(p,t) = Ap^{-\lambda}\exp\left(\int_0^t \gamma_s ds\right)$, (10) is equivalent to

$$g\lambda > -\gamma_t$$

for all $t \geq 0$. Thus (10) can hold even if population is declining, as long as its rate of decline, γ_t, never exceeds $g\lambda$.

2.0.1 Two implications that we shall test

The next two propositions state the results that we shall test. The first result follows directly from implicitly differentiating (9) and is derived in the Appendix:

Proposition 2 *The replacement cycle is shorter if technological progress is faster:*

$$\frac{\partial T}{\partial g} = -\frac{\frac{r+c}{c} + (gT-1)e^{gT}}{g^2\left(e^{gT} - e^{-rT}\right)} < 0. \tag{11}$$

That industries with a high g should have a low T means that industries with higher productivity growth should have an earlier shakeout. This will be the first relation that we shall test.

The second relation emerges between two variables that in the model should be equal to each other. Let

$$y \equiv \frac{X_0}{q_T}$$

be the fraction of capacity replaced at the first shakeout, and let

$$z \equiv \frac{q_0}{q_T}$$

be the initial output of the industry relative to its output at the shakeout date. From (7), $p_0 = ce^{gT}$. Then from (5),

$$q_0 = D\left(ce^{gT}, 0\right) = X_0$$

Since *all* of the initial capacity is replaced at T, it must mean that $X_0 = q_0$ and that therefore

Proposition 3 *The fraction of capacity replaced at T depends inversely on output growth over the period $[0,T]$:*

$$y = z. \tag{12}$$

The restriction in (12) holds regardless of why output has grown. Output may have grown because the demand curve shifted out, or it may have grown because demand is elastic so that the price decline led output to grow. The end result in (12) is the same.

2.0.2 Some other implications

We also have an exact rate at which investment echoes or investment spikes die off. Let X_n be the size of the n'th investment spike. We then have

Proposition 4 *Investment spikes decay geometrically. That is,*

$$X_n = e^{-gTn} X_0 \tag{13}$$

for $n = 1, 2, ...$

Proof. Because $p(t)$ is continuous at T, the number of efficiency units replaced at the spikes is a constant, X_0. which means that $X_n = e^{-gT} X_{n-1}$. The absolute size of each spike is a fraction, e^{-gT}, of the previous one. ■

The spike dates remain T periods apart, at $0, T, 2T, 3T, ...$ The spikes occur regularly because technological progress occurs at the steady rate g. Thus we offer an explanation for periodic creative destruction episodes that, since it does not rely on recurring chocks, is different from the one that Caballero and Hammour offer. It also can be shown that if g were to decline with industry age, (T_n) would be a strictly increasing sequence.[4]

The remaining parameters of the model are the maintenance cost c and the rate of interest r. The following claim is proved in the Appendix:

Proposition 5

$$(i) \quad \frac{\partial T}{\partial c} < 0. \quad \textit{And, if } gc < r^2, \quad (ii) \quad \frac{\partial T}{\partial r} > 0.$$

A rise in the maintenance cost, c, reduces the lifetime of capital as one would expect. Since replacing capital constitutes an investment, when the rate of interest rises that form of investment is discouraged, and it will occur less frequently.

3 Tests

The Gort-Klepper data measure the industry's age from the date that the product was commercially introduced, i.e., from the date of its first sales. The shakeout period is defined as the epoch during which the number of firms is declining. An "exit" occurs when a producer of a product stops making it. The producer can continue to exist as a firm that makes other products.

[4]The proof takes up a lot of space and is available upon request. The proof is contained in Jovanovic and Tse (2005) which treats a more general model of the industry life cycle.

So far the model has only capital and no firms.[5] How does the replacement of capital relate to the exit of firms? We hypothesize that a fraction of machines replaced are in firms that will themselves exit and be replaced by new firms. The argument is that if the firm needed to exit for some unrelated reason, then the appropriate time to do so would be when its plant and equipment need to be replaced. This is more likely to be so if used capital markets do not function well, or if the firm to find it hard to sell off an unwanted division. Rather than replacing its equipment, the firm could then simply exit at T, and let others bring the new equipment into the industry.

Whether exit dates and optimal capital replacement dates do actually coincide, is hard to show directly. What the model does imply – and we shall elaborate on its productivity implications in Section 3.4 – is that it is the least efficient, the least high-tech production units that replace their capital. In line with our hypothesis, Baily *et al.* (1993, Table 3) show that the least productive plants are the most likely to exit and Deily *et al.* (2000, Table 2) show that the least-efficient private hospitals are more likely to close. Tying the exit decision to the age of capital is confounded by the tendency for the age of the capital stock to be correlated with experience. In the meat-packing industry, however, Anderson et al. (1998) find that after age twenty a plant's likelihood of exit increases steadily with its age. The authors conclude that newer meat-packing plants generally have superior capital than do older plants, but that the operators of older plants have greater experience, and that the positive effects of the latter dominate technology effects until the plant is about twenty years old. Even when we proxy for the age of the capital by the age of the firm[6], Agarwal and Gort (1996, Figure 2) show that hazard rates for *firm* exit tend to rise once age exceeds 30 or 40 years.

As suggested by this discussion, we shall proxy the age, T, of an industry at its first replacement spike, by the date at which the shakeout of firms in that industry begins, and the rate of technological progress, g, by the rate at which the price of the product declines. The data are taken from Gort and Klepper (1982).

[5] One way to introduce firms into the model without changing its implications for industry-wide investment is as follows. Assume that each firm can own exactly one unit of capital, and that a firms dies with its capital. Since output is continuous at T, this means that at that date X_0 producers are replaced by $e^{-\gamma T} X_0$ new producers. Then gross exit is X_0 firms, and net exit is $\left(1 - e^{-\gamma T}\right) X_0$ firms. As a fraction of the peak number of producers, net exit would be

$$y \equiv \frac{\left(1 - e^{-\gamma T}\right) X_0}{X_0 + \int_0^T x(t)\, dt}. \tag{14}$$

[6] In our context this would be an imperfect proxy simply because firms may enter laterally from other markets, and the capital stock they would buy for the market at hand would be younger than the firm itself.

3.1 Testing (11): The timing of the shakeout

Since replacement episodes are in the model caused by technological progress, we first check if industries with higher productivity growth experience earlier shakeouts. That is, we ask whether industries with a high g have a low T as Proposition 2 claims and, if so, how well the solution for T to (9) fits the cross-industry data on g and T. Gort and Klepper report eight industries for which they can measure the rate of price declines over the entire pre-shakeout stage; we reproduce the numbers in Table 1:

PRODUCT NAME	$\hat{g}$	$\hat{T}$
DDT$^{(-)}$	0.20	12
Electric shavers$^{(-)}$	0.05	8
Fluorescent lamps$^{(+)}$	0.21	2
Penicillin$^{(+)}$	0.57	7
Phonograph records$^{(-)}$	0.03	36
Styrene$^{(+)}$	0.06	31
Television$^{(-)}$	0.05	33
Zippers$^{(-)}$	0.04	63

Table 1: THE DATA IN FIGURES 1 AND 3.

We now describe the procedure by which we chose the model's parameters. By Proposition 1, a unique solution to (9) for T exists, denoted it by $\tilde{T}(r, c, g)$. We estimate T to be the date that Gort and Klepper find that Stage 4 (the shakeout stage) begins in their various industries. For industry i we call this variable $\hat{T}_i$. In view of (6), we estimate g by the rate of product-price decline in that industry. For industry i we call this variable $\hat{g}_i$. We do not have observations on r and c, so we set $r = 0.07$ and estimate c to minimize the sum of squares

$$\min_c \left\{ \sum_{i=1}^{9} \left[\hat{T}_i - \tilde{T}(0.07, c, \hat{g}_i) \right]^2 \right\}.$$

The estimate is $c = 0.039$ with a standard error of 0.013.

The the data and the implicit function $\tilde{T}(g, r, c)$ fitted to them are shown in Figure 3. The fit is quite good, in that the $R^2 = 0.53$. Thus, the capital-replacement motive seems to partly explain the timing of the shakeout.

3.1.1 Alternative explanations for the negative $\hat{T} - \hat{g}$ relation

Demand declines.—On the face of it, this seems like a plausible alternative explanation for the negative relation between $\hat{T}$ and $\hat{g}$. The gist of the argument can be

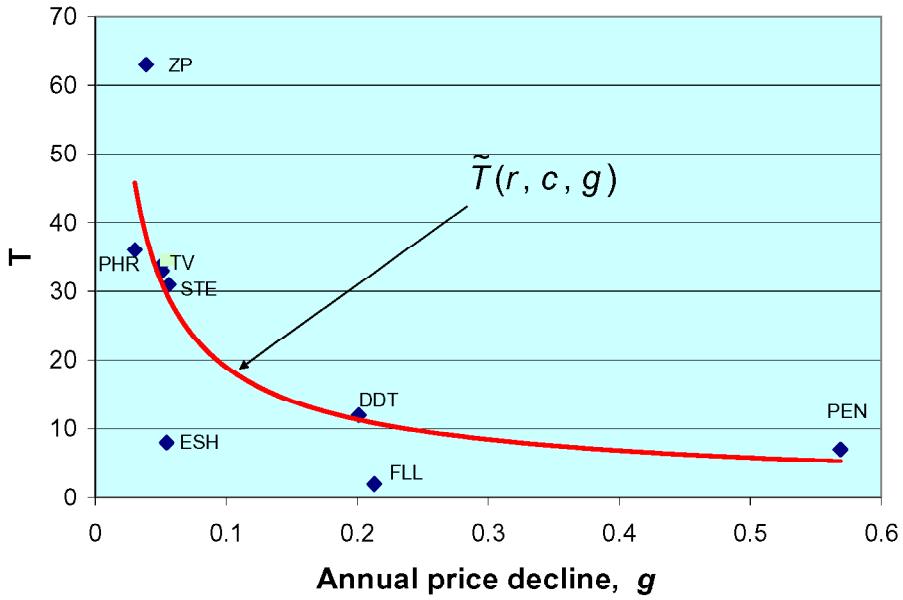


Figure 3: THE RELATION BETWEEN $\hat{T}$ AND $\hat{g}$

collapsed into a two-period example, suppose demand either stays constant or that it declines, and that the outcome is unpredictable. Then a demand decline produces a shakeout ($T = 2$) and a decline in p, whereas without a decline there is no shakeout ($T = +\infty$) and no decline in p. This seemingly plausible argument does not, however, meet with support in GK's sample. Among the industries where output fell during the shakeout, which in Table 1 are marked with a superscript $^{(-)}$, the average pair $\left(\hat{g}^{(-)}, \hat{T}^{(-)}\right) = (0.07, 30.4)$, whereas among the industries where output rose, marked with a superscript $^{(+)}$, the average pair $\left(\hat{g}^{(+)}, \hat{T}^{(+)}\right) = (0.28, 13.3)$. Yet the demand hypothesis implies the opposite: Prompted by the decline in output, $\hat{g}^{(-)}$ should have exceeded $\hat{g}^{(+)}$, and $\hat{T}^{(-)}$ should have been lower than $\hat{T}^{(+)}$.

Exit after learning through experience.—The argument of Horvath *et al.* (2003) would need to be augmented with some reason for why a run-up in entry should come sooner in those industries where p is declining rapidly. Such an argument would then explain why the shakeout itself would occur earlier.

Consolidations for other reasons.—Again, one would need a reason for why consolidations should occur sooner in industries where p is falling faster.

Technological change.—Our argument uses technological change in a vintage-capital context. Both components are needed for the conclusions to hold. We would argue that the second component, which is missing in the other papers we listed under this heading, is essential. It is the new vintages of capital that must replace

old vintages sooner when technological progress is faster. Progress manifests itself through a falling p, and so the relation emerges in a straightforward way. It may be possible to extend the arguments of Hopenhayn, Klepper, Jovanovic and MacDonald and come up with an explanation not involving vintage capital, but that remains to be seen.

3.2 Testing (12)

Testing Proposition 3 requires that we estimate the fraction of capacity replaced and the growth of output when the industry reaches age T. Equation (12) says that if we regress y on z we should get a perfect fit. We calculate $\widehat{z}_i$ from GK's table 5 as the inverse of the rate of output increase from the date industry i was born up to the date when the number of firms in the industry peaked. But we do not have data on capacity replaced; we do have estimates on the fraction of firms that exit during the shakeout. We calculate $\hat{y}_i$ = GK's Table 4 column 3.[7] This is the total decrease in the number of firms for a period of time that lasted for on average 5.4 years, after which the number of firms in the industry peaked. Hence $\widehat{y}_i$ is the upper bound on the number of firms that exited at time T.

Because exiting firms are smaller than other firms, the fraction of firms that exits does not equal the fraction of industry capacity withdrawn by the exiting firms: Evans (1987) and Dunne, Roberts and Samuelson (1989) find that firm size is positively associated with survival. Since small firms have a greater tendency to exit than large firms, the fraction of capacity that is scrapped is smaller than the fraction of firms that exit. Table 2 of Dunne, Roberts and Samuelson (1989) shows that the size of exiting firms is about 35 percent of the size of non-exiting firms. Therefore the fraction of firms should be roughly three times the fraction of capacity replaced. In other words, Proposition 3 together with the size adjustment implies the regression equation $\hat{y}_i = 3\hat{z}_i + u_i$, where i denotes the industry. Firms may also exit for reasons that are outside the model.[8] Therefore we add a constant, α, to the RHS of the regression equation in order to represent these other forms of exit. Thus, we estimate the relation

$$\hat{y}_i = \alpha + \beta\hat{z}_i + u_i,$$

[7] This is net exit, whereas we would prefer to use gross exit as a measure of y. However,

$$\text{gross exit} = \text{net exit} + \text{entry},$$

and Table 3 of Agarwal and Gort (1996) shows no significant cross-industry correlation of entry and exit rates. This implies that net exit is an unbiaserd measure of gross exit.

[8] In our model, it is always the oldest capital that is replaced and, on these grounds, the oldest firms that exit. Now Figure 2 of Agarwal and Gort (1996) shows that this is not true in any of the five stages. The exit hazard does tend to rise beyond a certain age, however, indicating an additional reason for exit among the oldest firms in the sample. Nevertheless, the fraction of firms exiting is, at best, a noisy measure of the fraction of the capacity replaced, and that is how we interpret it

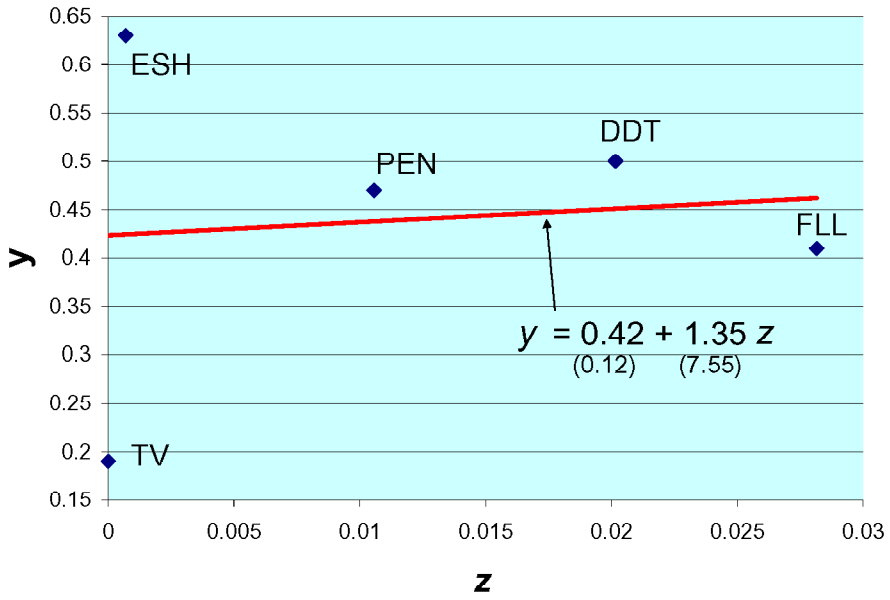


Figure 4: THE RELATION BETWEEN y AND z.

where β represent the fraction of firms that exit in response to capacity replacement.

The estimates are in Figure 4.[9] There are too few observations to allow us to tell if β differs significantly from 3.

3.3 Repeated echoes in Proposition 4

If the replacement of capital was the only reason for exit, and if industries were not subject to other shocks, then we should indeed expect to find shakeouts that diminish geometrically in absolute terms as (13) shows. However, Gort and Klepper did not report second shakeouts. This may partly be because noise in demand and costs caused investment to deviate from what our model predicts, and that the deviations would amplify with the passage of time and, hence, with industry age. It also may in part be because the Gort-Klepper data rarely cover industry age to the point $t = 2T$ where we ought to observe a second shakeout. That is, the data reveal at most one shakeout per industry, and this is because the first shakeout comes, in most cases, well after the midpoint of the sample period.

In 9 out of the 46 Gort-Klepper industries, however, data were collected at least up

[9] G-K's Table 5 reports complete information on the rate of output increase up to the shakeout for 10 industries. But only for five of these is there also information on the rate of firm exit during the shakeout in G-K's table 4.

to the point $t = 2T$ where we ought to observe a second shakeout. Even though GK do not report a second shakeout for any of these industries, one may have been hidden in the data. In 6 out of these 9 industries, the average annual rates of net exit were negative for the period of time after the first shakeout. The temporal concentration of these exits around a short period of time would point to the existence of a second shakeout. These second and subsequent concentrated spells of exits, if any of them did occur, apparently were not large enough for Gort and Klepper to label as shakeouts. Furthermore, if we have allowed for g to decline over time, the second shakeout would only come at some time $t > 2T$. In this case, the time for the second shakeout might not yet have come for a number of these 9 industries. Some evidence for industry-wide retooling spikes is in Cooper and Haltiwanger (1993).

There are forces that would cause the spikes to dissipate fairly rapidly. First and foremost, gT ranges from 0.43 (Fluorescent Lamps) to 3.98 (Penicillin), with a median of 1.68 (Television). This means that the amount of capacity replaced at the second shakeout in the median industry should have been a fraction $e^{-1.68} = 0.19$ of the amount replaced at the first shakeout.

Second, there may be capital-adjustment costs even at the industry level, such as are modeled by Caballero and Hammour (1994), so that the cost of a capital is an increasing function of the rate at which capital is being brought into the industry. If this modification were introduced into the present model, then the X_n would cease to be spikes and would start to look like waves. The higher the adjustment cost, the less spiky and more bell shaped the investment echoes would appear. If T were firm-specific because, say, c were an i.i.d. over time but firm-specific random variable, then even if the initial investment was truly a spike, the first echo would have a bell shape, the second a more dispersed shape, and so on, so that, depending on how dispersed the c's were, the echoes could rapidly go away.

A second reason why echoes would dissipate is that many exiting firms do not simply dissolve and auction off their assets, but are instead purchased by other industry incumbents. These incumbents now would own capital of more than one vintage, and would be less likely to exit when only a subset of their capital stock was to be replaced. [10]

3.4 Implications for productivity growth

The model has no labor, and so TFP is simply the productivity of capital.[11] If the price of capital used to deflate investment spending were per unit of quality, then all the technological progress would appear in the capital stock, the resulting measure

[10] In this case, however, we still ought to see recurring investment spikes characterized by (13).

[11] One way to add labor to the model is to have one worker per machine, and to interpret w as the wage. Labor productivity would then be the same as the productivity of the physical units of capital and would equal the expression in (15). A shakeout would be accompanied by a sharp decline in employment.

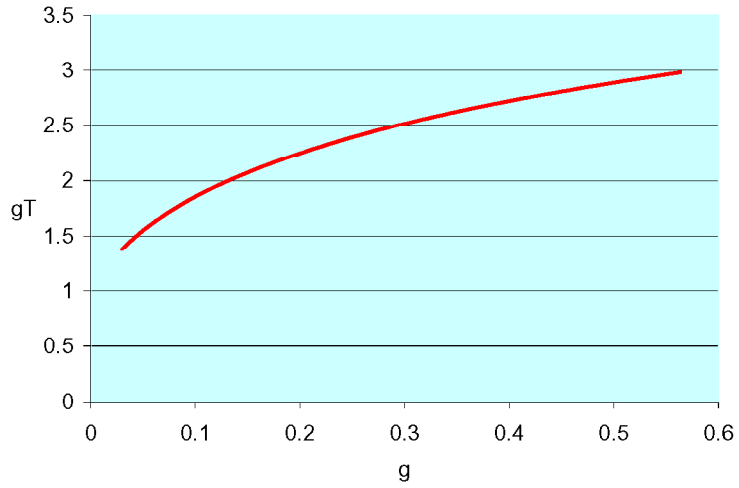


Figure 5: TFP GROWTH AND THE MAX/MIN TFP RATIO

of the capital stock would be given by (1) and would equal output, with TFP being constant at unity. On the other hand, if no adjustment is made for quality so that the price index used to deflate investment was unity, then the capital stock would be given by $K(t,t)$ as given by (2) when evaluated at $s = t$, and TFP would equal

$$\text{TFP}_t = \frac{q_t}{K(t,t)}. \tag{15}$$

Thus TFP rises smoothly until the shakeout date, and then it experiences an upward jump during the shakeout when the number of physical units of capital falls, but the number of their efficiency units stays unchanged.

3.4.1 TFP growth vs. dispersion

Faster TFP-growing industries should show greater TFP dispersion in our model. Dwyer (1998) has derived the same relation in a similar model. He also estimated the relation in a group of textile industries. The relation was positive and significant.[12] Dwyer measured dispersion by the TFP ratio of the tenth percentile plants to the ninetieth percentile plants. TFP growth ranged from about two percent to about eight percent and the TFP ratio ranged between 2.4 to 4.6.

Our model cannot generate such large dispersion, even when we measure dispersion as the ratio of the most productive to the least productive producer. Figure 5 shows the comparative steady state relation. The relation is positive, but for the

[12] At the two-digit level, Oikawa (2006) finds a positive relation between TFP growth and TFP dispersion measured by the standard deviation of the logs.

less-than-ten-percent range of TFP growth (which would include all of Dwyer's industries), the most we can explain is a ratio of about 1.75.

A further difficulty with the model as an explanation of TFP dispersion is that the dynamics of TFP do not quite match the dynamics of actual plants in the data. In a word, too much *leapfrogging* goes on. In the data TFP-rank reversals are common, but they typically do not give the "last shall be first" implication. In our model, as in many other vintage-capital models, when a producer updates his capital, his productivity rises from the last percentile to the first percentile. A plant would move smoothly down the distribution of percentiles until it reached the last percentile, and it then would suddenly jump back up to the first percentile, and so on. Analysis of productivity transitions by Baily, Hulten, and Campbell (1993, Table 3) does not bear this out. Plants move up and down the various quintiles of the distribution and the transition matrix is fairly full. Moreover, births tend to be below average productivity.

On the positive side, a supporting fact in Baily *et al.*'s transition matrix is that more plants (14 percent) move from the bottom quintile to the top quintile than the reverse (5 percent). One reason why we do not see the dramatic "last shall be first" rank reversals is the tendency for a plant's TFP to grow as it ages.[13] Thus a new, cutting-edge plant does not immediately have the highest productivity. Rather, its productivity grows as it ages, as Bahk and Gort (1993), and Table 5 of Baily *et al.* (1993) documents. This would explain why a new plant does not enter at the top of the distribution of TFP in the way that this model predicts and why, as Table 5 of Baily et al. (1993) shows they tend to rise from the bottom to the top TFP quintile by the time they are 10-15 years old.

3.4.2 Transitional dynamics in productivity dispersion

The model implies that in the initial stages of an industry's life, the distribution of TFP across producers should be fanning out. How fast it does so, however, should depend on the rate of technological progress, i.e., on g. This is the rate at which the productivity of new capital gains at the expense of old capital. But after the industry reaches age T, no further fanning out takes place, because old capital begins to be withdrawn. Thereafter, dispersion remains constant at its steady-state level. In Figure 6 we show how the fanning out process depends on the industry's growth rate.

[13] Parente (1994) models learning of the technology with the passage of time, and Klenow (1998) models learning as a function of cumulative output.

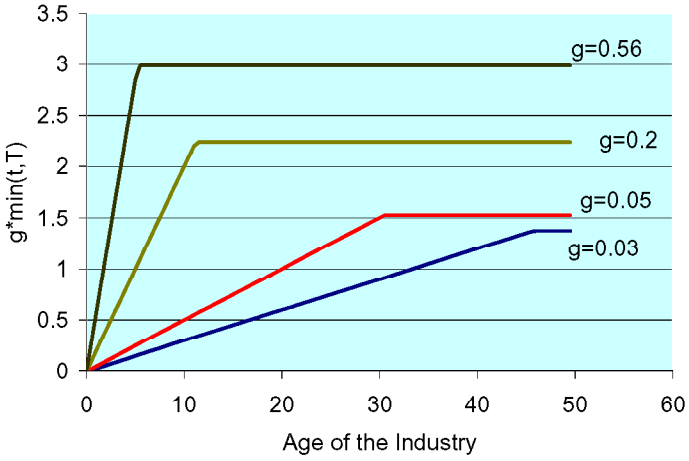


Figure 6: TRANSITIONAL DYNAMICS IN PRODUCTIVITY DISPERSION

4 Conclusion

We have shown evidence that shakeouts of firms occur earlier in industries where technological progress is faster. Evidence also shows that output continues to rise through the shakeout episodes, suggesting that shakeouts are replacement episodes. We presented a vintage-capital model that seemed to fit the main fact, namely that shakeouts occur earlier in industries in which the price of the product declines more rapidly. We briefly outlined several other arguments for shakeouts and found that they were not able to explain this fact.

A question on the shakeout that remains open is the actual mode of producer exit and its determinants. A firm shut down altogether, and all its assets may be dispersed. Alternatively, the firm may cease producing a good but may stay in business producing other goods. Finally, it may be acquired by a larger firm. The need to replace its capital may prompt any of the three modes of exit, but clearly a richer model is needed to analyze such things.

References

[1] Agarwal Rajshree, and Michael Gort "The Evolution of Markets and Entry, Exit and Survival of Firms." *Review of Economics and Statistics* 78, No. 3. (Aug., 1996): 489-98.

[2] Aizcorbe, Ana and Samuel Kortum. "Moore's Law and the Semiconductor Industry: A Vintage Model." December 2004.

[3] Anderson, Donald, Brian Murray, Jackqueline Teague and Richard Lindrooth. "Exit from the Meat-packing Industry: A Microdata Analysis." *American Journal of Agricultural Economics* 80, no. 1 (February 1998): 96-106.

[4] Andrade, Gregor; Mitchell, Mark and Stafford, Erik. "New Evidence and Perspective on Mergers." *Journal of Economic Perspectives,* Spring 2001, 15(2), pp. 103-120.

[5] Atkeson, Andrew and Patrick Kehoe. "Industry Evolution and Transition: The Role of Information Capital." FRB Minneapolis staff report 162 (August 1993).

[6] Byong-Hyong Bahk; Michael Gort. "Decomposing Learning by Doing in New Plants." *Journal of Political Economy*, Vol. 101, No. 4. (Aug., 1993), pp. 561-583.

[7] Caballero, Ricardo and Mohamad L. Hammour. "The Cleansing Effect of Recessions." *American Economic Review* 84, no. 5 (December 1994):1350-68.

[8] Deily, Mary, Niccie McKay and Fred Dorner. "Exit and Inefficiency: The Effects of Ownership Type." *Journal of Human Resources* 35, no. 4 (Autumn, 2000): 734-747.

[9] Dunne, Timothy, Mark Roberts; Larry Samuelson. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *RAND Journal of Economics* 19, no. 4 (Winter, 1988): 495-515.

[10] Dwyer, Douglas. "Technology Locks, Creative Destruction, and Nonconvergence in Productivity Levels." *Review of Economic Dynamics* 1, no. 2 (April 1998): 430-473.

[11] Evans, David. "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries." *Journal of Industrial Economics* 35, No. 4 (June 1987): 567-81.

[12] Gort, Michael and Steven Klepper "Time Paths in the Diffusion of Market Innovations," *Economic Journal* 92 (1982): 630-53.

[13] Hopenhayn, Hugo. "The Shakeout." wp#33 Univ. Pompeu Fabra, April 1993.

[14] Horvath, Michael, Fabiano Schivardi and Michael Woywode. "On Industry Life-Cycles: Delay, Entry, and Shakeout in Beer Brewing." *International Journal of Industrial Organization* 19, no. 7 (2003): 1023-1052.

[15] Jovanovic, Boyan. "Selection and the Evolution of Industry." *Econometrica* 50, no. 3 (June 1982): 649-70

[16] ———, and Saul Lach "Entry, Exit, and Diffusion with Learning by Doing." *American Economic Review* 79, no. 4 (September 1989): 690-99.

[17] ———, and Glenn MacDonald. "The Life Cycle of a Competitive Industry." *Journal of Political Economy* 102 (April 1994): 322-47.

[18] ———, and Peter L. Rousseau. "General-Purpose Technologies." *Handbook of Economic Growth* vol. 1B Amsterdam: Elsevier (2006): 1181-224.

[19] ———, and Chung-Yi Tse. "Vintage Capital at the Industry Level." February 2006.

[20] Klenow, Peter. "Learning Curves and the Cyclical Behavior of Manufacturing Industries." *Review of Economic Dynamics* 1, no. 2 (May 1998): 531-50.

[21] Klepper, Steven, "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *American Economic Review* 86 no. 3 (June 1996): 562 - 83.

[22] ———, and Kenneth Simons. "Industry Shakeouts and Technological Change." October 2001.

[23] Mitchell, Matthew. "Technological Change and the Scale of Production." *Review of Economic Dynamics* 5, no. 2 (April 2002): 477-88.

[24] Oikawa, Koki. "Uncertainty-Driven Growth." New York University, January 2006.

[25] Parente Stephen L., 1994. "Technology Adoption, Learning-by-Doing, and Economic Growth," *Journal of Economic Theory*. Elsevier, vol. 63(2), pages 346-369.

[26] Utterback, James, and Fernando Suárez, "Innovation, Competition, and Industry Structure," *Research Policy* 22 (1993): 1-21.

5 Appendix

Proof of Proposition 2.—Rewrite (9) as

$$ge^{-rT} + re^{gT} - \left(\frac{r+c}{c}\right)(r+g) \equiv \Phi(T, g, c, r) = 0.$$

We have $\frac{\partial \Phi}{\partial T} = rg\left(e^{gT} - e^{-rT}\right) > 0$. Therefore (11) follows if $\frac{\partial \Phi}{\partial g} > 0$. Now

$$\begin{aligned}
\frac{\partial \Phi}{\partial g} &= e^{-rT} + rTe^{gT} - \frac{r+c}{c} \text{ and, by eliminating } e^{-rT}, \\
&= \left(\frac{r+c}{c}\right)\frac{r+g}{g} - \frac{r}{g}e^{gT} + rTe^{gT} - \frac{r+c}{c} \\
&= \left(\frac{r+c}{c}\right)\frac{r}{g} - \frac{r}{g}e^{gT} + rTe^{gT} = \frac{r}{g}\left(\frac{r+c}{c} - e^{gT}\right) + rTe^{gT} \\
&= \frac{r}{g}\left(\frac{r+c}{c} + (gT-1)\,e^{gT}\right) > 0.
\end{aligned}$$

The strict inequality follows because (i) The function $(gT-1)\,e^{gT}$ is increasing in gT, with derivative $e^{gT}\left[1+(gT-1)\right] = gTe^{gT}$, and because (ii) As $gT \to 0$, $(gT-1)\,e^{gT} \to -1$, so that at its smallest point, $\frac{\partial \Phi}{\partial g} = \frac{r^2}{gd} > 0$. Then, since $\frac{\partial T}{\partial g} = -\frac{\partial \Phi}{\partial g}/\frac{\partial \Phi}{\partial T}$, and since $\frac{\partial \Phi}{\partial T} > 0$, (11) follows. ■

Proof of Proposition 4.—(i) We showed that $\Phi_T > 0$ in the proof of Proposition 2. Moreover, $\frac{\partial \Phi}{\partial c} > 0$ because the ratio $\frac{r+c}{c}$ is decreasing in c. Therefore $\frac{\partial T}{\partial c} = -\frac{\partial \Phi}{\partial c}/\frac{\partial \Phi}{\partial T} < 0$. (ii) When $gc < r^2$, $\frac{\partial T}{\partial r} = -\frac{\partial \Phi}{\partial r}/\frac{\partial \Phi}{\partial T}$

$$\begin{aligned}
\frac{\partial \Phi}{\partial r} &= -Tge^{-rT} + e^{gT} - \left(\frac{r+g}{c}\right) - \left(\frac{r+c}{c}\right) \text{ and, by eliminating } e^{gT}, \\
&= -Tge^{-rT} - \frac{ge^{-rT}}{r} + \left(\frac{r+c}{c}\right)\frac{(r+g)}{r} - \left(\frac{r+g}{c}\right) - \left(\frac{r+c}{c}\right) \\
&= -Tge^{-rT} - \frac{ge^{-rT}}{r} + \left(\frac{r+c}{c}\right)\frac{g}{r} - \left(\frac{r+g}{c}\right) \\
&= \frac{g}{r}\left(1 - e^{-rT}\right) - \left(\frac{r}{c} + Tge^{-rT}\right) < \frac{g}{r} - \frac{r}{c}.
\end{aligned}$$

so that then $\frac{\partial T}{\partial r} = -\frac{\partial \Phi}{\partial r}/\frac{\partial \Phi}{\partial T} > 0$. ■