NBER WORKING PAPER SERIES

CREATIVE DESTRUCTION IN INDUSTRIES

Boyan Jovanovic
Chung-Yi Tse

Working Paper 12520
http://www.nber.org/papers/w12520

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2006

We thank M. Gort, S. Greenstein, C. Helfat, O. Licandro, S. Klepper, and S. Kortum for comments, R. Agarwal for comments and for data, and the NSF for support. The views expressed herein are those of the author(s) and d not necessarily reflect the views of the National Bureau of Economic Research.

Creative Destruction in Industries
Boyan Jovanovic and Chung-Yi Tse
NBER Working Paper No. 12520
September 2006
JEL No. L11

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 shakeouts seem to be classic creative destruction episodes. Shakeouts of firms tend to occur sooner in industries where technological progress is rapid. Existing models do not explain this. Yet the relation emerges naturally in a vintage-capital model in 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 to the Gort-Klepper data and to Agarwal's update of those data.

Boyan Jovanovic
New York University
Department of Economics
269 Mercer Street
New York, NY 10003
and NBER
Boyan.Jovanovic@nyu.edu

Chung-Yi Tse
School of Economics and Finance
University of Hong Kong
Pokfulam Road
Hong Kong
tsechung@econ.hku.hk

Creative Destruction in Industries*

Boyan Jovanovic and Chung-Yi Tse†

June 11, 2007

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 shakeouts seem to be classic creative-destruction episodes.

Shakeouts of firms tend to occur sooner in industries where technological progress is rapid. Existing models do not explain this. Yet the relation emerges naturally in a vintage-capital model in 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 to the Gort-Klepper (1982) data and to Agarwal's (1998) update of those data.

1 Introduction

Schumpeter argued that creative destruction happens in waves. By "creative destruction" he meant the replacement of old capital – physical and human – by new capital and of old firms by new firms. The driving force behind these waves of creative destruction is technological change, thought Schumpeter.

Whether such creative-destruction waves are to be seen at the aggregate level or not, they have been documented in many industries. The "shakeout" episodes that Gort and Klepper (1982, henceforth GK) document 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 and new entrants.

*We thank A. Gavazza, M. Gort, S. Greenstein, C. Helfat, O. Licandro, S. Klepper, S. Kortum and M. Kredler for comments, R. Agarwal for providing data, and the NSF for support.

†NYU and Hong Kong University, respectively.

Several explanations exist for these creative-destruction episodes. When the entire economy is involved, the shock that prompts the creative-destruction 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). At the level of an individual industry, there are four sets of explanations:

(*i*) *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. Our explanation does not rely on recurring shocks to generate recurring replacement.

(*ii*) *Technological advances.*—When a new technology raises the efficient scale of firms, it crowds some of them out of the industry; so argue Hopenhayn (1993) and Jovanovic and MacDonald (1994). The assembly-line technology, e.g., probably 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 that a dominant design emerges: The winning model forces out other models and their producers. In a similar vein, Klepper (1996) argues that larger firms do more R&D than small firms because they can spread its results over a larger number of units; because they invent at a faster rate, large firms then drive smaller firms out.

(*iii*) *Exit after learning through experience.*—GK observed that a shakeout usually comes after a wave of entry. Horvath, Schivardi and Woywode (2003) argue that if a run-up in entry occurs at some point in the industry's life, then some time later, a fraction of the entrants will have found themselves unfit to be in that industry – a type of learning that Jovanovic (1982) stresses – and will then exit *en masse*.

(*iv*) *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 to a sharp fall in the numbers of producers.

Our hypothesis.—One obvious explanation for creative-destruction episodes seems, however, to have gone unnoticed: Producers sometimes 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. But at the replacement date, the firm may not be able to finance the new plant and equipment, or for some other reason the entrepreneur may then decide that the moment is then right to move out of the industry and let some other firm – incumbent 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

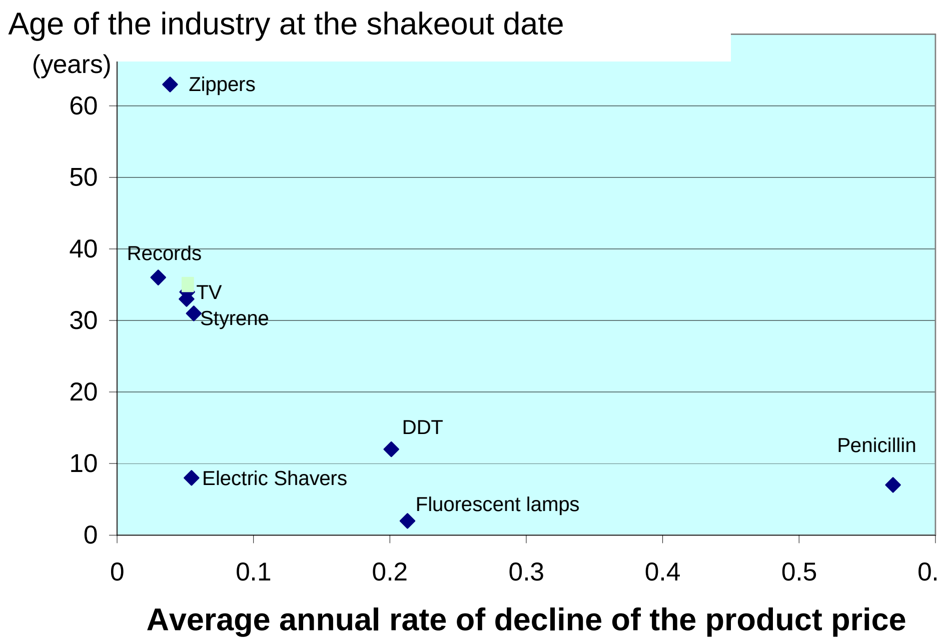


Figure 1: Technological progress and industry age at shakeout

in Figure 1.[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.[2] As we shall show, the same pattern arises in subsequent echoes – the faster the technological progress, the more frequent the echoes.

We also provide direct evidence that firms with old capital are more likely to exit: In the air-transportation industry, exiting firms have capital that is on average eight years older than the capital of the surviving firms; in Section 3 we shall display this highly significant relation both for the U.S. and for the world as a whole. A related pattern emerges at the two-digit-industry level: Sectors that face more rapidly declining equipment-input prices experience higher rates of entry and exit (Samaniego, 2006). In other words, a sector that enjoys a high rate of embodied technological change will have a high rate of entry and exit, or what one would normally understand to be a higher level of creative destruction.

All technological change in our model is embodied in capital, which means that

[1] GK time the start of the shakeout when net entry becomes negative for an appreciable length of time. The shakeout era typically begins when the number of producers reaches a peak and ends when the number of producers again stabilizes at a lower level.

[2] Table 7 of GK 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 6.

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.

The model is a standard vintage-capital model; Mitchell (2002) and Aizcorbe and Kortum (2005) use it to 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 investment spikes that may be caused by initial conditions have vanished. Jovanovic and Lach (1989) use it to analyze transitional dynamics but they cannot generate a shakeout or repeated investment echoes. We shall derive damped investment echoes that relate to the constant investment echoes derived by Boucekkine, Germain and Licandro (1997, henceforth BGM) in a similar GE model and by Mitra, Ray and Roy (1991) in a model where there is no progress but in which capital is replaced because it eventually tends to wear out. Our contribution is to interpret the shakeout data using, for the first time, a model in this class.

Plan of the paper.—Section 2 presents the model. Section 3 provides evidence on the hypothesis that capital replacement and firm exit occur at the same time. Sections 4 and 5 test the model's main implications. Section 6 presents the model's implications for TFP growth and Section 7 concludes the paper. The Appendix contains some proofs.

2 Model

Consider a small industry that takes as given the rate of interest, r, and the price of its capital. 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 takes prices as given. Let p be industry price, q industry output, and $D(p,t)$ the demand curve at date t, assumed continuous in both arguments. Production of the good becomes feasible at $t = 0$.

The price of capital is unity for all t. 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. We assume that willingness to pay at small levels of q is high enough to guarantee that investment will be positive immediately.[3]

Industry output.—Capital is the only input. Let $K_t(s)$ be the date-t stock of active capital of vintage s or older, not adjusted for quality. Industry output at time

[3]Sufficient for this is that $\int_0^\infty e^{-rt} D^{-1}(0,t)\,dt > 1$.

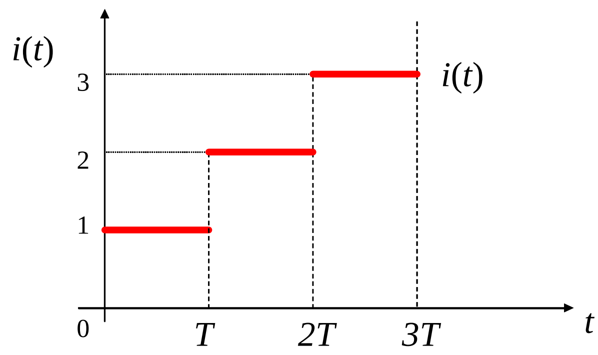


Figure 2: THE NUMBER OF SPIKES, $i\,(t)$.

t is the sum of the outputs of all the active capital

$$q\,(t) = \int_0^t e^{gs} dK_t\,(s)\,. \tag{1}$$

Capital ceases to be active when it is scrapped.

Evolution of the capital stock.—Because capital is supplied to the industry at a constant price, the investment rate will exhibit damped echoes. Any mass point that occurs will repeat itself, though in a damped form. That is, if a mass-point of investment ever forms, will recur at a periodicity of T, and the size of the mass point will diminish over time. Moreover, there must be an initial mass point at $t = 0$ because no capital is in place when the industry comes into being. After that initial mass point, capital evolves smoothly until date T when the original capital is completely replaced by vintage-T capital. This is the second industry investment spike. The third investment spike then occurs at date $2T$, when all the vintage-T capital is replaced, and so on. 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 $i = 1, 2, \ldots$. Let $i\,(t)$ be the integer index of the most recent spike.[4] We plot $i\,(t)$ in Figure 2.

We shall show that equilibrium is indeed of the form described in the previous paragraph: All the capital created at one spike date is replaced at the following spike date, T periods later. Therefore the capital stock at date t comprises capital created at the most recent spike date $i\,(t)$, plus the flow of investment, $x\,(t)$, over the preceding T periods.

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 $X_{i(t)}$, and the date-t *cumulative* distribution

[4]Formally, $i\,(t) = \max\{i \in \text{ integers} \mid i \geq 1 \text{ and } (i-1)\,T \leq t\}$.

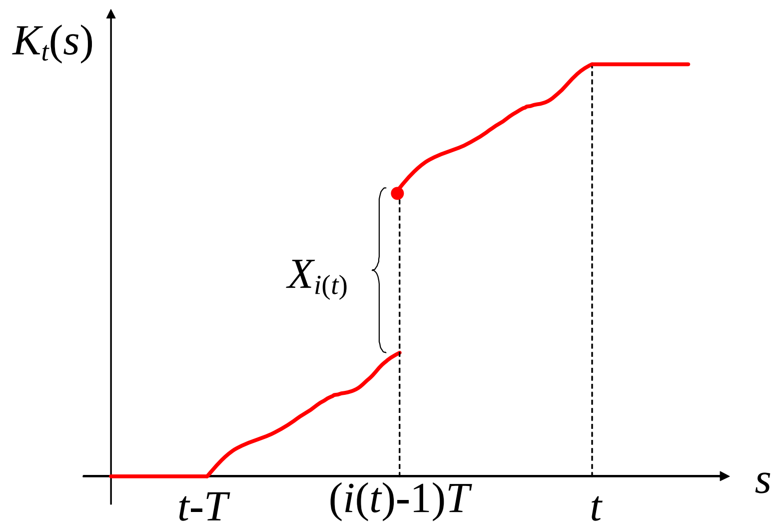


Figure 3: The date-t distribution of capital by vintage, s.

of capital by vintage is

$$K_t(s) = X_{i(t)} + \int_{\max(0,t-T)}^{s} x(u)\,du. \tag{2}$$

We portray $K_t(s)$ in Figure 3. It has exactly one discontinuity at $(i(t)-1)T$.

2.1 Equilibrium

The definition of equilibrium is simple if $x(t) > 0$ for all t. BGL call this the "no holes" assumption because when it holds, the vintage distribution of capital in use has no gaps in it. 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*: The revenue of a vintage-t machine at date $t' \in [t, t+T]$ is $e^{gt}p(t')$. 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 the RHS of (4) were ever less than unity, $x(t)$ would be zero.[5]

3. *Market clearing*:

$$D(p(t), t) = e^{g(i(t)-1)T} X_{i(t)} + \int_{\max(0,t-T)}^{t} e^{gs} x(s)\, ds. \tag{5}$$

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. ■

Having got T, we now use (7) to solve for the last unknown: $p_0 = ce^{g\tilde{T}(r,c,g)}$.

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. That condition is that for all t,

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

[5] Proposition 4 of BGM covers that case which arises when there is too high an initial stock of capital. This cannot be true at $t = 0$ in our model, and we shall state conditions in (10) that exclude it as an equilibrium possibility at any date.

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 $g\lambda$.

2.2 The nature of the investment spikes

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:*

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

Thus, industries with higher productivity growth should have an earlier shakeout.[6]

The second relation concerns the rate at which investment echoes or investment spikes die off:

Proposition 3 *Investment spikes decay geometrically. That is,*

$$X_n = e^{-gT(n-1)} X_1 \tag{12}$$

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

Proof. Because $p(t)$ is continuous at T, the number of efficiency units replaced at the spikes is a constant, X_1, which means that $X_n = e^{-gT} X_{n-1}$, and (12) follows. ■

The spike dates remain T periods apart, at $0, T, 2T, 3T, \ldots$. The spikes occur regularly because technological progress occurs at the steady rate g. Asymptotically, however, the spikes vanish and the equilibrium becomes like the one that Mitchell (2002) and Aizcorbe and Kortum (2005) analyze.

[6] We interpret g as an invariant property of an industry. But it can be interpreted as applying only to a given epoch in the lifetime of a given industry, and subject to occasional shifts. In Aizcorbe and Kortum (2005), e.g., one can think of such shifts as tracing out the relation between technological change and the lifetime of computer chips.

2.3 Other properties of the model

The next proposition relates the size of the shakeout to the ratio of the industry's date-zero to date-T output. To measure the size of the shakeout, we shall use the variable

$$y \equiv \frac{X_1}{q_T},$$

i.e., the fraction of capacity replaced at the first shakeout. Next, 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),

$$q_0 = D\left(ce^{gT}, 0\right) = X_1$$

Since *all* of the initial capacity is replaced at T, it must mean that $X_1 = q_0$ and that therefore

Proposition 4 *The fraction of capacity replaced at T depends inversely on output growth over the period $[0, T]$:*

$$y = z. \tag{13}$$

The restriction in (13) 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 (13) is the same. Proposition 4 will be tested in Section 5.

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.

2.4 Relation to GE vintage-capital models

Our partial equilibrium structure is not a special case of any GE vintage-capital model in which there is but one final good. Before explaining why, let us begin with the similarities. Johansen (1959) and Arrow (1962) assume a production function for the sole final good that is Leontieff in capital and labor. In that case the effective maintenance cost is the wage multiplied by the labor requirement per machine. Since wages rise at the same rate as the rate at which the labor requirement declines, and the maintenance rate is then constant in units of the consumption good. Thus our maintenance cost, c, has an exact counterpart in these models. Since these models have no costs of rapid adjustment, the infinitely-elastic supply of capital that we assume amounts to the same thing.

Now for the main and perhaps only fundamental difference between our model and the one-consumption-good GE models. In a GE model, for the interest rate to remain constant in the face of variations in the rate of investment that inevitably occur along the transition path from arbitrary initial conditions, the instantaneous utility of consumption must be linear. Then, if the rate of technological progress is also constant, the investment echoes will have the constant periodicity that they also have in our model, but they will not be damped. Rather, the investment profile simply repeats itself every T periods – see Mitra *et al.* (1991) and BGL.

3 Evidence on capital replacement and exit

The model assumes that capital is operated by measure-zero firms of indeterminate size. We hypothesize that a fraction of the 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 replacing. Instead of replacing its equipment at T, the firm can exit and let others bring the new equipment into the industry.

Reliable data on the age of capital are hard to find, and there are few, if any, studies linking firm exit to the age of its capital stock. We therefore conducted our own study using data on aircraft where accurate information on age is available. Figure 4 shows that in the U.S., the airplanes of exiting airlines were on average 7 years older than those of the surviving airlines. After weighing by the number of observations, the difference is highly significant. Balloon size is for each series proportional to the square root of the number of observations, but the constant of proportionality is larger for the exiting capital series so as to allow us to see how sample size of that series too evolves over time.

Figure 5 shows almost as strong a difference among all the world airlines. Capital of exiting airlines was four years older than the capital of surviving airlines and the difference in the means is again highly significant. Surprisingly, perhaps, the U.S. carriers operated older planes than the carriers based in other countries. A description of the data is in the Appendix, but a summary is in Table 1:

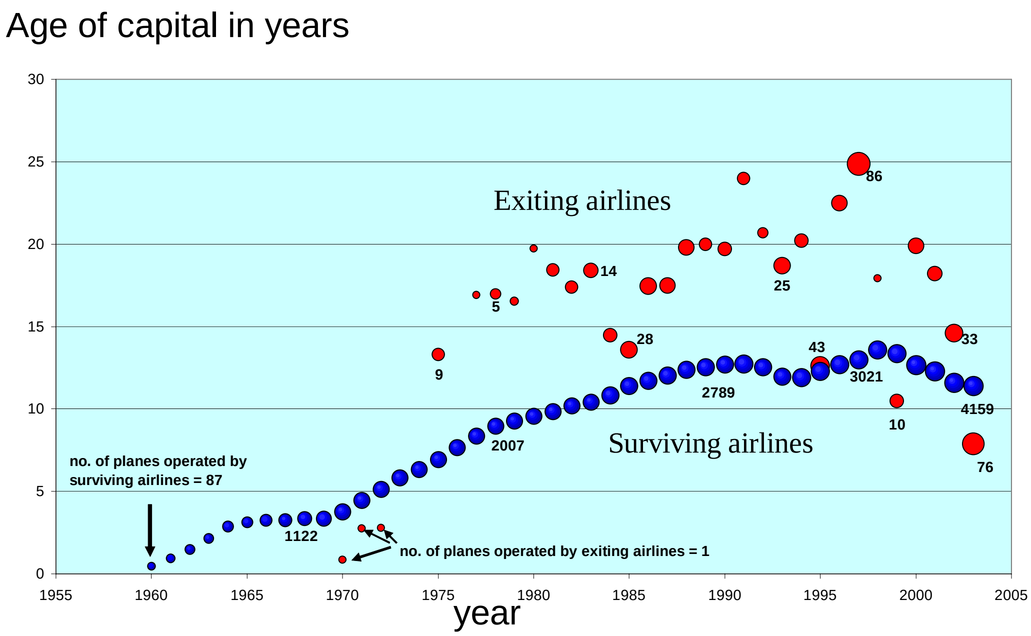


Figure 4: AGE OF CAPITAL AMONG EXITERS AND SURVIVORS IN THE U.S. AIRLINE INDUSTRY, 1960-2003.

	U.S.		**World**	
	Survivors	*Exiters*	*Survivors*	*Exiters*
Average age (yrs)	10.4	17	9.8	13.7
Std. dev. of age	7.5	9.4	7.4	9.4
# of airlines	231	181	1140	659
# of plane-year obs.	95,147	530	213,650	2,092

Table 1: THE DATA IN FIGURES 4 AND 5.

As reported in the table, a plane is counted as one observation for each year of its life. An airline is counted at most twice – as a survivor and then possibly as an exiter.

Three other pieces of evidence link exit decisions to the need to replace capital to the decision to exit

1. *Plant exit.*—Salvanes and Tveteras (2004) find that old plants are (i) less likely to exit, but (ii) more likely to exit when their equipment is old. Fact (i) they attribute to the idea that plants gradually learn their productivity and exit if the news is unfavorable as Jovanovic (1982) argued. Fact (ii) they attribute to the vintage-capital effect on exit.

2. *The trading of patent rights.*—The renewal of a patent is similar to replacing a piece of capital in the sense that a renewal cost must be paid if the owner of

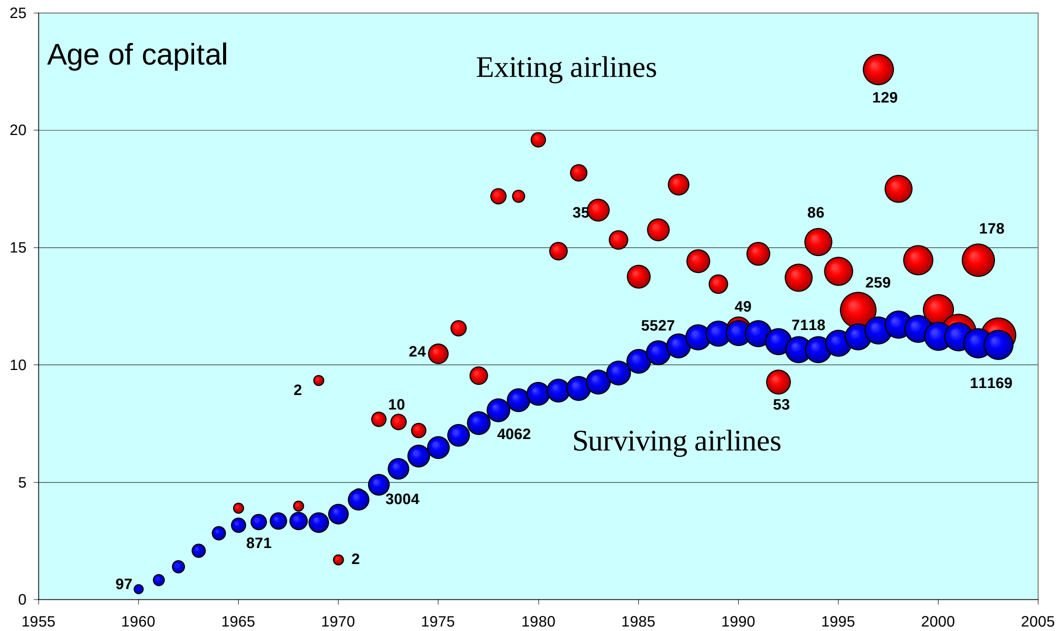


Figure 5: Age of capital of exiters and survivors in among all the world's airlines, 1960-2003.

the patent is to continue deriving value from it. If that owner sells the patent right, he effectively exits the activity that the patent relates to. Serrano (2006) finds that the probability of a patent being traded rises at its renewal dates, indicating that the decision to exit is related to the wearing out of a patent right.

3. *Higher embodied technical progress raises exit.*—A faster rate of decline in the price of capital makes it optimal to replace capital more frequently. If replacement sometimes leads to exit, we should see more exit where there is more embodied progress. Samaniego (2006) indeed finds that in sectors where the prices of machinery inputs fall faster, the firms using those machines experience higher rates of exit.

4 Empirics on the negative $g - T$ relation

GK measure an 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. GK say that an "exit" occurs when a firm stops making the product in question, even if that firm continues to make other products.

4.1 Testing (11): The timing of the shakeout

As suggested by the discussion in Section 3, we shall begin by proxying the age, T, of an industry at its first replacement spike, by the industry's age at which the shakeout of its firms begins. This measure will underlie the first test of the model. And as suggested by (6), we shall proxy the rate of technological progress, g, by the rate at which the price of the product declines. 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.

PRODUCT NAME	$\hat{g}^{(1)}$	$\hat{T}^{(1)}$	$\hat{g}^{(2)}$	$\hat{T}^{(2)}$
Auto tires			0.03	21
Ball-point pens	0.07	>28	0.07	>18
CRT			0.04	24
Computers			0.07	19
DDT	0.20	12	0.13	16
Electric shavers	0.05	8	0.01	5.5
Fluorescent lamps	0.21	2	0.21	2.5
Nylon	0.03	>34	0.03	>21
Home & farm freezers			0.02	18
Penicillin	0.57	7	0.57	8.5
Phonograph records	0.03	36	0.02	8.5
Streptomycin			0.31	22
Styrene	0.06	31	0.07	12
Television	0.05	33	0.08	8
Transistors			0.17	15
Zippers	0.04	63	0.05	38

Table 2: THE DATA IN FIGURES 6 AND 7.[7]

We now describe the procedure by which we choose the model's parameters. By Proposition 1, a unique solution to (9) for T exists, denoted it by $\tilde{T}(r, c, g)$. In Table 2, $\hat{T}^{(1)}$ is the date that GK find that Stage 4 (the shakeout stage) begins in their various industries. In a couple of cases the shakeout had not yet begun, and they are censored. The annual rate at which the price declines, averaged over the period

[7]The variable definitions are: $\hat{T}^{(1)}$ = age at the start of shakeout, $\hat{g}^{(1)}$ = average annual price decline on $\left[0, \hat{T}^{(1)}\right]$, $\hat{T}^{(2)}$ = the number of years elapsed between the start of stage 2 and the midpoint of the shakeout stage, and $\hat{g}^{(2)}$ = annual price decline over stages 2 and 3.

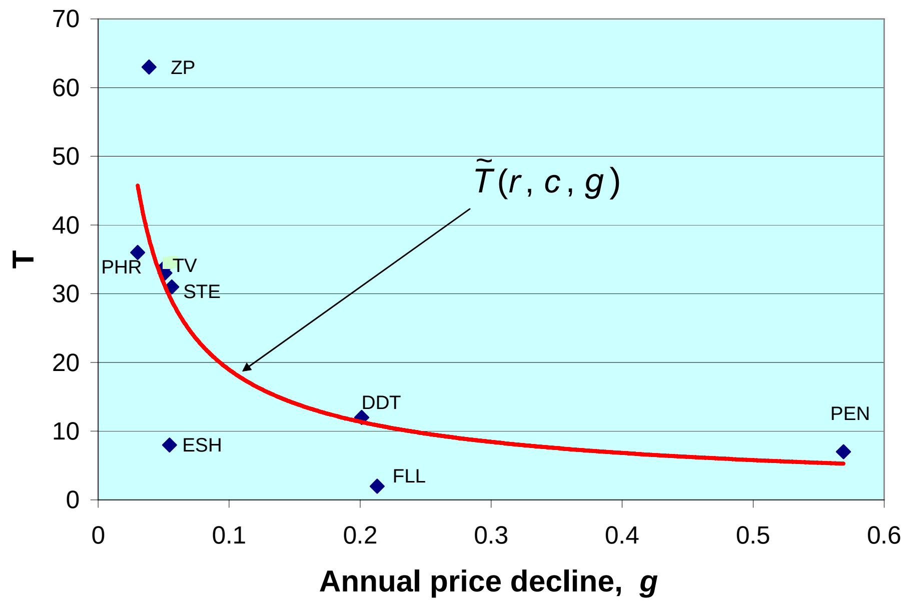


Figure 6: The relation between $\hat{T}$ and $\hat{g}$ for the eight uncensored GK observations; censored observations not used here.

$\left[0, \hat{T}^{(1)}\right]$ is $\hat{g}_i^{(1)}$. We do not have observations on r and c, so we set $r = 0.07$ and estimate c to minimize the sum of squared deviations of the model from the data for the eight industries for which we have complete information on both g and T:

$$\min_c \left\{ \sum_{i=1}^{8} \left[\hat{T}_i^{(1)} - \tilde{T}\left(0.07, c, \hat{g}_i^{(1)}\right)\right]^2 \right\}.$$

The data and the implicit function $\tilde{T}(g, r, c)$ fitted to them are shown in Figure 6. The estimate is $c = 0.039$ (s.e. $= 0.013$, $R^2 = 0.53$).

Our estimate of c is in the range of typical maintenance spending. McGrattan and Schmitz (1999) report that in Canada, total maintenance and repair expenditures have averaged 5.7 percent of GDP over the period from 1981 to 1993, and 6.1 percent if one goes back to 1961. These estimates are relative to output, however, whereas ours are relative to the purchase price of the machine which is normalized to unity. Relative to output, maintenance costs are one hundred percent at the point when the machine is retired (this is equation [3]). Maintenance costs are constant over the machine's lifetime, whereas the value of the machine's output relative to the numeraire good is e^{gT} when the machine is new. Now e^{gT} averages around $e^{1.7} = 5.5$ – see Figures 13 and 14 – so that as a percentage of output, maintenance spending ranges between 18 and 100 percent. Therefore, c must stand partly for wages to workers as a fixed-proportion input as in the original vintage-capital models like Arrow (1962) and Johansen (1959) that had a fixed labor requirement.

4.1.1 Testing (11) using an alternative definition of T

We now entertain a different definition of T; one that may provide a better test of the model, and one that will provide us with more observations. The main reason for doing this is that GK's "Stage 1," defined as the period during which the number of producers is still small (usually two or three), may not contain what we would call an investment spike. During stage 1, only a few firms enter, a number that is in some industries – autos and tires, e.g., – much smaller than the number of firms that exit during the shakeout.

The model predicts a date-zero investment spike $X_1 = D(p_0, 0)$, without which there would be no exit spike at date T. Not all the GK industries will fit this, however. Indeed, GK state that rarely is a product's initial commercial introduction immediately followed by rapid entry. Autos, e.g., had very low sales early on, and it took years for sales to develop.[8] Therefore the spike is better defined at or around the time when the entry of firms was at its highest. Moreover, in GK, for many industries, the shakeouts were not completed until a few years after they first began. It may thus be more appropriate to designate the shakeout date as the midpoint of the shakeout episode instead of as the start date of the shakeout episode.

In light of this, let $\hat{T}^{(2)}$ be the time elapsed between the industry's "Takeoff" date (which is when Stage 2 begins) and the midpoint of the the industry's shakeout episode (the midpoint of Stage 4). This revised definition for $\widehat{T}$ calls for adjusting $\widehat{g}$ to be the rate of the average annual price decline between the takeoff date (which comes after the industry has completed stage 1), and the first shakeout date. We call this variable $\hat{g}^{(2)}$. This allows us to enlarge the sample. The additions are as follows:

1. For six of the GK industries, complete information for price-declines in Stages 2 and 3 (but not Stage 1) is available. They can now be added to the analysis;

2. The two censored observations listed in Table 2 will also be added;

3. We replace the GK information for the TV by that reported in Wang (2006), who compiled the data from the Television Factbook. This change may better reflect the history of the TV industry as GK dated the birth of the industry as early as 1929, while according to Wang, the commercial introduction of TV starts only in 1947.

[8] Klepper and Simons (2005), however, do find an initial spike for TVs and Penicillin – both start out strong after WW2. There may be a problem with the TV birthday being set at 1929 as GK have it. During WW2 the Government had banned the sale of TVs.

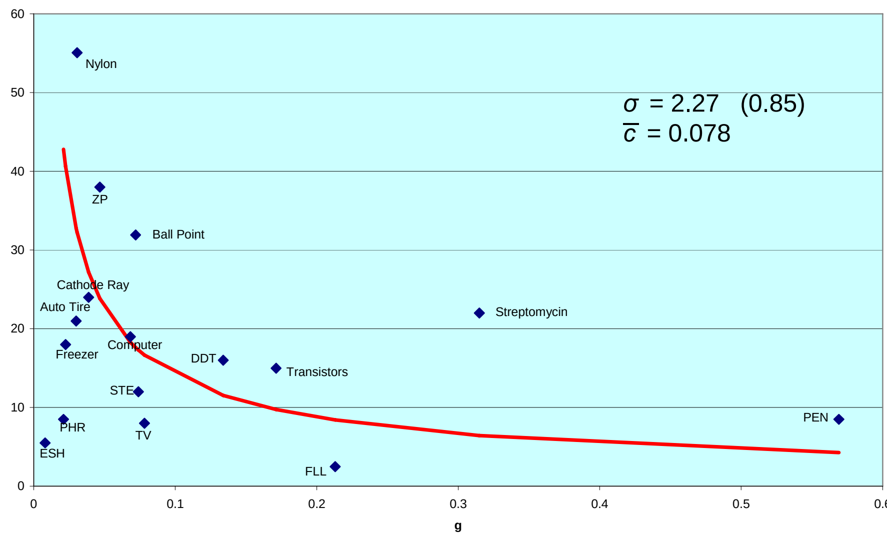


Figure 7: $\hat{T}$ AND $\hat{g}$ USING THE ALTERNATIVE DEFINITION OF T

The estimation procedure.—The inclusion of the two censored observations leads us to use maximum likelihood. We set $r = 0.07$ and assume that the distribution of c over industries is log- normal: For all firms in industry i, $\ln c_i$ is a draw from $N(\ln \overline{c}, \sigma^2)$. We then estimate $\overline{c}$ and σ to fit the model to the data; the details are in the Appendix.

The estimate of $\bar{c} = 0.078$ is a bit higher than with the previous sample. This is because T is smaller under the new definition, and a higher replacement cost is needed to generate the earlier replacement. Figure 7 plots the predicted T for the industry for which $c = \overline{c}$. For the two truncated observations, Ball–point pens and Nylon, the data points represent the respective means conditional on their truncated values.

4.2 Testing (11) using subsequent entry and Exit Spikes

In our model industry-wide investment spikes continue to occur, yet Gort and Klepper report at most one shakeout per industry. Shakeouts should diminish geometrically in absolute terms as (12) shows. Moreover, exit spikes should coincide with entry spikes. This section tests these implications using Agarwal's extension and update of the GK data, described fully in Agarwal (1998). The products are listed and some statistics on them presented in Table 3.

The evidence hitherto is mixed: Cooper and Haltiwanger (1996) find industry-wide retooling spikes, but GK did not report second shakeouts, though this may in part be because the GK data rarely cover industry age to the point $t = 2T$ where we ought to observe a second shakeout. In any case, Agarwal's data cover more years and contain entry and exit separately, and we shall use them to study this question.

A procedure for detecting spikes must recognize the following features of the data:

(A) *Length of histories differ by product.*—Coverage differs widely over products, from 18 years (Video Cassette Recorders) to 84 years (Phonograph Records).

(B) *The volatility of entry and exit declines as products age.*—The model predicts that the volatility of entry and exit should decline with industry age. Other factors also imply such a decline: (i) Convex investment costs at the industry level, as in Caballero and Hammour (1994), and (ii) Firm-specific c's. Both (i) and (ii) would transform our X_n from spikes into waves and, eventually, ripples.[9]

Hodrick-Prescott residuals in entry and exit rates.—Roughly speaking, we shall say that a spike in a series Y_t occurs whenever its HP residual is more than two standard deviations above its mean. "Roughly", because of adjustments for (A) and (B) above. We constrain industry i's trend, τ, by

$$\sum_{t=2}^{A_i} (\tau_t - \tau_{t-1}) \leq a A_i^b,$$

where A_i is the age at which an industry's coverage ends. We set $a = 0.005$ for both series. Because both series are heteroskedastic, with higher variances in earlier years, we chose $b = 0.7$ for both entry and exit (If b were unity, an industry with longer coverage would have a larger fraction of its observations explained by the trend). The trend therefore explains about the same fraction of the variation in short-coverage industries as in long-coverage industries.

This fixes problem (A), but not (B): The HP residual, $u_t \equiv Y_t - \tau_t$, is still heteroskedastic, the variance being higher at lower ages. To fix this, we assumed that the standard-deviation was age specific and equal to

$$\sigma_t = \sigma_0 t^{-\gamma}, \tag{14}$$

where $\sigma_0 \geq 0$ and $\gamma \geq 0$ are product-specific parameters estimated by maximizing the normal likelihood[10]

$$\prod_{t=1}^{A_i} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2}\left[\frac{u_t}{\sigma_t}\right]^2\right).$$

[9] Spikes may also dissipate because (i) A positive shock to demand would start a new spike and series of echoes following it; these would mix with the echoes stemming from the initial investment spike, (ii) Random machine breakdowns at the rate δ would transform (12) into $X_n = e^{-(g+\delta)T(n-1)}X_1$, which decays faster with n.

[10] Although the HP residuals are not independent and probably not normal either, this procedure still appears to have removed the heteroskedasticity in the sense that the spikes were as likely to occur late in an industry's life as they were to occur early on.

The spike-detection algorithm.—If at some date the HP residual is more than two standard deviations above its mean of zero, then that date is a spike date. But we shall allow for the possibility that unusually high replacement will take up to three periods. Thus we shall say that a series Y_t in a certain time window is above 'normal' if one or more of the following events occurs

$$\begin{array}{ll} \text{1-period spike:} & u_t > 2\sigma_t, \\ \text{2-period spike:} & u_t > \sigma_t \;\text{ and }\; u_{t+1} > \sigma_{t+1}, \\ \text{3-period spike:} & u_t > \frac{2}{3}\sigma_t \;\text{ and }\; u_{t+1} > \frac{2}{3}\sigma_{t+1} \;\text{ and }\; u_{t-1} > \frac{2}{3}\sigma_{t-1}. \end{array}$$

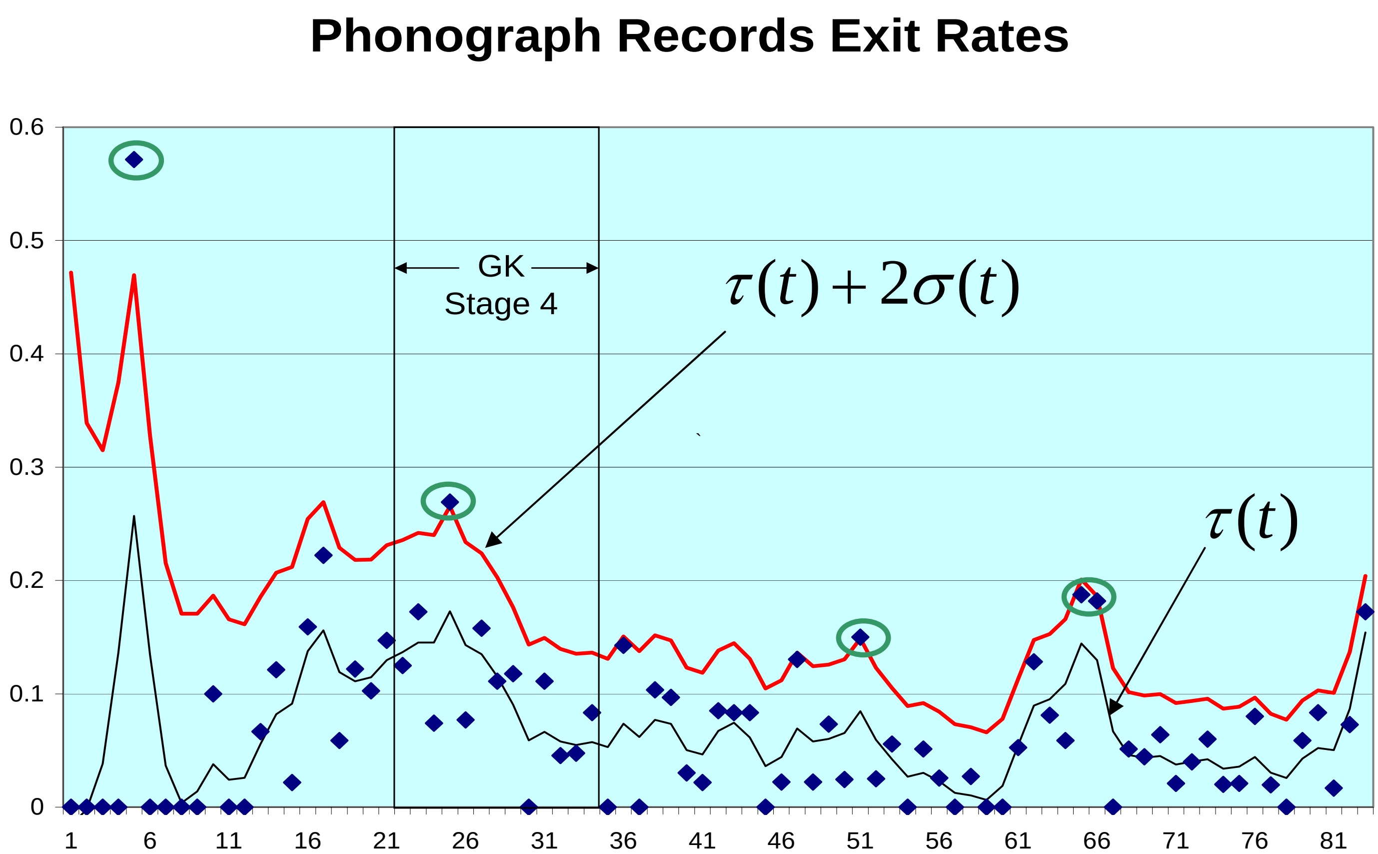
The cutoff levels of 2, 1, and $\frac{2}{3}$ times σ_t were chosen in the expectation that each of the three events would carry the same (small) probability of being true under the null. The latter depends on the distribution and the serial correlation of the u_t which we do not know. But, again for the normal case, these probabilities turned out to be roughly the same. That is,

$$1 - \Phi(2) = 0.023, \quad (1 - \Phi(1))^2 = 0.025, \text{ and } \left(1 - \Phi\left(\frac{2}{3}\right)\right)^3 = 0.016.$$

Table 3 summarizes the results in more detail. The 33 products are listed alphabetically and are so numbered. To explain the table, let us focus on product 23, Phonograph Records and read across the row. Records were first commercialized, i.e., sold, in 1908. Being the oldest product, it also is the product for which we have the most observations, 84, since (with one exception) the series all end in 1991. The next four entries are the exit and entry spikes, by age of industry and by calendar year. There are seven one-year spikes and one two-year spike, this being the last exit spike. The 1934 exit spike is labelled in red because it falls in the GK shakeout region the dates of which are in the last column of the table. See Figure 7 where the GK shakeout region is shaded. The remaining columns report the correlations between the entry and exit series.[11] The raw series are negatively correlated – when the industry is young, entry is higher than exit, and later the reverse is true – but the correlation is slight (-0.07). The trends (i.e., the τ's) are more negatively correlated (-0.27). The HP residuals, on the other hand, are *positively* correlated; our model suggests that this should be so because the spikes should coincide.[12] The correlations averaged across products are at the bottom of the table.

[11] Since there are no incumbents at date zero, the entry spike at date zero is infinite in both the data and the model. Since neither the entry nor the exit rate is defined in the first year of the raw series, both series begin in the second year. Thus "Year t" of the entry and exit rate series refers to element $t+1$ of the raw data series.

[12] While the entry and exit spikes do not always coincide, a clustering test finds a highly positive and significant correlation in their timing.



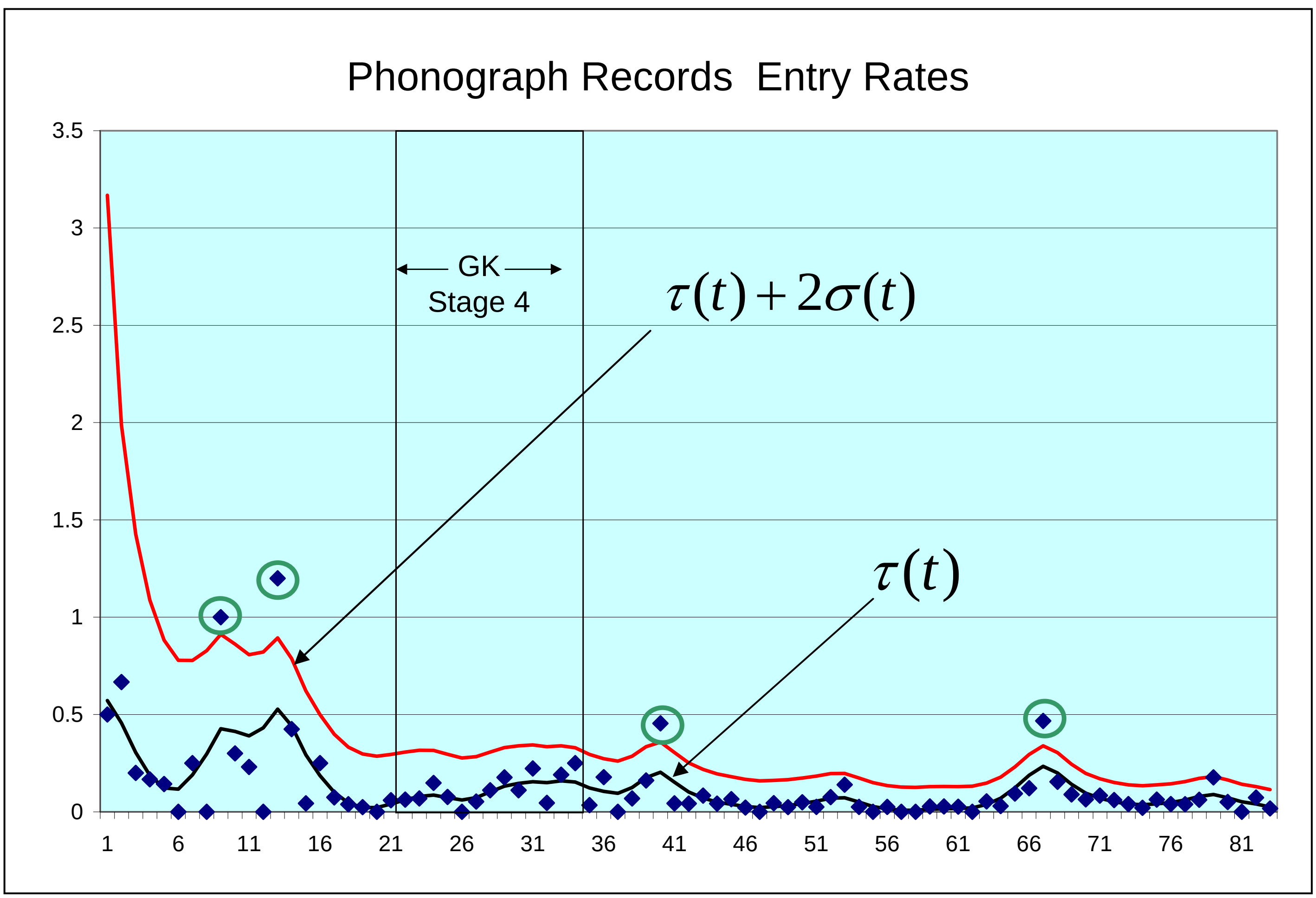


Figure 7: How the spikes were determined for the case of the Phonograph-Record Industry

	Product & yr of comm intr.	Length of raw series	EXIT		ENTRY				Corr btw entry and exit rates	G-K Stage 4
			spike dates		spike dates		Corr btw smoothd entry and exit rates	Corr btw entry and exit residuals		
			in yrs since comm intro	in calendar yr	In yrs since comm intro	in calendar yr				
1	Antibiotics	44	7	1955	2	1950	-0.46	-0.04	-0.30	not in G-K
	1948		19	1967	39	1987				
			32	1980						
2	Artificial Christmas Trees	54	8	1946	35	1973	-0.52	-0.02	-0.21	1968-1969
	1938		18	1956	49	1987				
			43	1981						
			45	1983						
3	Ball-point Pens	44	23	1971	9-10	1957-58	0.81	-0.32	0.55	S4 not reached
	1948				26	1973-74				
4	Betaray Gauges	36	7	1963	7	1963	0.18	0.34	0.26	1973-
	1956				33	1989				
5	Cathode Ray Tubes	57	54	1989	10	1945	0.92	0.55	0.79	1963-1967
	1935				15	1950				
					52	1987				
6	Combination Locks	80	13	1925	21	1933	0.41	0.01	0.21	not in G-K
	1912		16	1928	29	1941				
			65	1977	53	1965				
			75	1987						
7	Contact Lenses	56	10	1946	6	1942	0.09	-0.15	-0.13	not in G-K
	1936		29	1965	12	1948				
			35	1971	30	1966				
					39	1975				
8	Electric Blankets	76	3	1919	6-7	1922-23	-0.13	0.07	-0.09	1962-1973
	1916		35	1951	30	1946				
			41	1957	46	1962				
			63	1979	70	1986				
			69-70	1985-86						
9	Electric Shavers	55	36	1973	36-39	1973-76	0.22	-0.09	-0.11	1938-1945
	1937				49	1986				
10	Electrocardiographs	50	6	1948	6	1948	0.25	0.41	0.37	1964-1969
	1942		32	1974	48	1990				
11	Freezers	46	27	1973	40	1986	0.41	-0.13	0.23	1947-1957
	1946									

Red = Is or may be in GK Stage 4 **Blue = Within 1 year of exit spike**

Table 3: Entry and Exit Statistics in Agarwal's data

	Product & yr of comm intr.	Length of raw series	EXIT spike dates: in yrs since comm intro	EXIT spike dates: in calendar yr	ENTRY spike dates: in yrs since comm intro	ENTRY spike dates: in calendar yr	ENTRY: Corr btw smthed entry and exit rates	ENTRY: Corr btw entry and exit rsiduals	Corr btw entry and exit rates	G-K stage 4
12	Freon Compressors	57	5	1940	3	1938	-0.21	-0.12	-0.17	1971-1973
	1935		**37**	**1972**	46	1981				
			52	1987						
13	Gas Turbines	48	20	1964	2	1946	-0.42	0.00	-0.22	1973-
	1944				36	1980				
					41	1985				
14	Guided Missiles	41	2	1953	8	1959	-0.32	-0.43	-0.41	1965-1973
	1951				35	1986				
15	Gyroscopes	77	10	1925	4	1919	0.36	0.10	0.24	1966-1973
	1915		31-32	1946-47	**30-32**	1945-47				
					39-40	1954-55				
16	Heat Pumps	38	5	1959	25	1979	-0.42	-0.08	-0.24	1970-1973
	1954		**16**	**1970**						
			32	1986						
17	Jet Engines	44	6	1954	**6**	1953-54	-0.26	0.42	0.07	1960-1962
	1948		22	1970	32	1980				
					35	1983				
18	Microfilm Readers	52	5-6	1945-46	8	1948	0.37	-0.24	-0.12	S4 not reached
	1940		16	1956	22-23	1962-63				
					33	1973				
					45	1985				
19	Nuclear Reactors	37	30	1985	**30**	1985	-0.57	0.12	-0.38	1965-1973
	1955		33	1988	35	1990				
20	Outboard Motors	79	22	1935	2-4	1915-17	-0.18	0.07	-0.08	1921-1923
	1913		67	1980	28	1941				
					34	1947				
					72	1985				
21	Oxygen Tents	60	**41**	**1973**	4	1936	-0.32	-0.17	-0.26	1967-1973
	1932		48	1980	56	1988				
			52-53	1984-85						
22	Paints	58	**39**	**1973**	13-14	1947-48	-0.41	0.04	-0.20	1967-1973
	1934		43	1977	20	1954				

Red = Is or may be in GK Stage 4 **Blue = Within 1 year of exit spike**

Table 3, continued

	Product & Yr of Comm Intr.	Length of raw series in Agarwal	EXIT spike dates in yrs since comm intro	EXIT spike dates in calendar yr	ENTRY spike dates in yrs since comm intro	ENTRY spike dates in calendar yr	Corr btw smoothed entry and exit rates	Corr btw entry and exit residuals	Corr btw entry and exit rates	G-K stage 4
23	Phonograph Records	84	5	1913	9	1917	-0.23	0.08	-0.07	1923-1934
	1908		25	1933	13	1921				
			51	1959	40	1948				
			65-66	1973-74	67	1975				
24	Photocopying Machines	52	34	1974	42	1982	-0.30	-0.20	-0.25	1965-1973
	1940		39	1979						
25	Piezoelectric Crystals	51	13	1953	3	1943	-0.28	-0.06	-0.21	1955-1957
	1940		40	1980	45	1985				
26	Polariscopes	64	15	1943	18	1946	0.20	-0.26	-0.03	1964-1967
	1928		45	1973	23	1951				
					25	1953				
					44	1972				
27	Radar Antenna Assemblies	40	15	1967	4	1956	-0.50	-0.10	-0.32	1957-1968
	1952		21	1973	33	1985				
28	Radiant Heating Baseboards	45	25	1972	27	1974	-0.42	0.02	-0.19	1972-1973
	1947		37	1984						
29	Radiation Meters	43	no spike dates		6	1955	-0.28	0.20	-0.13	not in G-K
	1949				37	1986				
					41	1990				
30	Recording Tapes	40	11	1963	33	1985	-0.53	0.06	-0.31	1973-
	1952		28	1980						
31	Rocket Engines	34	32	1990	7	1965	-0.34	-0.08	-0.19	1973-
	1958				33	1991				
32	Styrene	54	17	1955	4-5	1942-43	-0.55	-0.09	-0.27	1966-1973
	1938				20-1	1958-59				
33	Video Cassette Recorders	18	no spike dates		7	1981	-0.33	-0.18	-0.11	not in G-K
	1974				11	1985				
	average	51.94					-0.11	-0.01	-0.07	
	max	84.00					0.92	0.55	0.79	
	min	18.00					-0.57	-0.43	-0.41	
	std deviation	14.97					0.40	0.22	0.27	

Red = Is or may be in GK Stage 4 **Blue = Within 1 year of exit spike**

Table 3, continued

Figure 7 plots the exit and entry series for Phonograph records, in each case plotting the HP trend and the two-standard-deviation band. Spikes are circled in green. Let us note the following points:

1. For both entry and exit the spikes are evenly distributed, and this is true in most industries. This suggests that the heteroskedasticity adjustment in (14) is adequate.

2. The number of entry and exit spikes is equal – four entry and four exit spikes. But only the final, fourth spikes coincide in that they are within one year of each other. There were only five other products (10, 11, 25, 27, and 32) for which this was so.

3. The second exit spike is well in the GK region, but there should also have been an entry spike in that region. Over all the industries the number of entry spikes (67 in all) is slightly less than the number of entry spikes (79 in all), for the nine industries in which the GK region does contain an Agarwal spike, it is *always* an exit spike – see the red numbers in Table 2.

4. Just as our model predicts, however, $T_{2,i} - T_{1,i}$ is negatively related to $\hat{g}_i$. That is, analogously to the result in Figures 6, the first exit spike is followed sooner by the second exit spike in those industries i where prices decline faster. We now calculate $\widehat{g}$ as the rate of average price decline during the years covered by $T_2 - T_1$. To maximize the number of observations, we include products for which price information is available for as little as 70% of the time during the years covered by $T_2 - T_1$. For *Outboard Motors* and *Recording Tapes*, however, price went up during those years which is inconsistent with the model. But they both had above-average $T_2 - T_1$ values, 42 and 17 respectively, and while we did not include them in the estimation routine, we include them in Figure 9, setting $\hat{g}$ to zero in both cases.

5. The same test is also done on entry spikes. Once again, $T_{2,i} - T_{1,i}$ is negatively related to $\hat{g}_i$. The results are in Figure 10.[13] At first it may seem like a better fit could be obtained at a lower value of $\bar{c}$ which would raise the red curve upwards. The problem, however, is that a fall in c raises T by much more when g is low, as one can verify from Proposition 5 and more intuitively from (7), so that the curve moves clockwise.

4.3 Other explanations for the negative $g - T$ relation

Endogenous technological change.—Our model assumes that technological change is exogenous. Klepper (1996) assumes firms do research, and his model appears to imply

[13]For both plots $\bar{c}$ solves $\min_c \sum_{i=1}^{n} \left(\widehat{T}_i - \overline{T}\left(r, c, \widehat{g}_i\right)\right)^2$.

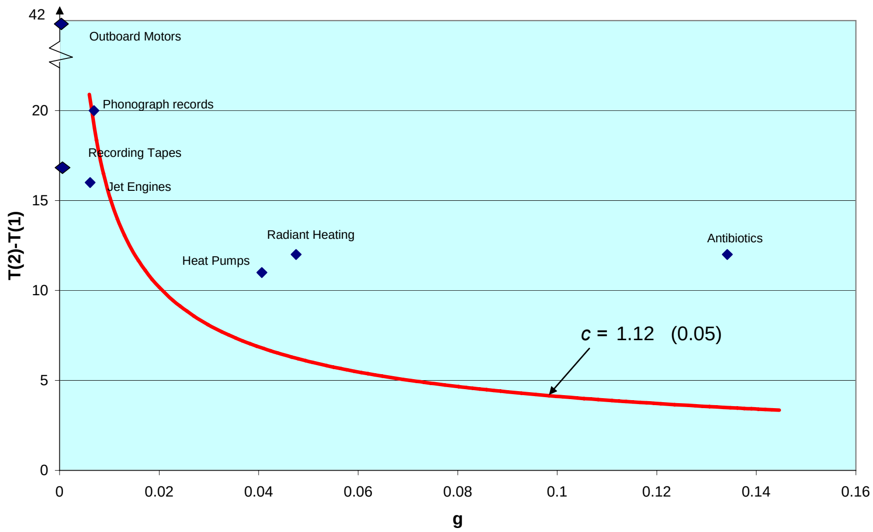


Figure 9: EXITS: THE RELATION BETWEEN $\hat{T}_2 - \hat{T}_1$ AND $\hat{g}$

that leading firms would squeeze out the inefficient fringe more quickly in industries where there is more technological opportunity and, hence, faster-declining product prices. The difference between our model and his concerns the fate of the firms in the first cohort of entrants: In our model the first cohort is the least efficient, whereas in Klepper's model, the first cohort is the *most* efficient because it has done the most research. In Grossman and Helpman (1991) and Aghion and Howitt (1992), monopoly incumbents are periodically replaced by more efficient entrants. Tse (2001) extends the model to allow for more than a single producer, but he too has a fixed number of producers over time and cannot explain shakeouts.

Demand declines.—Consider the following explanation for the negative relation between $\hat{T}$ and $\hat{g}$. Let there be just two periods of different length: $[0, 1]$ and $(1, \infty)$. Suppose demand either stays constant or that it declines, and that the outcome is unpredictable. A demand decline produces a shakeout ($T = 1$) and a decline in p. If demand does not decline then there is no shakeout ($T = +\infty$) and no decline in p. This explanation does not, however, get support from GK's sample. For the analysis in Figure 6, among the industries where output fell during the shakeout,[14] the average pair was $\left(\widehat{g}^{(-)}, \widehat{T}^{(-)}\right) = (0.07, 30.4)$, whereas among the industries where output rose,[15] the average pair $\left(\widehat{g}^{(+)}, \widehat{T}^{(+)}\right) = (0.28, 13.3)$. For the analysis in Figure 7, the corresponding figures are $\left(\widehat{g}^{(-)}, \widehat{T}^{(-)}\right) = (0.05, 17.3)$[16] and $\left(\widehat{g}^{(+)}, \widehat{T}^{(+)}\right) =$

[14] DDT, Electic shavers, Phonograph records, TV, and Zippers.
[15] Fluorescent lamps, Penicillin, and Styrene.
[16] The industries in note 14 plus Auto tyres and CRT.

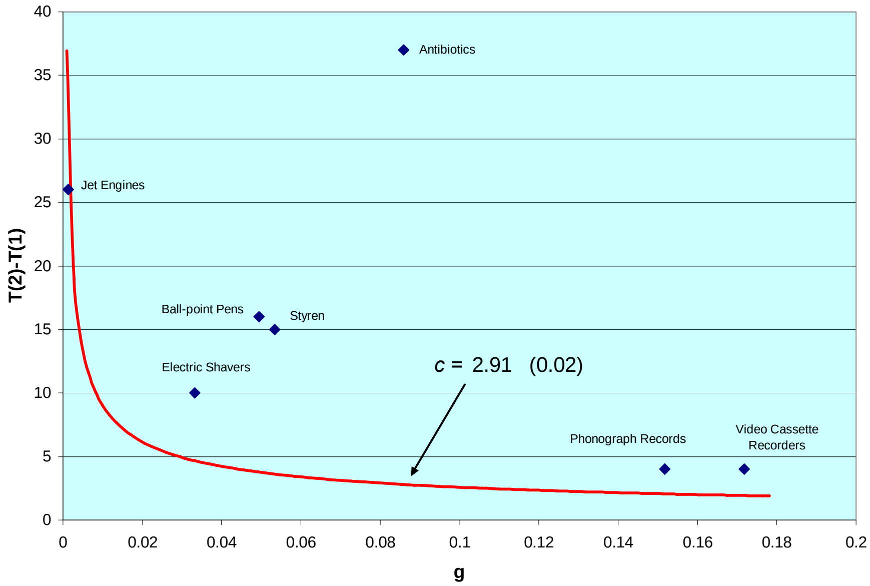


Figure 10: ENTRY: THE RELATION BETWEEN $\hat{T}_2 - \hat{T}_1$ AND $\hat{g}$

$(0.21, 11.2)$.[17] 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 measure the learning period by T. But then one needs to explain why learning would be faster in those industries where p is declining rapidly.

Consolidations for other reasons.—Again, one would need a reason for why consolidations should occur sooner in industries where technological change is more rapid.

5 Testing (13)

Testing Proposition 4 requires that we estimate the fraction of capacity replaced and the growth of output when the industry reaches age T. Equation (13) 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. This is the total decrease in the number of firms for a period of time that lasted for on average 5.4 years, after

[17]The industries in note 15 plus Freezers and Transistors. GK did not report whether output fell or rose during the shakeout for Computers and Streptomcyin.

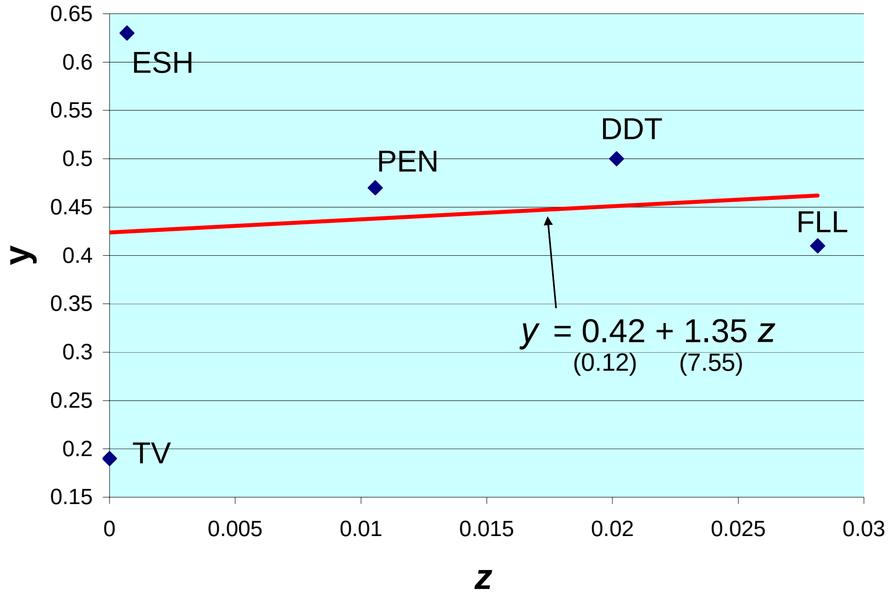


Figure 11: TESTING (13): THE RELATION BETWEEN y AND z.

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 surviving 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 (1988) find that firm size is positively associated with survival. Table 2 of Dunne *et al.* (1988) 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 4 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. We add a constant, β_0, to the RHS of the regression equation in order to represent these other forms of exit. Thus, we estimate the relation

$$\hat{y}_i = \beta_0 + \beta_1 \hat{z}_i + u_i,$$

where β_1 represent the fraction of firms that exit in response to capacity replacement. The estimates are in Figure 11. There are too few observations to allow us to tell if β_1 differs significantly from 3.

5.0.1 Tests of (13) with the alternative definition of T

The alternative definition of T takes the date of the first spike to be the takeoff date of an industry, i.e., the start of GK's Stage 2. In that case, if we ignore as negligible

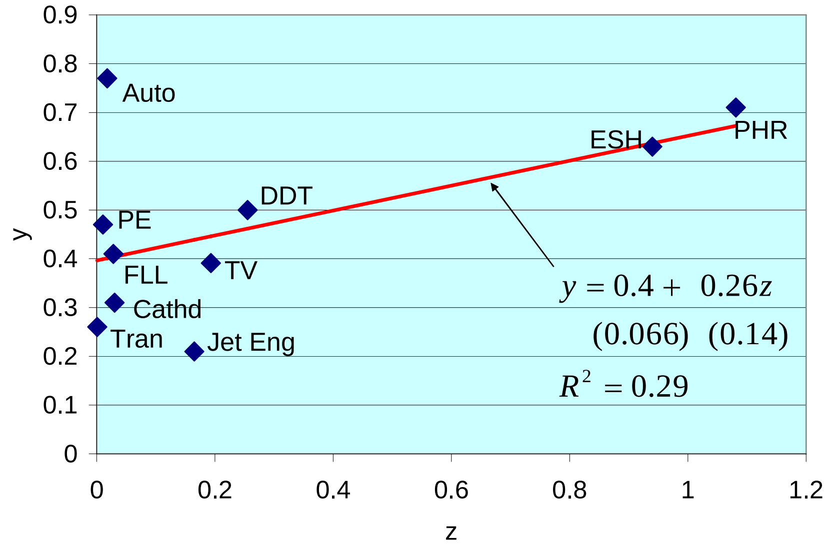


Figure 12: Testing (13) with the alternative definition of T

the capacity created before the takeoff date, $\widehat{z}$ should equal the inverse of the rate of output increase between the takeoff date and the start of Stage 4, i.e., the date at which the number of firms in the industry peaked. There are 5 industries in GK for which output data, as well as information on $\widehat{y}$, are available from the takeoff date to the shakeout date but not earlier. These industries can now be added to the analysis. For the TV industry we use the information from Wang (2006) in place of GK. The results are described in Figure 12. The slope is marginally significant, but far lower than the model predicts.

6 Implications for productivity growth

The model has no labor, and so TFP is simply the productivity of capital.[18] 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 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

[18] One way to add labor to the model is to have one worker per machine, and to interpret c 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.

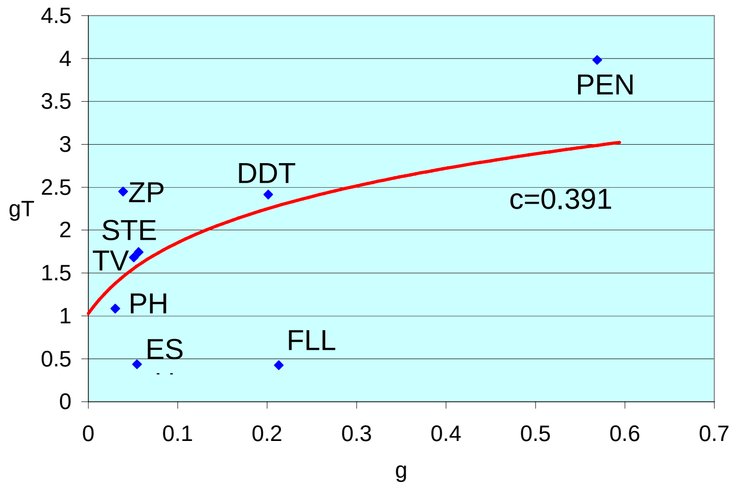


Figure 13: TFP growth and the max/min TFP ratio – predicted and actual

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.

6.0.2 TFP growth vs. dispersion

Faster TFP-growing industries should show greater TFP dispersion in our model. Dwyer (1998) derives the same relation in a similar model. In a group of textile industries he found that the relation was positive and significant.[19] 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 13 plots the logarithm of

$$\ln \frac{\text{TFP}_{\max}}{\text{TFP}_{\min}} = gT$$

[19] 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.

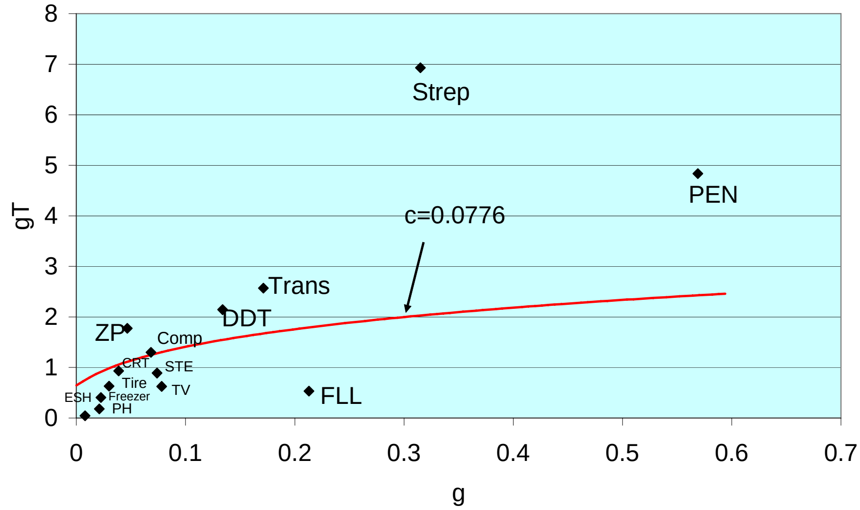


Figure 14: TFP GROWTH AND THE MAX/MIN TFP RATIO WITH THE ALTERNATIVE DEFINITION OF T – PREDICTED AND ACTUAL

as a function of g. It shows, in other words, the comparative steady-state relation between inequality and growth.[20] The relation is positive, but for the 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. Table 5 of Aizcorbe and Kortum (2005) reports a result similar to the one portrayed in Figure 13 – when g is larger, gT should go up.

Figure 13 is based on the information in Figure 6, and on the least-squares estimate $c = 0.039$. If we switch to measuring T by the time between the industry's takeoff and the midpoint of the shakeout episode while adjusting the measure of g accordingly, and if we use the ML estimate of $\bar{c}$ reported in Figure 7, we get the results shown in Figure 14. Note that Figures 13 and 14 portray the very same information portrayed in Figures 6 and 7 – no new information is added. The difference is merely that the variable plotted on the vertical axis is gT instead of T. Since the data do not have T declining with g as fast as the model predicts, the relation between g and gT is steeper than the model predicts. This is especially evident in Figure 14.

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 sometimes, but we hardly ever see the "last shall be first" phenomenon of the least productive

[20]Moreover, the model's implications for inequality have a curious discontinuity at $g = 0$. Namely, if g is really zero, then everyone should be using the same quality machine and there should be no TFP inequality. But a steady state with a very small g would have TFP inequality of at least unity.

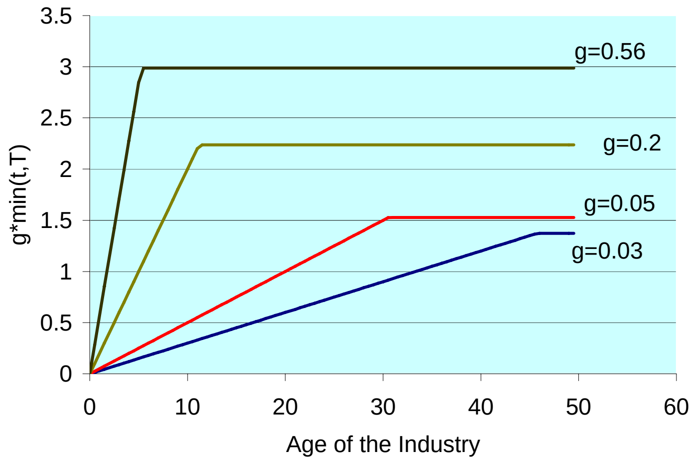


Figure 15: TRANSITIONAL DYNAMICS IN PRODUCTIVITY DISPERSION

producer suddenly becoming the most productive. In our model, 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 (1992, 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 of 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.[21] 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) document. 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.

6.0.3 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

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

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 15 we show how the fanning out process depends on the industry's growth rate by plotting, for four different values of g, the logarithm of

$$\ln \frac{\mathrm{TFP}_{\max}}{\mathrm{TFP}_{\min}} = g \min(t, T)$$

where t is the age of the industry. In contrast, learning models like that of Jovanovic (1982) are ambiguous on this score: The dispersion among survivors can become more narrow over time if certain distributional assumptions are met, as was the case for the distribution that Jovanovic (1982) used – the variance of the truncated normal distribution is smaller than the variance of the untruncated distribution.

7 Conclusion

This paper started out with a graphical display of evidence that industry shakeouts of firms occur earlier in industries where technological progress is faster. We argued that other models of shakeouts were not able to explain this fact, whereas our vintage-capital model does so by predicting earlier replacement when capital-embodied technological progress is fast. We supported this claim with evidence from the airline industry that showed firm exits to be positively related to the age of the capital stock.

By inferring technological progress in the inputs from the decline in the price of the output as our model predicted, we found that the model fits fairly well the negative relation between technological progress and the onset of the shakeout. Moreover, we found that subsequent investment spikes, too, are also more frequent where technological progress is fast.

References

[1] Agarwal Rajshree. "Evolutionary Trends of Industry Variables." *International Journal of Industrial Organization* 16, no. 4 (July 1998): 511-25.

[2] Aghion, Philippe and Howitt, Peter, "A Model of Growth Through Creative Destruction," *Econometrica,* March 1992, 60(2), 323-351.

[3] Aizcorbe, Ana and Samuel Kortum. "Moore's Law and the Semiconductor Industry: A Vintage Model." *Scand. J. of Economics* 107, no. 4 (2005): 603–630.

[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] Arrow, Kenneth. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29, no. 3 (June 1962): 155 - 173..

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

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

[8] Baily, Martin, Charles Hulten and David Campbell "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity, Microeconomics* (1992): 187-249

[9] Boucekkine, Raouf, Marc Germain, and Omar Licandro. "Replacement Echoes in the Vintage Capital Growth Model." *Journal of Economic Theory* 74 (1997): 333-48.

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

[11] Cooper, Russell and John Haltiwanger "Evidence on Macroeconomic Complementarities." *Review of Economics and Statistics* 78, no. 1 (February 1996): 78-93.

[12] 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.

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

[14] 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.

[15] Gavazza, Alessandro. "Leasing and Secondary Markets: Theory and Evidence from Commercial Aircraft", Yale University, March 2007.

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

[17] Grossman, Gene M. and Helpman, Elhanan, "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, January 1991, 58(1), 43–61.

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

[19] 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.

[20] Johansen, Leif. "Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis." *Econometrica* 27, No. 2 (April 1959): 157-176.

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

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

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

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

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

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

[27] ———, and Kenneth Simons. "Industry Shakeouts and Technological Change." *International Journal of Industrial Organization* 23, no. 2-3 (2005): 23-43.

[28] McGrattan Ellen, and James A. Schmitz, Jr.. "Maintenance and Repair: Too Big to Ignore." *Federal Reserve Bank of Minneapolis Quarterly Review* 23, No. 4 (Fall 1999): 2–13.

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

[30] Mitra, Tappan, Debraj Ray and Rahul Roy. "The Economics of Orchards: An Exercise in Point-input, Flow-output Capital Theory." *Journal of Economic Theory*, 53, no. 1 (1993): 12-50.

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

[32] Parente Stephen L., 1994. "Technology Adoption, Learning-by-Doing, and Economic Growth," *Journal of Economic Theory.* Elsevier, vol. 63, no. 2 (1994): 346-69.

[33] Salvanes, Kjell and Ragnar Tveteras. "Plant Exit, Vintage Capital and the Business Cycle." *Journal of Industrial Economics* LII, no. 2 (June 2004): 255-76.

[34] Samaniego, Roberto M. "A Theory of Entry and Exit with Embodied Technical Change." Unpublished manuscript, George Washington University, 2006.

[35] Serrano, Carlos. "The Dynamics of the Transfer and Renewal of Patents." University of Toronto, March 2006.

[36] Tse, Chung Yi, "The Distribution of Demand, Market Structure and Investment in Technology," *Journal of Economics*, October 2001, 73(3), 275–297.

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

[38] Wang, Zhu. "Learning, Diffusion and Industry Life Cycle", manuscript, Federal Reserve Bank of Kansas City, 2006.

8 Appendix

8.1 The aircraft data

The data that Figures 4 and 5 portray come from a file, complied by a producer of a computer based aviation market information system and safety management software, that records the history of every non–military aircraft manufactured. They are described by Gavazza (2007, Sec. 5.1) and at *http://www.flightglobal.com/StaticPages/Acas.html* . Coverage begins in 1942 and ends at April 2003. For each aircraft, it records the usual identification information, such as the manufacturer, the model, etc., and (most important for our purposes) the first and the last date that each aircraft was operated by individual carriers. We aggregate such information at the level of an individual carrier to calculate the average age of capital of the carrier at a given point in time. The analysis is restricted to narrow-body (100 seats +) and wide–body passenger jets.

Measurement of entry and exit.—We say that a carrier has "entered" in the year in which it was first recorded as having operated a passenger jet. A carrier has "exited" in the year in which it was last recorded as having operated a passenger jet. A carrier has "survived" a certain year if it entered before that year and if it has not exited by the end of the year. The data contain information on mergers; a company that was acquired is not counted as an exit.

Measurement of age.—The age of a jet at a point in time is the time elapsed since the jet was first delivered. The average age of a carrier's capital is the average age of all the jets that it operated on January 1 of the year.

8.2 Proofs and derivations

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 5.—(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 $\frac{\partial T}{\partial r} = -\frac{\partial \Phi}{\partial r} / \frac{\partial \Phi}{\partial T} > 0$. ■

Derivation of the Likelihood Function used for the estimates in Figure 7.—. For each industry i, if we know c_i and $\widehat{g}_i$, we know $\widetilde{T}_i = \widetilde{T}(r, c_i, \widehat{g}_i)$. We would like to solve $\widetilde{T}_i$ as a function of c_i, so that we can derive the distribution of $\widetilde{T}_i$ from the assumed log–normal distribution on c_i. But since $\widetilde{T}(r, c_i, \widehat{g}_i)$ does not admit a closed form solution, the exact functional relationship between $\widetilde{T}_i$ and c_i is unknown, and we linearize it in c. After substituting $x_i \equiv \log c_i$ and $g = \hat{g}_i$ into (9), it reads

$$\left(\frac{r}{e^{x_i}} + 1\right)(r + \hat{g}_i) = \hat{g}_i e^{-r\widetilde{T}_i} + re^{\hat{g}_i \widetilde{T}_i}. \tag{16}$$

Taking total derivatives at $x_i = \overline{x}$, where $\overline{x} = \log \overline{c}$, gives

$$-re^{-\overline{x}}\left(r + \hat{g}_i\right) dx_i = r\hat{g}_i \left(e^{\hat{g}_i \overline{T}_i} - e^{-r\overline{T}_i}\right) dT_i$$

where $\bar{T}_i \equiv \tilde{T}\left(0.07, e^{\overline{x}}, \hat{g}_i\right)$. Then

$$-\frac{dT_i}{dx_i} = e^{-\overline{x}} \frac{r + \hat{g}_i}{\hat{g}_i} \frac{1}{e^{\hat{g}_i \bar{T}_i} - e^{-r\bar{T}_i}} \equiv \beta_i$$

Hence $T_i \simeq \tilde{T}\left(0.07, \overline{x}, \hat{g}_i\right) - \left(x_i - \overline{x}\right)\beta_i$. If $x_i \sim N\left(\overline{x}, \sigma^2\right)$, then approximately, $T_i \sim N\left(\bar{T}_i, \beta_i^2 \sigma^2\right)$. The density of T_i is $f\left(T_i\right) = \frac{1}{\sqrt{2\pi}\beta_i \sigma} \exp\left(-\frac{1}{2}\left[\frac{T_i - \bar{T}_i}{\beta_i \sigma}\right]^2\right)$. Letting $F\left(T_i\right)$ denote its CDF, the likelihood is

$$L = \prod_{i=1}^{2}\left(1 - F_{T_i}\left(T_i\right)\right) \prod_{i=3}^{n} f_{T_i}\left(T_i\right),$$

where $i = 1, 2$ denote the two censored observations.

Conditional Expectation in Figure 7.—Take the Nylon observation for example. The observation is truncated at 21. As we showed in the previous paragraph, $T_i \sim N\left(\bar{T}_i, \beta_i^2 \sigma^2\right)$ approximately. By setting $\overline{x}$ and σ equal to their respective point estimates of -2.56 and 2.27, we obtain estimates of $\bar{T}_i$ and $\beta_i^2 \sigma^2$. Then we calculate the expected value of T_i conditional on $T_i \geq 21$.