Machine Replacement and the Business Cycle: Lumps and Bumps

By RUSSELL COOPER, JOHN HALTIWANGER, AND LAURA POWER*

This paper explores investment fluctuations due to discrete changes in a plant's capital stock. The resulting aggregate investment dynamics are surprisingly rich, reflecting the interaction between a replacement cycle, the cross-sectional distribution of the age of the capital stock, and an aggregate shock. Using plant-level data, lumpy investment is procyclical and more likely for older capital. Further, the predicted path of aggregate investment that neglects vintage effects tracks actual aggregate investment reasonably well. However, ignoring fluctuations in the cross-sectional distribution of investment vintages can yield predictable nontrivial errors in forecasting changes in aggregate investment. (*JEL* E22, E32)

This paper investigates the aggregate implications of the nonconvex adjustment of equipment at the plant level. Our emphasis contrasts quite sharply with the neoclassical model of investment where the accumulation of capital reflects the slow adjustment of capital to its desired value.[1] The model we pursue assumes that small adjustments of the capital stock are either infeasible or undesirable. In particular, many investment projects (e.g., the construction of a new plant or the purchase of large machines) are not possible in small quantities. Further, the costs of adjusting the capital stock may be nonconvex, as described, for example, by Michael Rothschild (1971). As a consequence, at the plant level one may see periods of low investment activity followed by bursts of investment activity, i.e., investment spikes. Moreover, to the extent that these periods of high activity are not smoothed by aggregation, the nonconvex cost of adjustment model may have interesting aggregate implications.[2] In particular, the behavior of aggregate investment can be highly dependent on the cross-sectional distribution of the age of the capital stock.

As documented further below, there are two

* Cooper: Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215; Haltiwanger: Department of Economics, 3105 Tydings Hall, University of Maryland, College Park, MD 20742; Power: Office of Tax Analysis, U.S. Department of the Treasury, 1500 Pennsylvania Avenue, NW, Washington, DC 20220. All three authors are research associates at the Center for Economic Studies, U.S. Bureau of the Census, at which the empirical work for this project was conducted. We are most grateful to Mark Doms and Tim Dunne for sharing data with us and to Lucia Foster and Alok Johri for superb research assistance. Paul Beaudry, Steve Berry, Ricardo Caballero, David Card, Zvi Eckstein, Douglas Gale, Simon Gilchrist, Zvi Hercowitz, John Kennan, Sam Kortum, John McClelland, Ariel Pakes, Michael Riordan, Mahlon Straszheim, and two anonymous referees provided helpful comments and suggestions during this research. Financial support from the National Science Foundation and the Census Fellows Program is gratefully acknowledged. We thank seminar participants at the Board of Governors of the Federal Reserve System, Boston University, the U.S. Bureau of the Census, the European University Institute, Johns Hopkins University, Harvard University, IGIER, Princeton University, Rice University, the University of Bologna, the University of Illinois, the University of Pennsylvania, the Stanford Institute for Theoretical Economics, UCLA, the University of Virginia, the University of Wisconsin, Yale University, and the NBER 1994 Summer Institute for helpful comments.

[1] This point is neatly seen in the transition path dynamics displayed by Robert G. King et al. (1988), where the speed of adjustment to the steady state is largely determined by the curvature of the utility function. Andrew B. Abel (1990) provides a synthesis of the neoclassical investment model with convex adjustment costs. Abel and Janice C. Eberly (1994) model the investment choice at the firm level allowing for fixed costs in the adjustment process in a stochastic environment. They provide conditions for inaction on investment due to these fixed costs of adjustment. Our emphasis is on the timing of discrete investment relative to the business cycle and the aggregation of discrete choices.

[2] For macroeconomists, this is a critical point. Giuseppe Bertola and Ricardo J. Caballero (1990) discuss the role of aggregate shocks in creating synchronization of discrete decisions, while Cooper and Haltiwanger (1992) stress the role of aggregate complementarities.

key observations that motivate this view of investment activity. First, as reported by Mark Doms and Timothy Dunne (1994), a significant fraction of investment activity at the plant level is associated with large variations in the capital stock: i.e., investment is largely a lumpy activity.[3] Second, aggregate variations in investment are associated with an increase in the frequency of investment spikes at the plant level. That is, investment activity at the extensive margin plays a major role in aggregate investment behavior.

We study a variant of the machine replacement model analyzed by Cooper and Haltiwanger (1993) in which individual plants decide upon the timing of machine replacement. Our earlier paper stressed deterministic cycles and argued that downturns were a good time for retooling since the opportunity costs were lower and, due to the deterministic cycles, an upturn in activity was sure to follow. In contrast, the emphasis here is on embedding the machine replacement problem in a dynamic stochastic environment. The theme that retooling should take place in periods of low activity must be qualified in this more general environment. The relationship between the gains to retooling and the state of the economy depends on the specification of the underlying stochastic process. Further, the costs of retooling depend on the nature of the adjustment cost specification: lump sum versus proportional costs. We find that recessions are not always the optimal time to retool. In particular, if shocks are persistent and adjustment costs are largely fixed, investment spikes are predicted to be procyclical.

Our empirical approach is to explore models using thresholds: i.e., to define investment spikes as episodes of relatively large investment expenditures.[4] This allows us to analyze the relationship between the probability of a large investment episode and the time elapsed since the last spike (we refer to this time as the plant's investment age). We use the Longitudinal Research Database (LRD) to investigate this relationship and then consider the consequences of observed lumpy investment for the behavior of aggregate investment.

Our first empirical exercise examines the implications of the interaction of the empirical hazards (by year, the fraction of plants having an investment spike in each investment age class) and the associated cross-sectional age distribution for aggregate investment fluctuations. We find that the predicted path of aggregate investment generated by *ignoring* fluctuations in the cross-sectional distribution of the age of the capital stock tracks aggregate fluctuations in actual investment quite well. That is, investment movements are largely explained by factors other than the evolution of the cross-sectional age distribution. However, ignoring the fluctuations in the cross-sectional age distribution can yield predictable nontrivial errors in forecasting changes in aggregate investment after large swings in aggregate investment. It is during such periods that the cross-sectional age distribution diverges substantially from the steady-state distribution.

A second empirical exercise is to obtain estimates of plant-level hazard functions using a flexible semiparametric specification. The key result that emerges at the micro level is that the probability of a plant experiencing a large investment episode is increasing in the time since the previous episode. Thus, at the plant level, bursts of investment are followed, on average, by periods of low investment. This is in contrast to the usual presumption of positive serial correlation in investment activity which emerges from the standard convex adjustment cost model.

I. A Machine Replacement Model

Consider the problem of an individual producer with a given stock of capital. The productivity of this capital is influenced both by a

[3] Øivind Nilsen and Fabio Schiantarelli (1998) find similar evidence in Norwegian micro data. For production units, they find zero equipment investment in 21 percent of their observations.

[4] Put differently, assume that the cost of adjustment function is logistic so that adjustment costs are convex only for small changes. Then our investment threshold occurs at an investment level near the reflection point of the adjustment cost function. As discussed in our conclusion, the alternative is to alter the theoretical model to accommodate both forms of investment flows and then test this model at the plant level. Abel and Eberly (1996) take a version of this approach and analyze firm-level Compustat data. They find evidence of nonlinearities.

shock to total factor productivity and the "age" of the capital. Given the state of productivity, the agent makes the discrete choice between replacing existing capital with a new machine or continuing to use the capital for another period.[5] In making this decision, the producer calculates the discounted expected benefits of more productive capital relative to the current costs of replacement. The gain to replacement is that a new piece of machinery is more productive as it reflects some aspects of technological progress.[6] There are two types of costs of replacement. One is the direct loss of output associated with the acquisition and installation of new capital goods. Second, the process of installing the new machines and retraining workers reduces productivity in the plant. As we shall see, the nature of the adjustment costs and the structure of the stochastic process governing the shocks jointly determine the timing of replacement activity.

Our specification focuses on lumpy investment by assumption and thus fits into the ongoing literature on nonconvex adjustment processes.[7] Formally, consider the optimization problem of an individual agent. Producer $i = 1, 2, \cdots I$ maximizes

$$E_0 \sum_{t=0}^{\infty} B^t Y_t^i \tag{1}$$

subject to:

$$Y_t^i = A_t^i \theta_t^i f(K_t^i) - z_t^i F_t. \tag{2}$$

$$K_{t+1}^i = \begin{cases} (1-\delta)K_t^i & \text{if } z_t^i = 0 \\ \kappa_{t+1} & \text{if } z_t^i = 1 \end{cases} \tag{3}$$

where $\kappa_t = \mu \kappa_{t-1}$ and $\mu \geq 1$ is the pace of exogenous technological progress. The choice variable in this problem is z_t^i where $z_t^i = 1$ means that replacement is chosen in period t.

The first equation is the objective function for the individual producer. In this simplified model, the producer maximizes the discounted present value of profits defined as output less adjustment costs, where the discount rate is $B \in (0, 1)$. The right side of (2) represents the production process, net of the adjustment costs. Gross output is an increasing, concave function of the stock of capital, given by $f(\cdot)$ but also depends on the state of productivity, denoted by A_t^i, which contains an idiosyncratic and a common component.[8]

The production relation also includes the two costs of adjusting the capital stock described above. First, there is an opportunity cost associated with the diversion of labor and other inputs away from production and into adjustments in the capital stock. This is parameterized by θ_t^i which equals 1 during nonreplacement periods and equals $\lambda < 1$ during each replacement period.[9] Second, there is a fixed cost of adjustment given by $z_t^i F_t$ which reflects both the lost output from the acquisition of the new capital and other fixed adjustment costs, where $z_t^i = 1$ if and only if (iff) replacement is occurring in period t.

[5] This problem is similar to that studied by John Rust (1987), though we emphasize the aggregate implications of the replacement decision.

[6] Hence this model implies embodied technological progress as a source of growth.

[7] Caballero and Eduardo M. R. A. Engel (1993) synthesize one approach to this type of problem. In their model, the starting point of the analysis is a hazard function which depends on the difference between the current and target value of a particular variable, such as employment or capital. This hazard function may be parameterized and estimated from either micro or macro data. This procedure has been used to characterize employment adjustments [Caballero and Engel (1993) and Caballero et al. (1997)], and to study investment flows [Caballero and Engel (1994) and Caballero et al. (1995)]. The current paper shares many themes with these other efforts, though it characterizes a hazard function that depends on different state variables: the time since last replacement (age of the capital) and the aggregate state. Accordingly, there are some important differences across these approaches in terms of what the hazard function and the cross-sectional distribution represent. As discussed in Section III, subsection B, the differences in approaches lead to quite different empirical strategies as well.

[8] That is, assume $A_t^i = A_t \varepsilon_t^i$. Of course, at this point in the analysis this distinction is not important. It will become very relevant for understanding the aggregate implications of this model. While we consider a model with technology shocks, one could instead consider an economy in which the shocks to the production process represent changes in relative product demands either across producers or over time.

[9] Note that this naturally implies heterogeneity in adjustment costs across firms if $\lambda < 1$, a point emphasized by Caballero and Engel (1994) as well.

The final equation expresses the time path of capital and hence its link to the producer's choice. We denote by κ_t the leading-edge technology; however, each producer's actual capital (K_t) is generally not the best technique available. In every period, the producer has the choice of replacing its current machine with the leading technology.[10] If the producer chooses replacement ($z_t^i = 1$), then the adjustment costs described above are borne in period t and the producer's capital in the next period is κ_{t+1}. If the producer chooses not to replace then the machine depreciates at rate δ.[11] Note then that the gains to replacement reflect both the pace of technological progress (μ) and the rate of physical depreciation of a machine.

There are two interpretations of this optimization problem. The most direct is that the economy consists of I independent producers financing investment from their own output and consuming the remainder. A second interpretation comes from the centralized problem of a planner making machine replacement decisions at I sites with a representative agent consuming the difference between output and investment. In the case of linear utility, the choices across sites are independent and (1)–(3) would then represent the planner's problem at one of the sites. When utility of the representative agent is strictly concave, our formulation misses the congestion effects that would arise in the planner's problem. Solving this problem is more difficult since the state space would then include the entire cross-sectional distribution of plant ages.[12]

The underlying exogenous technological progress in this economy makes this into a nonstationary problem. For our analysis, it is convenient to consider a stationary version of this economy. Let $x_t = X_t/\kappa_t$ so that lowercase letters represent values which are normalized by the current value of the leading technology. Further, suppose that the fixed cost of installing the capital is proportional to the leading-edge technology, i.e., $F_t = F\kappa_t$ and that $f(\cdot)$ exhibits constant returns to scale. Then, (1)–(3) can be rewritten as

$$E_0 \sum_{t=0}^{\infty} \beta^t y_t^i \tag{4}$$

subject to:

$$y_t^i = A_t^i \theta_t^i k_t^i - z_t^i F. \tag{5}$$

$$k_t^i = \begin{cases} \rho k_{t-1}^i & \text{if } z_{t-1}^i = 0 \\ 1 & \text{if } z_{t-1}^i = 1. \end{cases} \tag{6}$$

In this version of the optimization problem, the discount rate (β) equals $B\mu$. We assume that the rate of technological progress (μ) is not too big so that $\beta < 1$. In (6), $\rho = (1 - \delta)/\mu$ and lies between zero and one, reflecting both physical depreciation and obsolescence. With our normalization, machine replacement ($z_t^i = 1$) implies that the state of the machine is 1 in the next period; otherwise, the capital is only a fraction ρ of its size in the previous period.

To analyze this problem, we utilize a dynamic programming approach where the states are the age of the capital stock (k), the aggregate shock (A), and the idiosyncratic shock (ε). The value function $V(k, A, \varepsilon)$ satisfies the following functional equation:

$$V(k, A, \varepsilon) = \max[V^R(k, A, \varepsilon), V^N(k, A, \varepsilon)], \text{ where} \tag{7}$$

$$V^N(k, A, \varepsilon) = A\varepsilon f(k) + \beta E_{A'|A,\varepsilon'} V(\rho k, A', \varepsilon'), \text{ and}$$

$$V^R(k, A, \varepsilon) = A\varepsilon f(k)\lambda - F + \beta E_{A'|A,\varepsilon'} V(1, A', \varepsilon')$$

[10] For simplicity we assume that the replacement process renders previous vintages worthless both in the plant and to others. The model could be supplemented to allow for the interaction of multiple vintages or a resale market for used capital. Further, the replacement decision is irreversible.

[11] The model could be easily supplemented to allow for some small levels of investment (e.g., to replace broken machines and so forth) even when large investments are not taking place. In particular, the production function could be interpreted as net of any stochastic depreciation. In this case, the agent would, by assumption, incur investment expenditures necessary to restore the capital.

[12] Cooley et al. (1994) analyze a version of this more complicated planner's problem.

where the expectation over A' is taken using the conditional distributions, $\Phi(A'|A)$. In the analysis, we assume that the idiosyncratic shock is independently and identically distributed (i.i.d.) and that the aggregate shock follows a first-order Markov process.[13] The current aggregate shock plays two roles in this dynamic programming problem. First it has a direct influence on current productivity and second it is informative about future productivity through $\Phi(A'|A)$.

Our first result is the existence of a solution to this functional equation. Proofs of this and other propositions are in the Appendix.

PROPOSITION 1: *There exists a solution to the functional equation.*

The solution to (7) entails replacement iff $V^R > V^N$ given the state vector, (k, A, ε). Given the state, the probability of an investment spike is either zero or one. Because the idiosyncratic shock is not observable to the econometrician, we characterize this solution by a hazard function $H(k, A)$, which is the probability of replacing if the current capital stock is k and the *aggregate* state of productivity is A. Note that the hazard function conditions on the aggregate state only so that $H(k, A) \in [0, 1]$.

The policy rules for this problem reflect two forces: an underlying replacement cycle and the response of the agent to shocks in productivity. For the nonstochastic version of this problem, Cooper and Haltiwanger (1993) find there will be a deterministic replacement cycle in which the time between replacement will depend, inter alia, on the proportional and fixed costs of adjustment, λ and F. In the stochastic case, this underlying deterministic cycle will imply that the older the capital, the more likely is replacement.[14] Formally,

PROPOSITION 2: *$H(k, A)$ is decreasing in k.*

Further, the replacement decision will be influenced jointly by the realized value of the technology parameter and the age of the capital stock. The response of investment to this random variable depends on both the nature of the adjustment costs (λ and F) and the persistence of the shock. The agent would prefer to replace machines during a period where inputs are not very productive (reflecting $\lambda < 1$) and would also prefer to have a new machine available when productivity is high.

As indicated by Proposition 3, if aggregate shocks are i.i.d. and there are no opportunity costs of adjustment, then the replacement probability is independent of A.

PROPOSITION 3: *If $F > 0$, $\lambda = 1$ and A is i.i.d., then $H(k, A)$ is independent of A.*

However, if adjustment costs are proportional to output, then the opportunity cost argument implies that replacing investment in good times is quite costly; i.e., $H(k, A)$ is decreasing in A. Proposition 4 verifies this claim, retaining the assumption of i.i.d. shocks. Finally, Proposition 5 shows that if shocks are positively serially correlated and adjustment costs are fixed, then replacement investment is procyclical since expected benefits are higher due to the persistence in A.

PROPOSITION 4: *If $F = 0$, $\lambda \in (0, 1)$ and A is i.i.d., then $H(k, A)$ is decreasing in A.*

PROPOSITION 5: *If $F > 0$, $\lambda = 1$ and $\Phi(A'|A)$ is decreasing in A, then $H(k, A)$ is increasing in A.*

II. Simulations

Obtaining analytic solutions to the stochastic dynamic programming problem to study the aggregate implications of the replacement problem is quite difficult. This reflects an inability to obtain a closed-form decision rule for the individual producer and, more importantly,

[13] Though A and ε enter multiplicatively in the production function, they have different stochastic properties and hence enter as separate state variables.

[14] As in John Rust (1987), the optimization problem yields policy functions that depend on the state variables. Equivalently, the probability of taking an action depends on the current period costs relative to the future benefits, given by the value function. Thus, the specification of the hazards in terms of the state variables k and A is directly linked to a specific characterization of the optimization problem. This contrasts with the semireduced form approach pursued, for example, in Caballero et al. (1995), where the adjustment rate function is posited to be a function of the difference between the desired and actual stock of capital.

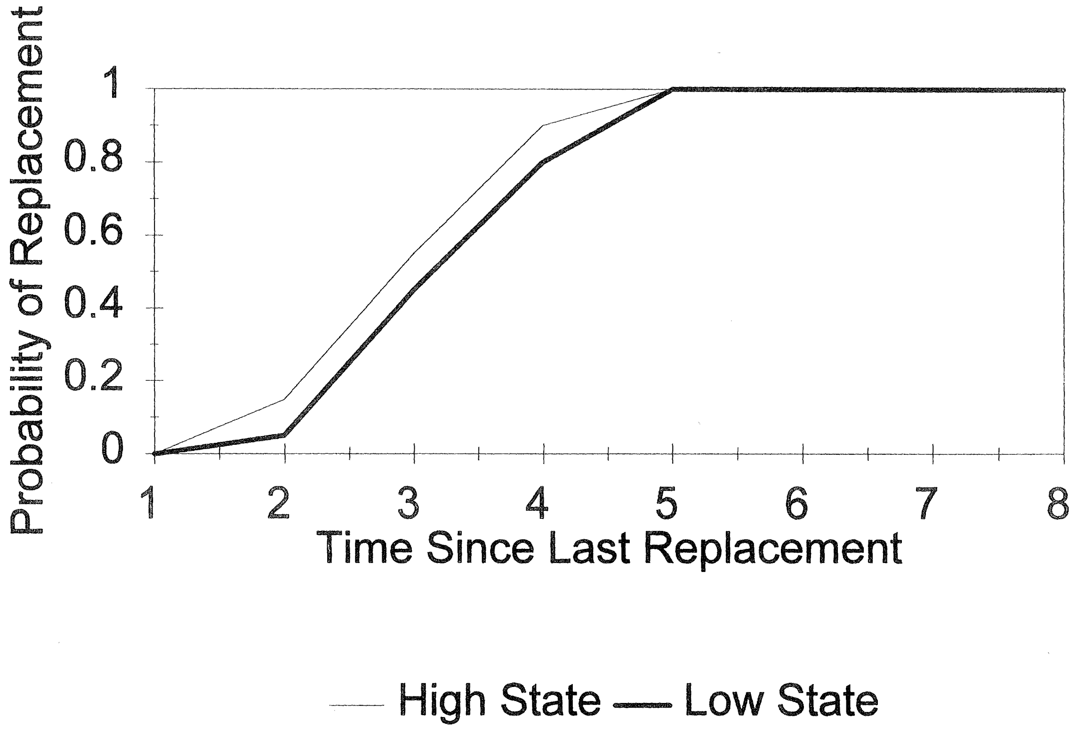

FIGURE 1. THEORETICAL HAZARD FOR MACHINE REPLACEMENT

difficulty dealing with the heterogeneity across multiple producers. So, to better understand the implications of the solution to this problem for aggregate investment, we have produced simulation results.

A. *Approach and Individual Optimization Results*

For the numerical analysis, we solve the dynamic programming problem given in (7) using value function iteration. The state space has three components. First, the capital stock lies in the discrete set given by $\{1, \rho, \rho^2, \rho^3, \cdots\}$ where ρ reflects both depreciation and obsolescence. Second, we assume that there are two aggregate shocks $\{A_H, A_L\}$ with $A_H > A_L$.[15] Third, we assume that the idiosyncratic shocks take values in a discrete set.[16] The idiosyncratic shocks are important to the analysis since they generate a nondegenerate distribution of the capital stock.

We assume that the rate of capital depreciation (δ) is 0.1, the discount rate for the agent (B) is 0.9 and the technological progress parameter (μ) is 1.00. The aggregate shock is determined by a first-order Markov process where $\text{Prob}(A_{t+1} = A_j | A_t = A_j) = 0.9$ for $j = L, H$. Since we have set $\mu = 1$, $\rho = (1 - \delta)$ and $\beta = B = 0.9$. The fixed cost (F) is set at 0.2 and the opportunity cost of replacement $(1 - \lambda)$ is 0.25.

Figure 1 presents the hazard functions for this economy conditional on the aggregate state. The horizontal axis measures the time since last replacement. Thus, from Proposition 2, the hazard is increasing in the time since last replacement, given A. In the discussion that follows,

[15] For the simulations, $A_L = 0.75$ and $A_H = 1.25$.

[16] The idiosyncratic shocks were multiplicative and took on 20 values between 0.4 and 1.6, with a mean of one. Throughout our experiments, these shocks were i.i.d.

we term this an "increasing hazard" to focus on the relationship between the probability of replacement and the time since last replacement given A.[17] Further, fixed costs and persistent shocks combine to imply a procyclical hazard: the probability of replacement increases during periods of high profitability.

B. *Aggregate Investment*

One of our primary motivations is to provide a link between discrete investment choices at the producer level and aggregate investment, defined here as the fraction of plants having an investment spike. That relationship is greatly enriched by the interaction of aggregate shocks and underlying heterogeneity across plants. To make this point, we study each effect in isolation and then focus on their interaction. Throughout this section, we use the parameters described earlier.

Before proceeding, it is useful to have a specification for the evolution of the capital stock and a measure of aggregate investment. Denote by $g_t(k)$ the period t cross-sectional distribution of capital, indexed by vintage (k). The aggregate investment rate in period t, given an aggregate shock (A), is

$$I_t(A) = \sum_k H(k, A) g_t(k). \tag{8}$$

To understand this expression, consider a given vintage of capital. For that vintage and aggregate productivity A, the hazard $H(k, A)$ is the probability of investment. Summing this over k, weighted by the fraction of producers in each state, yields aggregate investment.[18] From this expression, the aggregate investment rate is determined jointly by the cross-sectional distribution as well as by the current state of productivity. Of course, the importance of the cross-sectional distribution is dependent upon the shape of the hazard function. Clearly, if $H(k, A)$, is independent of k, then the aggregate level of investment is independent of $g_t(k)$.

This accounting framework highlights the two influences on aggregate investment: shifts in the hazard and movements along a fixed hazard due to changes in the cross-sectional distribution. The key issue in the empirical analysis is disentangling these two forces.

The evolution of the cross-sectional distribution is given by

$$g_{t+1}(1) = I_t(A) \tag{9}$$

$$g_{t+1}(\rho k) = [1 - H(k, A)] g_t(k)$$

where $I_t(A)$ is given above. Thus the fraction of agents with new capital ($k = 1$) is simply the investment rate in a given period. Further, the fraction with capital of vintage ρk in period $t + 1$ is the fraction of producers with capital k in period t who choose not to invest.

A useful way to visualize the evolution of the cross-sectional distribution is given in Figure 2. Here the state space is depicted as a ladder. Producers on a particular rung in period t either go to the top of the ladder with probability $H(k, A)$ (i.e., they invest) or move down one rung with probability $1 - H(k, A)$. For a fixed value of A, the cross-sectional distribution will evolve to a stationary distribution. Aggregate shocks disrupt this process so that the resulting time path of investment reflects the response of investment to these shocks and the dynamics of the adjustment process of the cross-sectional distribution.

To better understand the underlying dynamics of this economy, consider first an extreme case in which there are no aggregate shocks in the system. The presence of the idiosyncratic shocks will lead to a nondegenerate distribution of the capital stock across agents. However, there will be no aggregate dynamics in that the economy will settle on a steady-state distribution.

Figure 3 displays the behavior of aggregate investment starting from a uniform distribution of producers across the capital state space in an economy without aggregate uncertainty. Initially there is a burst of investment since many producers start with relatively old capital. Over time the distribution evolves until a stationary distribution is obtained and the level of aggregate investment is constant. Each producer undertakes machine

[17] Still, the hazard depends on the two state variables (k, A) along with the idiosyncratic shock and thus should not be confused with a deterministic time-dependent policy function.

[18] Recall that there is no choice on the size of investment so that we normalize a "project" to be an investment level of 1, which means that the aggregate investment rate is the aggregate level of investment.

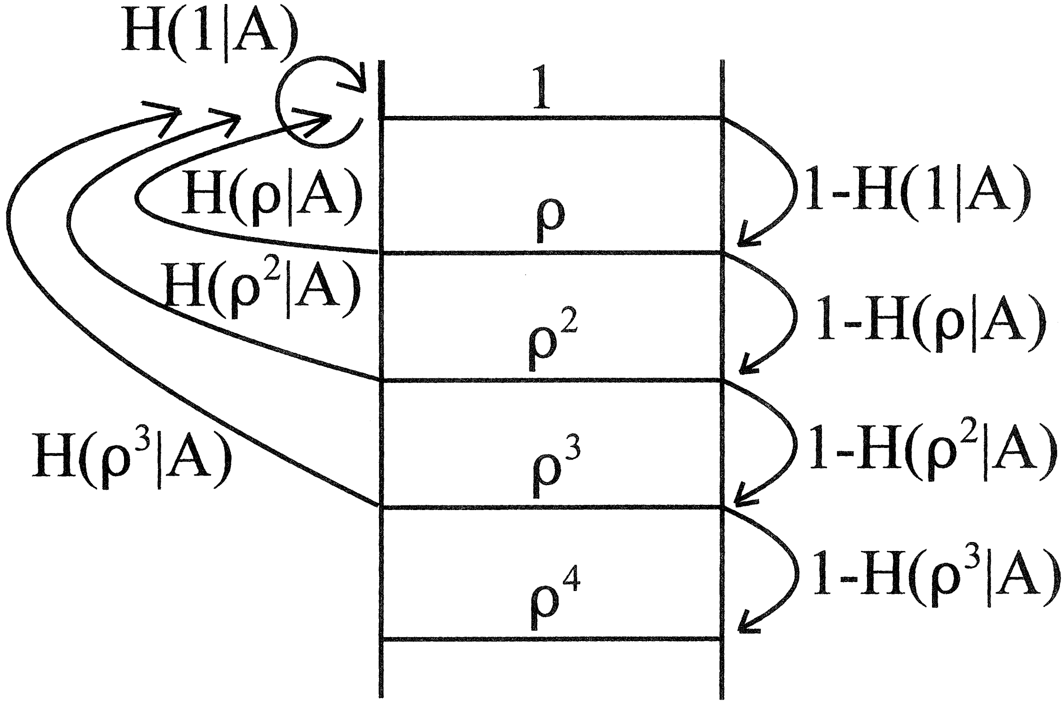

FIGURE 2. EVOLUTION OF THE CROSS-SECTIONAL DISTRIBUTION

Notes: H represents the hazard rate for investment, A is the aggregate shock, and ρ is the rate of depreciation. The rungs in the ladder represent different vintages of capital. The top rung is new capital, the second rung is one-year-old capital, the third rung is two-year-old capital, etc.

replacement at stochastic intervals of time: the deterministic machine replacement system is made stochastic by the producer-specific shocks. Yet, due to the assumed large number of plants, the economy has a stationary distribution with constant aggregate investment.

Note that the dynamics of adjustment towards the steady state leads to cycles in aggregate investment. These cycles arise from the interaction of the upward-sloping hazards, as in Figure 1, and the evolution of the cross-sectional distribution of capital vintages. For example, the initial burst of investment shown in period 1 of Figure 3 creates a relatively large fraction of "young plants" in periods 2 and 3. Since, from the hazards shown in Figure 1, young plants have low spike probabilities, the initial burst of investment is followed by a couple of periods of below-average investment.

Consider now the other extreme in which there are only aggregate shocks to the system. In this case all fluctuations in aggregate investment are driven by these shocks. As there is no heterogeneity, initial differences in the age of capital disappear over time as replacement becomes synchronized due to the common shocks.[19]

The most interesting case, of course, arises when there are both aggregate and idiosyncratic shocks. This is clearly the empirically relevant case since differences across plants given the state of the aggregate economy are quite significant.[20] Furthermore, it is the interaction between the cross-sectional distribution and the aggregate state that gives rise to such rich dynamics in the machine replacement problem. Note that when both types of shocks are present, the economy will have fluctuations in aggregate investment as well as a nondegenerate cross-sectional distribution of the capital stock.

[19] This is the type of synchronization described in Bertola and Caballero (1990).

[20] We discuss the empirical evidence on plant-level heterogeneity in some detail in the next section.

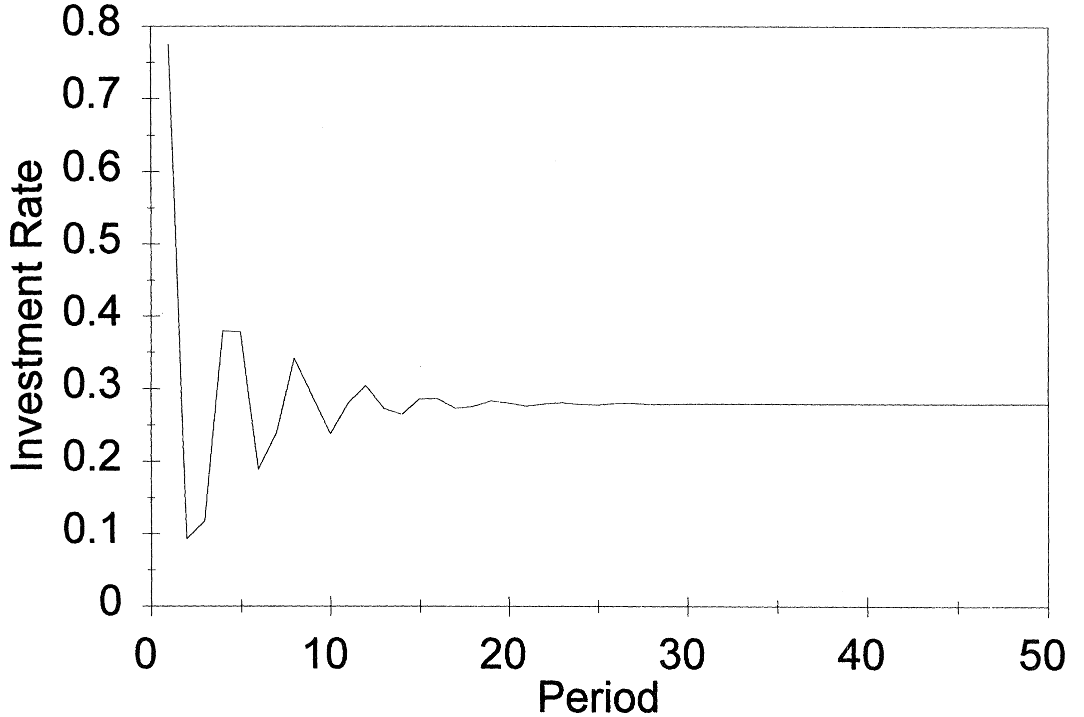

FIGURE 3. CONVERGENCE WITHOUT AGGREGATE SHOCKS—BASELINE PARAMETERS

A simulation in which there are two aggregate states and multiple idiosyncratic shocks is displayed in Figure 4. In this figure, the aggregate state is either 0 (corresponding to $A_t = A_L$) or 1 (corresponding to $A_t = A_H$). Note that investment spikes when the aggregate state changes from low to high. However, these spikes are not of uniform height since the fraction of investors depends on the underlying distribution. This is one way in which the cross-sectional distribution influences aggregate investment. Furthermore, between variations in the aggregate state, the economy experiences cycles in investment, often termed "echo effects."[21] This is similar to the pattern portrayed in Figure 3 as the economy converged to a stationary distribution. In an economy with both idiosyncratic shocks and aggregate shocks, these same transition dynamics govern the system between the switches in the aggregate state.

[21] As pointed out to us by Jess Benhabib, echo effects in vintage capital markets dates back to very early discussions of business cycles, as in R. C. O. Matthews (1959) and the discussion therein of Marx's work on equipment in the textile industry.

C. *Convex Adjustment Costs: A Competing Model*

Our machine replacement model of Section II predicts an upward-sloping hazard with respect to the time since last replacement. In practice, we investigate this prediction by examining whether the probability of having a large investment episode at the plant level is increasing in the time since the last large investment episode. Further, we take the evidence of large investment episodes, highlighted by Doms and Dunne (1994), as motivation for our study.

A competing hypothesis is that episodes of large plant-level investment simply reflect the skewed distribution of idiosyncratic shocks rather than an upward-sloping hazard. To examine this point further, consider a simple convex cost of adjustment model in which a firm maximizes the discounted present value of profits

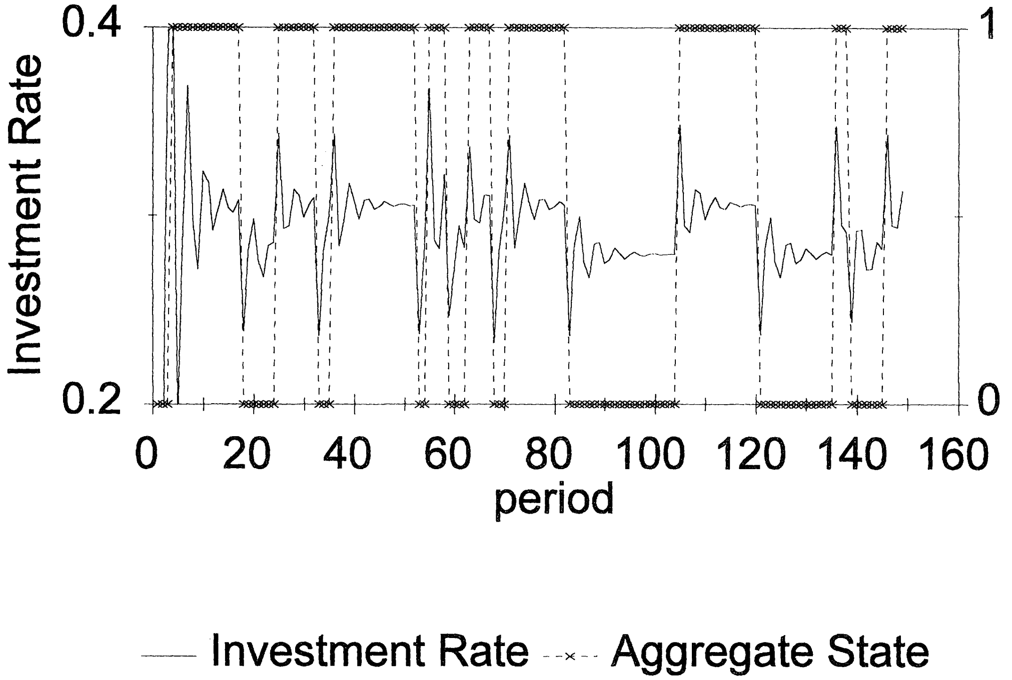

FIGURE 4. AGGREGATE INVESTMENT FLUCTUATIONS—BASELINE SIMULATION

(Π) given by $\Pi = SK - \frac{1}{2}\gamma I^2$, where K is the stock of capital, S is the state of productivity, and I is the level of investment. The flow of profits is equal to the flow of output from the capital less the quadratic costs of adjusting the capital stock. The capital stock is assumed to depreciate at a rate of $\delta \in (0, 1)$. The solution to the problem of maximizing the discounted value of profits is $I = a + bS$, where a and b are constants that depend on the parameters of the problem including those that characterize the stochastic process for S.

Figure 5 illustrates the results of simulating this model assuming that S is positively serially correlated.[22] The investment rate (vertical axis) fluctuates due to variations in the productivity of capital and, with enough dispersion in S, investment rate spikes can emerge. Note, however, that these investment spikes occur in bunches, reflecting the persistence in S and therefore implying that the probability of large investment in the current period is higher if there was an investment spike in the previous period. In this sense, this model and others with a convex cost of adjustment structure will not generally imply upward-sloping hazard functions.

III. Empirical Evidence on Investment Patterns

A. *Measurement and Basic Facts*

To evaluate the predictions of our theory, we use longitudinal data for approximately 6,900 plants in the U.S. manufacturing sector for the period 1972–1991. The data are a subset of the Longitudinal Research Database (LRD) representing large, continuously operating manufacturing plants over the sample. While entry and exit are obviously of interest in this context, in this study we focus our effort on the continuously operating plants.

Our objectives are threefold: (i) to characterize large-scale investment episodes at the plant

[22] For these simulations, the discount rate was set at 0.9, the rate of depreciation at 0.1, the serial correlation in S_t was set at 0.9, and γ was normalized at 1.

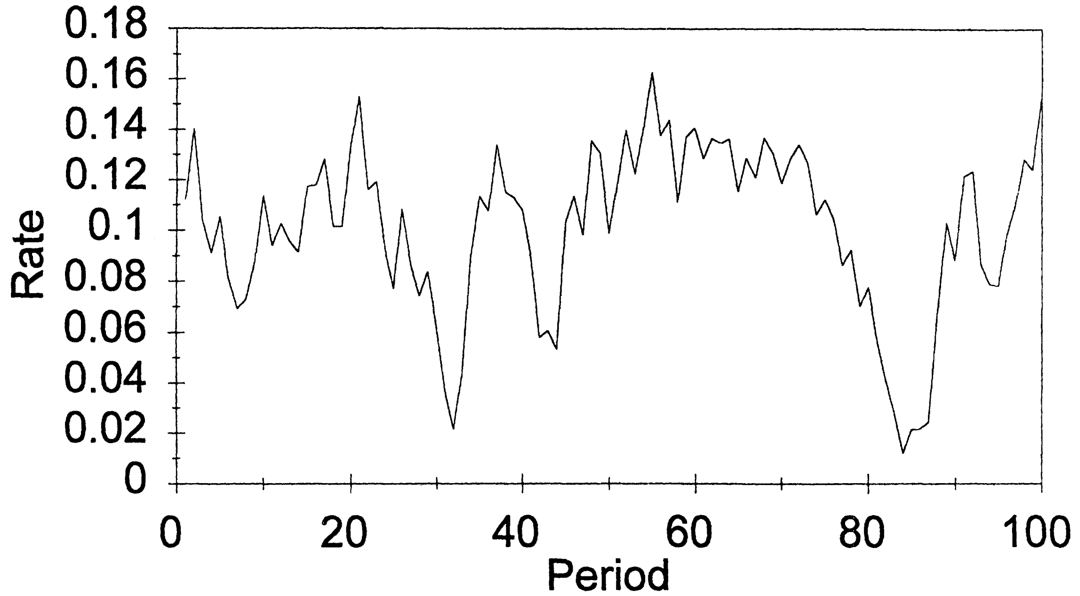

FIGURE 5. INVESTMENT RATE

level, (ii) to quantify the connection between the timing of large investment episodes at the plant level and aggregate fluctuations in investment, and (iii) to determine whether there is a systematic relationship in the timing of large investment episodes at the plant level consistent with the machine replacement model.

Four recent studies (Doms and Dunne, 1994; Power, 1994, 1998; Caballero et al., 1995) have used the LRD to study plant-level investment. Doms and Dunne conduct an exploratory data analysis characterizing the distributions of plant-level investment. For their 17-year sample (which is quite similar to ours), they find that plants concentrate about 50 percent of their cumulative 17-year investment in the three years surrounding the year with the largest investment. Power (1994) conducts an exploratory study of the timing of the large investment episodes and in this regard we build upon her work. Caballero et al. (1995) exploit the adjustment hazards approach developed by Caballero and Engel (1993). In this context, the latter approach involves estimating the plant-level relationship between investment and an estimate of the difference between the actual and desired capital stock. None of these papers address the issues that are the focus of this paper: the timing of large investment episodes at the plant level and the connection between the timing of large investment episodes at the plant level and aggregate fluctuations in investment.[23]

To begin our empirical analysis, we characterize the basic properties of the cross-sectional distribution of plant-level investment in our sample. Our analysis generally focuses on investment rates, defined as gross investment in machinery and equipment divided by the real capital stock at each of our 6,900 plants, as a partial control for plant size.[24] As discussed in greater detail in Cooper et al. (1995), our sample of 6,900 plants yields an aggregate

[23] Recently, Nilsen and Schiantarelli (1998) have conducted a very similar investigation using plant-level data from Norway. Their conclusions are quite similar to ours: lumpy investment is present, hazards are upward sloping, and aggregate shocks are the primary determinant of investment movements. Further, John McClelland (1997) studies the influence of permit requirements (a type of fixed cost) on the investment of plants in the pulp and paper industry, again finding upward-sloping hazards.

[24] To measure real capital, we initialized the real capital stock for each plant in 1972 by multiplying the book value of machinery in 1972 by the ratio of real to book value of the machinery for the two-digit industry in which the plant operates in 1972 and then the perpetual inventory method was used to calculate the real capital stock at each plant.

investment rate that mimics the aggregate investment rate of total manufacturing: for example, the correlation between aggregate gross investment for total manufacturing and gross investment measured in our sample is 0.982. Our plants are substantially larger than the typical plant—the average plant size in terms of employment in manufacturing is roughly 50 in 1987 while our average plant size is around 800 workers. These large plants constitute only a small fraction of the approximately 350,000 operating plants in U.S. manufacturing but account for about 45 percent of aggregate investment. The industrial mix of the plants in our sample corresponds quite closely to the industrial mix for all plants in manufacturing.

Figure 6 depicts a histogram of annual plant-level gross investment rates. The distribution is highly skewed to the right. Roughly 13 percent of plants have investment rates at or near zero (less than 0.02 gross investment rate), which reflects the fact that many plant-year observations involve little or no investment. The long right tail illustrates the fact that a relatively small, but important, fraction of plants experience a large investment episode in any given year.

To evaluate the contribution of the largest investment episodes to aggregate investment, we rank the investment rates for each plant over the 20-year sample period and examine the contribution of the ranked episodes to cumulative aggregate investment over the 20-year period.[25] Figure 7 depicts the results of this exercise. The sum of the investment associated with each plant's largest investment episode explains about 17 percent of cumulative aggregate investment. The top five investment years at each plant account for more than 50 percent of cumulative aggregate investment.[26]

Since our primary objective is to examine the timing of large investment episodes at the plant level and the implications of this behavior for aggregate investment, we must develop a method for defining a large or lumpy investment episode. That is, unlike the specification of the theoretical model, not all positive investment episodes are easily categorized as lumpy investment episodes.

To address this difficulty, we define a lumpy investment episode (hereafter termed an investment spike) as occurring if the *gross investment rate exceeds 20 percent.* This threshold is intended to eliminate routine maintenance expenditures (e.g., to repair and replace broken machinery) since, as noted in the discussion of our model, these expenditures may be reflected in measured investment. As long as the routine maintenance expenditures do not exceed 20 percent, the lumpy investment we study is distinct from these expenditures.

Of course this cutoff is somewhat arbitrary as is the use of this particular threshold. Cooper et al. (1995) argue that our results are robust to a variety of other definitions of investment spikes. Power (1994, 1998) studies spikes defined as abnormally high investment episodes relative to the typical investment rate experienced within a plant and obtains similar results on the shape of hazard functions. Finally, McClelland (1997) uses similar estimation techniques as those we use below and experiments with a number of investment-rate thresholds (20 percent, 25 percent, and 35 percent). He finds that the shapes of his estimated hazards are independent of threshold.

To provide some perspective on the time-series fluctuations and relative importance of large investment episodes, Figure 8 depicts the time-series fluctuations in the fraction of plants with gross investment rates in excess of 20 percent and their contribution to aggregate investment. The aggregate gross investment rate is plotted for comparison on the right vertical axis. A number of observations are worth making. First, plants with large investment episodes constitute about 20 percent of the plants but account for almost 50 percent of gross investment. The latter is a large percentage but far from 100 percent—that is, both lumpy and nonlumpy investment are important components of investment. Second,

[25] By cumulative aggregate investment we mean the sum of all investment for all plants and all years for our sample.

[26] These results appear to differ somewhat quantitatively (although not qualitatively) from the results by Doms and Dunne (1994). That is, they find that the largest investment episode at each plant accounts for 25 percent of cumulative investment over their sample for the typical plant. However, this primarily reflects differences in the length of our samples (their sample is 17 years, ours is 20 years). The difference in the length of the sample will by construction yield differences in the contribution of the largest investment episode to cumulative investment over the sample.

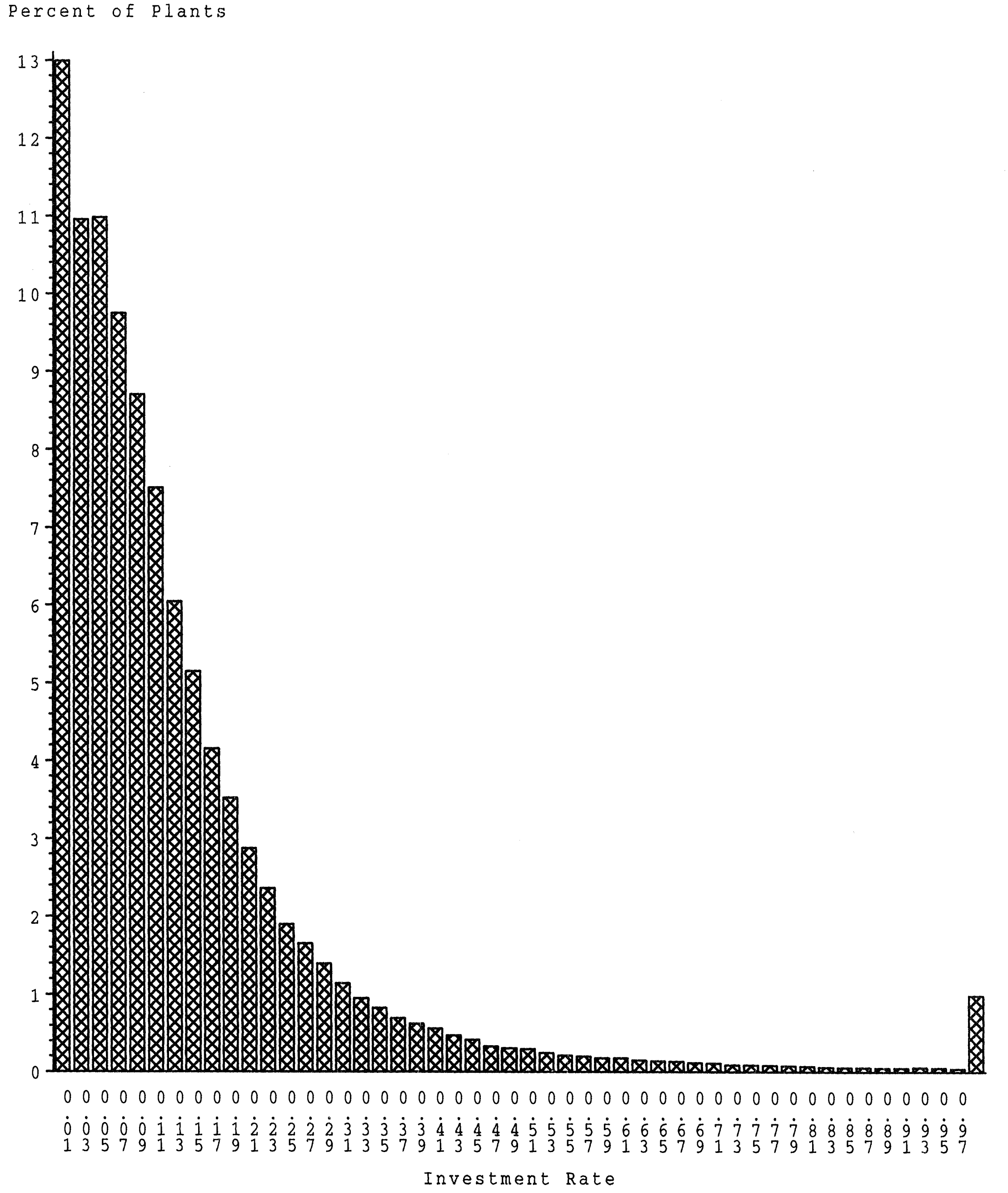

FIGURE 6. PLANT-LEVEL INVESTMENT DISTRIBUTION

Notes: Each bar represents the percent of plants with the depicted investment rate. The far right bar includes all plants with an investment rate greater than or equal to 0.98.

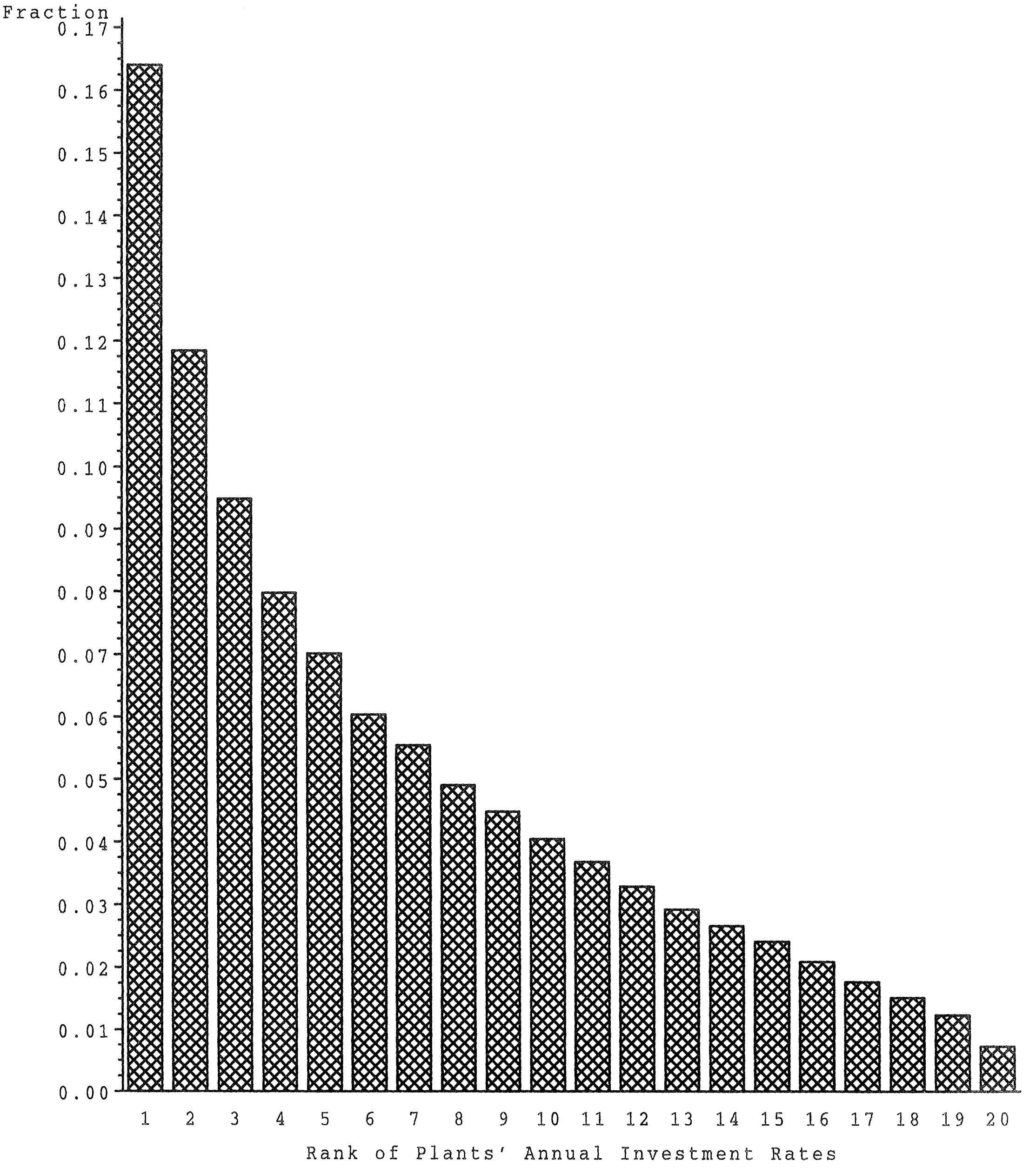

FIGURE 7. CONTRIBUTION OF RANKED ANNUAL INVESTMENT TO 20-YEAR TOTAL INVESTMENT

Notes: The bar above 2 represents the sum of investment associated with each plant's second-highest annual investment episode divided by the sum of each plant's total investment for the 20-year period. The bars over the other numbers represent analogous ratios.

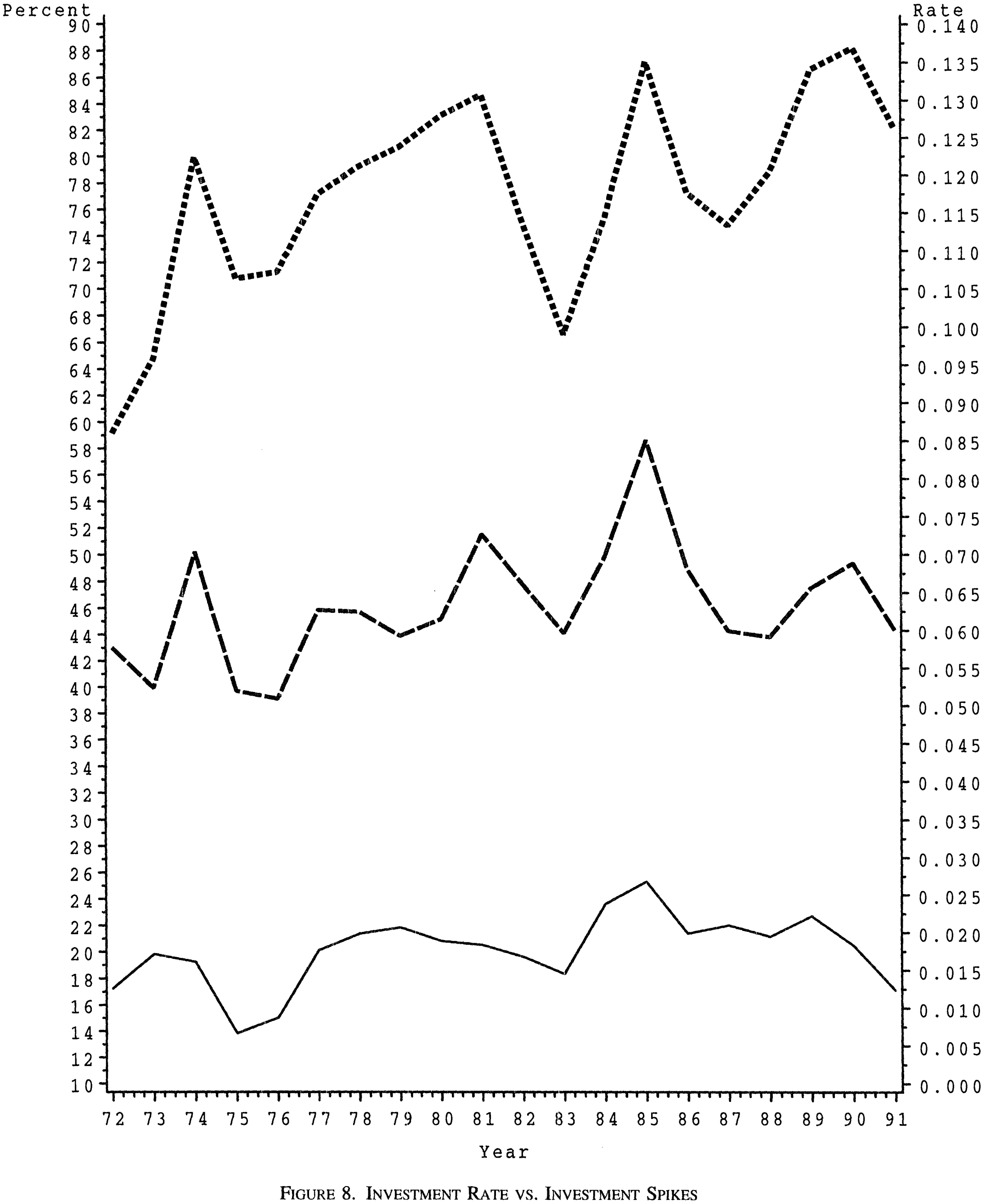

FIGURE 8. INVESTMENT RATE VS. INVESTMENT SPIKES

Notes: A plant is defined as having an investment spike if its investment rate is greater than 20 percent. The dotted line is the investment rate. The dashed line is the percent of investment accounted for by plants having investment spikes. The solid line is the percent of plants with investment spikes.

the fraction of plants having large investment episodes and the amount of investment accounted for by such plants is positively correlated with the aggregate investment rate.[27] This is important because much of the subsequent analysis is about the determinants of plants having large investment episodes and the cyclical fluctuations in such plants. These findings indicate that fluctuations in aggregate investment are closely linked to the fraction of plants experiencing large investment episodes.

B. *Aggregate Implications*

From equation (8), a key empirical issue is determining the relative importance of the two primary factors determining aggregate investment: shifts in the hazard functions and the evolution of the cross-sectional distribution of machine ages. To study these factors, consider the following partial representations of investment:

$$I_t^g = \sum_k H_{\mathrm{emp}}(k, A_t) g(k) \tag{10}$$

and

$$I_t^k = \sum_k H_{\mathrm{emp}}(k, \bar{A}) g_t(k). \tag{11}$$

Both (10) and (11) are variations on (8), which characterizes total investment as the interaction of a shifting hazard due to an aggregate shock (A_t) and an evolving cross-sectional distribution ($g_t(k)$). In (10), the hazard shifts due to variations in A_t, but the cross sectional is held fixed at its mean, $g(k)$. Thus the series I_t^g is the predicted pattern of investment assuming that investment changes only due to hazard shifts. Similarly, in (11), the hazard is assumed not to shift so that variations in investment arise solely through the evolution of the cross-sectional distribution. So, I_t^k is created by variations in the cross-sectional distribution, holding the hazard fixed at its average value.

The hazard function used in these calculations, denoted by $H_{\mathrm{emp}}(k, A_t)$, is the empirical hazard computed by taking, for each k, the ratio of plants that spike to the number of plants of that age in a given year. There is no structure imposed on the shape of the hazard by this calculation and hence we term this the empirical hazard. By construction, interacting this hazard with the actual cross-sectional distribution at any point in time ($g_t(k)$) yields the actual investment rate, I_t. Note that $H_{\mathrm{emp}}(k, \bar{A})$ represents the average hazard across years.

The top panel of Figure 9 depicts the empirical hazards for the years 1983–1987.[28] There are two features of the hazards worth noting. First, the empirical hazard is procyclical: during periods of high aggregate investment activity, the hazard is shifted up. This is analogous to the point raised in the discussion of Figure 8: periods of high investment are also times of more frequent investment spikes.

Second, the empirical hazard is downward sloping, indicating that replacement probability declines with capital age. The sharpest decline in the probability of having an investment spike occurs after investment age zero (i.e., plants who had a spike in the previous period). The large probability associated with having a spike in the period immediately following a spike reflects the multiyear spike phenomenon emphasized in the findings of Doms and Dunne (1994). That is, one of their key findings is that large investment episodes are often spread across two (or sometimes three) years. While this multiyear spike phenomenon may reflect economic factors, it is also possible that it reflects a form of measurement error induced by the calendar-year nature of the data. Large investment projects that start late in one calendar year and are completed in the subsequent year can easily yield this pattern. Results pre-

[27] The correlation between the aggregate investment rate and the fraction of plants with investment rates larger than 20 percent is 0.57. The correlation between the aggregate investment rate and the fraction of investment accounted for by plants with investment rates greater than 20 percent is 0.69.

[28] To minimize the impact of censoring in the decomposition exercise in this section, we perform the aggregate decompositions for the period 1980–1991. This permits nine investment age categories (0–8), where age 0 in a given year represents plants that had an investment spike in the previous period and age 8 represents plants that have not experienced an investment spike in the previous eight years. Figure 9 plots the empirical hazards for a selected number of years over this sample.

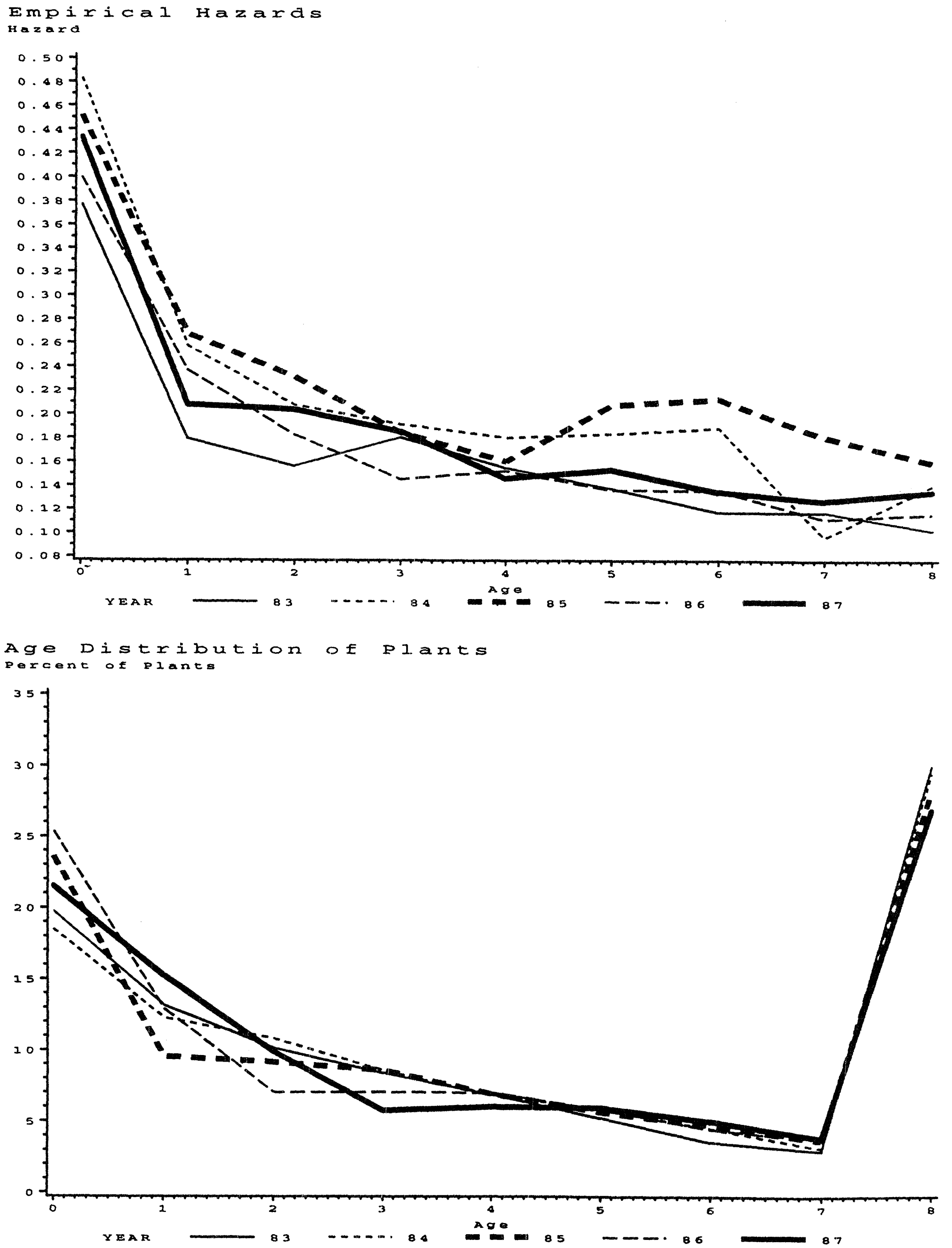

FIGURE 9. EMPIRICAL HAZARDS AND CROSS-SECTIONAL DISTRIBUTIONS

sented below provide further support for this interpretation.

On face value, the downward-sloping empirical hazard may appear more supportive of a convex cost of adjustment model than a lumpy investment model. However, these empirical hazards reflect a variety of forces beyond those we emphasized in the theoretical analysis. In particular, as is well known in the empirical labor economics literature, unobserved structural heterogeneity can yield downward-sloping hazards even if the hazard for any individual plant is upward sloping.

A similar effect is possible in the context of investment spikes. For example, suppose that there are three types of plants, where a type i plant undergoes machine replacement every T_i periods. Let π_i denote the fraction of type i plants in the population. If plant types are observable, then conditional on a type, the hazard function is clearly upward sloping. However, if types are not observable, then the probability of machine replacement at age 1 is π_1, at age 2 it is $\pi_2/(\pi_2 + \pi_3)$, and at age 3 and beyond, it is 1. Thus the hazard for all the plants together will be lower at age 2 than at age 1 iff $\pi_2 < \pi_1(1 - \pi_1)$. The decreasing part of the hazard is simply due to a composition effect: the fraction of type 2 plants is small relative to those at risk in the group of age 2 plants. Given the theoretical model, such structural heterogeneity could arise from fundamental quantitative differences in the productivity shock distribution, the pace of technological progress, and/or the fixed adjustment costs across producers.

Put differently, the empirical hazards depicted in Figure 9 reflect both between-plant and within-plant effects. In contrast, the prediction of an upward-sloping hazard is solely a within-plant prediction. This is a point we return to later. An attractive feature of the decomposition exercise considered in this section is that it does not depend on separating out the influences of heterogeneity and duration dependence. That is, we can characterize the aggregate implications of variations in the cross-sectional age distribution without separating out the influences of heterogeneity and duration dependence on the micro hazards.

The bottom panel of Figure 9 displays the age distribution of plants for 1983–1987. Age 0 plants are those that experienced a spike in the previous year. The low investment rate in 1983 (from Figure 8) implies that there are relatively few age 0 plants in 1984 and thus relatively few age 1 plants in 1985, few age 2 plants in 1986, and so on. Overall, there is substantial variation in the age distribution of plant capital. The variations in the age distribution depicted in this figure will have an effect on investment which is highlighted by the decomposition in (11).

The time series of I_t^k and I_t^g are plotted against the actual fraction of plants experiencing investment spikes in Figure 10.[29] I_t^g tracks aggregate investment closely implying that time-series variation in the cross-sectional age distribution has only a modest marginal contribution—or put conversely, fluctuations in the hazard are what matter most. In terms of measures of goodness of fit, the root mean-squared error associated with I_t^g (i.e., the difference between I_t and I_t^g) is 0.004. These are small forecast errors relative to the variation exhibited by aggregate investment (standard error of the latter is 0.021).

The flip side of these results is that the series I_t^k exhibits only modest movements over time. However, close examination makes it clear that during periods of the greatest movement in I_t^k there are some interesting and important differences between I_t and I_t^g. Specifically, periods in which I_t^k is relatively low (high) are precisely those periods that I_t^g overpredicts (underpredicts) the actual increase in I_t.

Consider, for example, the years 1983–1987. From Figure 9, the empirical hazard shifts down substantially in 1983 and then shifts up substantially in 1984 and 1985. The lower panel of Figure 9 indicates that these shifts yielded nontrivial fluctuations in the cross-sectional age distribution. Accordingly, we observe that ignoring the fluctuations in the cross-sectional distribution (as in I_t^g) yields a pattern of overprediction and underprediction for the next several years. This pattern of overprediction and underprediction is precisely what one would expect if the machine replacement model was correct but the implied influence of the cross-sectional age distribution was ignored.

[29] In generating the series for this figure, we standardized all of the series to have the mean of the actual investment series. That is, we eliminated any mean differences between the series.

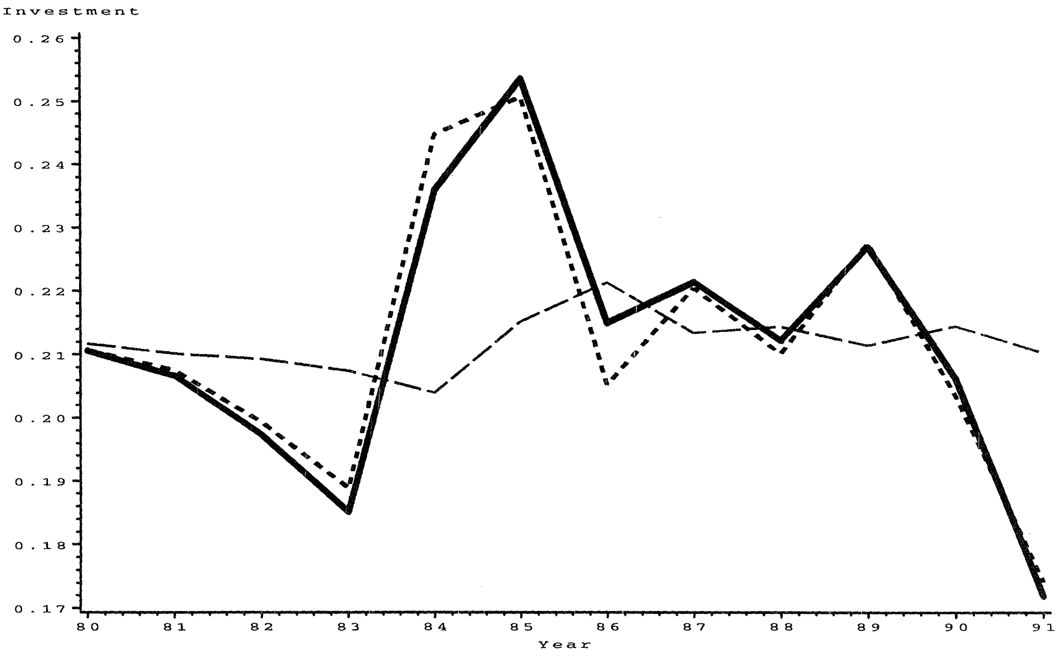

FIGURE 10. ACTUAL AND PREDICTED INVESTMENT SPIKES

Notes: The solid line is actual investment. The short dashed line is I_t^g—investment holding the cross-sectional distribution constant. The long dashed line is I_t^k—investment holding the hazard fixed.

Recall from the time-series simulation that the influence of the cross-sectional age distribution is twofold. First, the impact of aggregate shocks depends on the state of the cross-sectional age distribution. Second, endogenous cycles in investment emerge when the existing cross-sectional age distribution differs sharply from that associated with the steady-state distribution consistent with current aggregate conditions. Both of these effects are especially important in the periods immediately following large fluctuations in aggregate investment. Further, while both of these effects lead to prediction errors, it is this latter effect that generates a pattern of overprediction and underprediction when one ignores the influence of the cross-sectional age distribution and the nonflat hazard.

While we do find that fluctuations in the cross-sectional age distribution matter, it is nevertheless the case that much of the variation is coming from shifts in the hazard over time. At first glance, our finding that so much of investment is explained by shifts in the hazard seems at variance with the results reported in Caballero et al. (1995) (hereafter, CEH). Two key differences in approach are important in comparing and contrasting the results across these studies. For one, we focus on investment spikes while CEH do not restrict their attention to investment spikes. Second, the state spaces for the two models are quite different. In particular, in CEH, the state variable is the difference between actual and desired capital stocks and thus the cross-sectional distribution in CEH is defined in terms of this difference. This implies that aggregate shocks will shift the cross-sectional distribution in the CEH framework. Put in terms of our model, shifts in the cross-sectional distribution in CEH will reflect shifts in the distribution of k, A *and* the distribution of ε. In contrast, our model isolates the contribution of changes in the cross-sectional distribution of capital vintages, as measured by investment age.

One aspect of the results in CEH that closely corresponds to those reported here is that they find that taking into account fluctuations in the cross-sectional distribution (defined in their state space) is especially important during times of large swings of investment. This latter finding is close in spirit to the results reported here.

C. *Estimating Microeconomic Investment Spike Hazards*

The goal of this final section is to provide evidence on the hazard function at the plant level. While appropriate for our analysis of aggregate investment dynamics, the empirical hazards depicted in Figure 9 are inadequate for evaluating the prediction of our model that replacement is more likely for older machines. As noted before and discussed at considerable length in the voluminous literature on duration dependence, the fact that the empirical hazards may reflect unobservable heterogeneity will generally bias the hazard functions downward. Accordingly, we proceed with an estimation strategy that permits us to control for unobservable heterogeneity.

We use a semiparametric specification developed by Bruce D. Meyer (1990) to estimate the hazard from the distribution of durations between spikes.[30] The Meyer approach is well suited for our setting for a number of reasons. The estimation methodology yields a flexible semiparametric characterization of the hazard estimated from the distribution of durations defined over a fixed number of discrete intervals. In our setting with annual data, the durations fall naturally into a limited number of discrete intervals. Another appealing feature of the Meyer approach is that it is relatively straightforward to allow for unobserved heterogeneity in this framework.

Following Meyer's (1990) derivation, let T_i be the length of the spell for plant i. Then the hazard for plant i at time t, is defined by the equation:

$$(12)\qquad \lim_{\Delta\to 0}\frac{\text{prob}[t+\Delta > T_i \geq t | T_i \geq t]}{\Delta} = h_i(t).$$

The hazard is parameterized here using a proportional hazards form, i.e.,

$$(13)\qquad h_i(t) = h_0(t)\exp(z_i(t)'\beta)$$

where $h_0(t)$ is the baseline hazard at time t, which is unknown, z is a vector of covariates, and β is vector of unknown parameters.

The probability that a spell lasts until $t + 1$ given that it has lasted until t is given by:

$$(14)\qquad P[T_i \geq t+1 | T_i \geq t] = \exp[-\exp(z_i(t)'\beta + \gamma(t))]$$

where

$$(15)\qquad \gamma(t) = \ln\left\{\int_t^{t+1} h_0(u)\, du\right\}.$$

The likelihood function for a sample of N individuals can be written in terms of (14) as:

$$(16)\qquad L^1(\gamma, \beta) = \prod_{i=1}^{N} L_i(k_i, d_i) = \prod_{i=1}^{N}\left[[1 - \exp\{-\exp[\gamma(k_i) + z_i(k_i)'\beta]\}]^{d_i} \times \prod_{t=1}^{k_i-1}\exp\{-\exp[\gamma(t) + z_i(t)'\beta]\}\right]$$

where C_i is the censoring time and $d_i = 1$ if $T_i \leq C_i$ and 0 otherwise, and $k_i = \min(\text{int}(T_i), C_i)$. The first term in (16) represents the probability of exit in the interval $[k_i, k_i + 1]$ given that the spell has lasted until k_i and thus represents the discrete interval hazard rate. The second term in (16) represents the probability of observing a spell that lasts at least until k_i.

Estimation of this specification for a given choice of discrete intervals yields a nonparametric estimate of the baseline hazard but does not control for unobserved heterogeneity. Introducing unobserved heterogeneity in this specification can be done in a variety of ways. Here we focus on the approach suggested by Meyer (1990) and Peter J. Dolton and Wilbert van der Klaauw (1995) (based on an extension of James

[30] The approach of Meyer is essentially the same as that of Aaron Han and Jerry A. Hausman (1990).

Heckman and Burton Singer [1984]) in which the unknown distribution of heterogeneity is approximated by a discrete distribution with a finite number of mass points. The points of support and corresponding probabilities for the latter are estimated jointly with the baseline hazard and other parameters. In terms of the above, the likelihood function can then be written as the product of weighted sums of terms similar to $L_i(k_i, d_i)$ in (16):

$$(17)\qquad L^2(\gamma, \beta, \alpha, \mu, J) = \prod_{i=1}^{N} \sum_{j=1}^{J} \alpha_j L_i(k_i, d_i | \mu_j)$$

where μ_j, $j = 1, \ldots, J$ are the J points of support with probabilities α_j and $L_i(k_i, d_i|\mu_j)$ corresponds to the expression in (16) with terms like $\exp(\gamma(t) + z(t)'\beta)$ replaced with $\exp(\gamma(t) + z(t)'\beta + \mu_j)$. In addition to the baseline hazard, maximization of the log-likelihood in (17) yields estimates of the points of support (where μ_0 is normalized to zero) and $J - 1$ weights (the sum of the α's is constrained to one).

Table 1 reports the results from estimating (16) and (17) with our data. To implement (16) and (17) with our data, we selected all duration spells that commenced between 1972 and 1979. That is, for plants that experienced an investment spike in the 1972 to 1979 period, we tracked their investment age until their next investment spike and, in so doing, measured the duration of their spells. Spells that are ongoing in 1991 and/or are greater than or equal to an investment age of 11 are censored. The 1972–1979 cohorts yielded 9,309 spells that serve as the basis of our estimation.

We are primarily interested in the baseline hazards and only considered a minimal set of covariates. The controls we include are cohort year effects that are dummies for the year the spell began. In this fashion, we are able to control for common effects experienced by plants in the same cohort. In the table, the cohort year effects are denoted by $C(YR)$. Two columns are reported in Table 1. The first column reflects the implementation of (16) without any unobserved heterogeneity. The second column reflects the implementation of (17) with the discrete distribution specification for heterogeneity. We focus on the results for $J = 3$ although we considered $J = 2$ and $J = 4$ as well [and obviously $J = 1$ since this is (16)]. The estimation yielded increases in the likelihood until $J = 4$ (log

TABLE 1—SEMIPARAMETRIC HAZARD MODEL ESTIMATES

Variable	No heterogeneity	Discrete heterogeneity
$C73$	−0.02	0.02
	(0.04)	(0.07)
$C74$	−0.11	−0.40
	(0.04)	(0.07)
$C75$	−0.11	−0.25
	(0.04)	(0.08)
$C76$	−0.06	−0.08
	(0.04)	(0.07)
$C77$	0.02	0.02
	(0.04)	(0.07)
$C78$	0.01	0.06
	(0.04)	(0.07)
$\gamma 0$	−0.67	−4.14
	(0.03)	(0.52)
$\gamma 1$	−1.43	−4.47
	(0.04)	(0.52)
$\gamma 2$	−1.53	−4.11
	(0.04)	(0.51)
$\gamma 3$	−1.56	−3.43
	(0.05)	(0.44)
$\gamma 4$	−1.68	−2.91
	(0.05)	(0.35)
$\gamma 5$	−1.63	−2.57
	(0.05)	(0.32)
$\gamma 6$	−1.73	−2.46
	(0.06)	(0.29)
$\gamma 7$	−1.96	−2.50
	(0.07)	(0.27)
$\gamma 8$	−2.00	−2.38
	(0.08)	(0.25)
$\gamma 9$	−1.89	−2.06
	(0.06)	(0.20)
$\mu 2$		1.94
		(0.38)
$\mu 3$		4.17
		(0.49)
$\alpha 1$		0.21
		(0.03)
$\alpha 2$		0.20
		(0.04)
Log-likelihood	−18642	−18619

Notes: Number of spells: 9,309. Standard errors are in parentheses. Estimated coefficients in the no heterogeneity column are for equation (16). Estimated coefficients in the discrete heterogeneity column are for equation (17).

likelihood almost identical to that for $J = 3$) and thus we settled on $J = 3$.[31]

Our primary focus is on the shape of the baseline hazard. The top panel of Figure 11 presents the estimated hazard rate (i.e., the probability of having an investment spike at various investment ages conditional on not having a spike until that age) for the no-heterogeneity case. Like the empirical hazards reported in Figure 9, the hazard is downward sloping—driven primarily by the large probability of having an investment spike in the year immediately succeeding an investment spike. The hazard is relatively flat after investment age 1.

The picture changes dramatically when we control for unobserved heterogeneity. Allowing for $J = 3$ yields a significant increase in the log-likelihood with each of the groups having a significant weight and quite different hazard patterns. The lower panel of Figure 11 reports the implied hazards for the three groups associated with $J = 3$. As seen from Table 1, group 1 has a weight of 0.21 and has a very low average hazard rate that rises only slightly as investment age rises. Group 2 has a weight of 0.20 with a higher average hazard rate and a hazard rate that falls slightly from investment age 0 to investment age 1 and then rises steadily with investment age. Group 3 has a weight of 0.59, a much higher average hazard rate, a pattern of a decrease in the hazard from investment age 0 to investment age 1, and then a steady increase in the hazard with investment age.

In short, controlling for unobserved heterogeneity reveals patterns that are broadly consistent with the machine replacement model of Section II. Regardless of methodology, we observe a pattern of a high probability of having a spike in the year immediately following a spike. However, abstracting from this multiyear spike effect and controlling for unobserved heterogeneity, the main result is easily summarized: *the probability of having an investment spike is increasing in the time since the prior spike.*

IV. Conclusion

Our goal in this paper was to model the machine replacement problem of an individual producer and then to trace the implications of this discrete-choice problem for the behavior of aggregate investment. To do so, we analyzed the choice of an individual agent with emphasis on the timing of replacement relative to the business cycle. Here we found that replacement investment is more likely to be procyclical the more persistent are shocks and the more important are fixed adjustment costs. Further, the theoretical model predicted fairly rich dynamics in aggregate investment from the interaction of aggregate shocks and the evolution of the cross-sectional distribution of capital vintages towards its stationary distribution.

The empirical section of the paper provides evidence that these theoretical properties are apparent in the data. First, large investment episodes are an important feature of plant-level investment and constitute a large fraction of aggregate fluctuations in investment. Second, estimates of microeconomic hazards reveal that the probability of a large investment episode (denoted an investment spike) is increasing in the time since the previous spike. Third, investment spikes are procyclical. Putting these pieces together implies that ignoring fluctuations in the cross-sectional distribution of the vintage of the capital stock can lead to nontrivial errors in the predicted changes in the fraction of plants experiencing investment spikes. Strikingly, the relative importance of the prediction errors increases during periods of large investment fluctuations.

There are at least three directions in which to build upon our work. First, the model excludes any movements on the intensive margin by assuming that all investment projects are of the same size. While analytically convenient, the empirical implementation of this approach requires the creation of a discrete variable from observations on investment flows. An alternative approach, which we will pursue, is to build a richer model of investment which allows the agent to choose a path of investment with nonconvex costs of ad-

[31] The method of selecting J by increasing it until the likelihood fails to increase is advocated by Dolton and van der Klaauw (1995). The shape of the hazards is quite similar for $J = 2$, $J = 3$, and $J = 4$. That is, only permitting two groups yields a tendency for upward-sloping hazards. Note that we also considered a gamma distribution specification of the unobserved heterogeneity. This specification also yielded an upward-sloping baseline hazard. Note that the log-likelihood for the gamma heterogeneity case is −18632, which reflects an increase relative to the no-heterogeneity case but is lower than the $J = 3$ case.

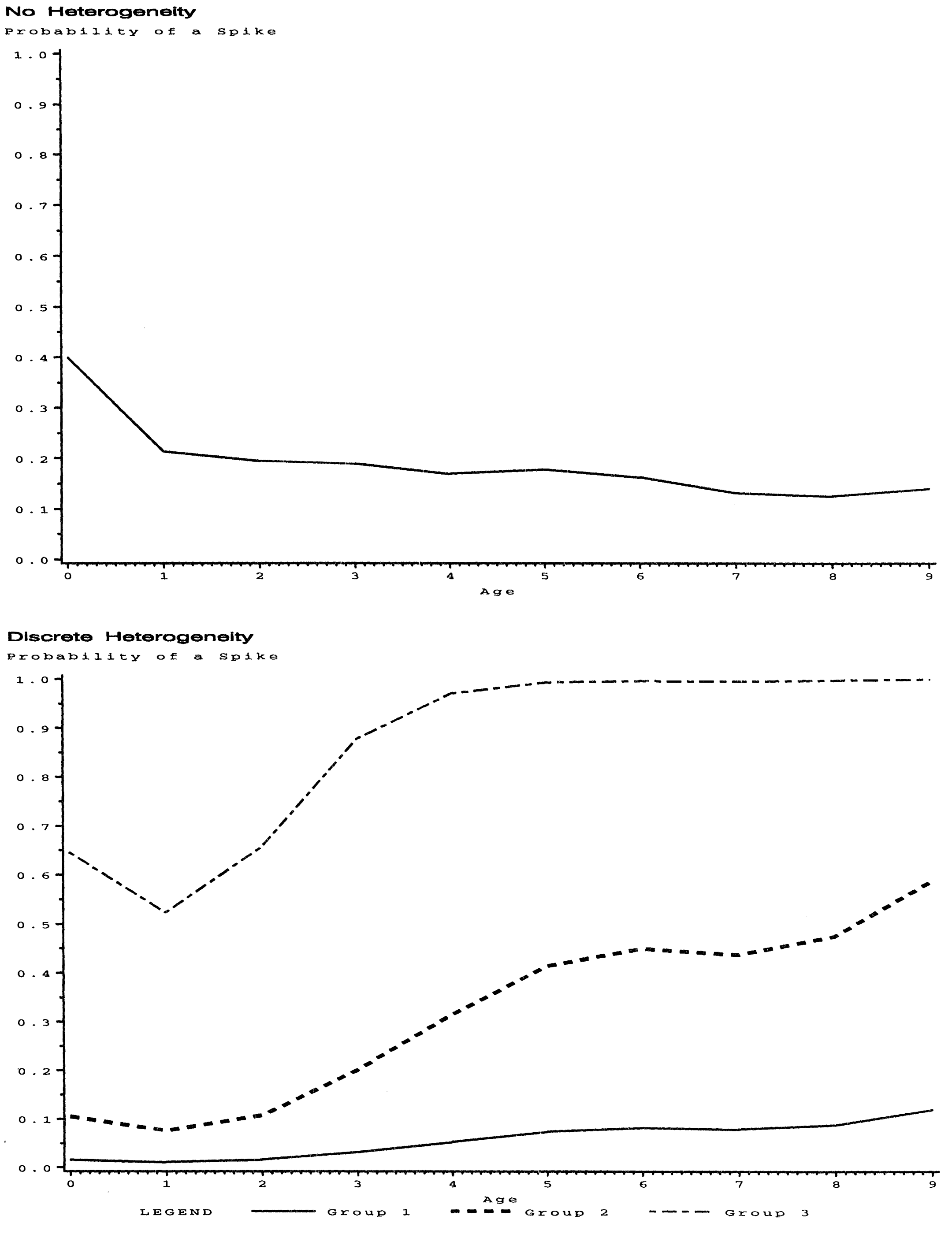

FIGURE 11. SEMIPARAMETRIC HAZARDS

justment. This may also include a description of the effects of capital-market imperfections. This model will be easier to implement empirically and will nest the standard neoclassical model of investment as a special case.

Second, working from the fundamental contribution of Rust (1987) and others, there is an alternative empirical strategy to pursue: the estimation of the structural model from the LRD. While our empirical work did not require the imposition of a particular hazard function, the approach does not provide a link between the estimates and the underlying structural model. In future work, we will make this link explicit and provide microeconomic estimates of an investment model with nonconvex costs of adjustment.[32] This seems particularly fruitful for the evaluation of policies such as investment tax credits.

The final issue concerns the links between capital, worker flows, and productivity. This paper concentrates on the timing of large investment episodes while others, e.g., Steve J. Davis and Haltiwanger (1992), focus on employment movements. An integrated model of investment activity and labor-market dynamics with implications for productivity is an obvious next step.

Appendix

PROOF OF PROPOSITION 1:

As long as μ is small enough that $\beta < 1$, then the existence of a solution to (7) is guaranteed by Theorem 9.6 in Nancy L. Stokey and Robert E. Lucas, Jr. (1989).

PROOF OF PROPOSITION 2:

Let A denote the productivity at the plant ignoring, for now, the decomposition of A. For a given value of A, let $k^*(A)$ satisfy $V^n(k, A) = V^r(k, A)$ where

$$\text{(A1)} \quad V^n(k, A) \equiv Ak + \beta EV(\rho k, A')$$

and

$$\text{(A2)} \quad V^r(k, A) \equiv Ak\lambda - F + \beta EV(1, A').$$

Letting $\Delta(k, A) \equiv V^r(k, A) - V^n(k, A)$, it is sufficient to show that $\Delta(k, A)$ is decreasing in k.

From (A1) and (A2),

$$\Delta(k, A) = Ak(\lambda - 1) - F + \beta E_{A'}[V(1, A') - V(\rho k, A')]$$

where $V(k, A) \equiv \max\{V^r(k, A), V^n(k, A)\}$. The first term in this expression is decreasing in k. Further, the last part of this expression, representing the expected gains to replacement, is also decreasing as k increases since $V(k, A)$ is an increasing function of k. Hence, $\Delta(k, A)$ is decreasing in k.

This proves that given the aggregate state A, the hazard is decreasing in k. Since it is true for each realized value of A, it is also true if A represents the product of an aggregate and an idiosyncratic shock and one integrates over the idiosyncratic components of the shock to obtain the hazard defined over the aggregate state.

PROOF OF PROPOSITION 3:

Using the definition of $\Delta(k, A)$, for the case of $F > 0$ and $\lambda = 1$, we have

$$\Delta(k, A) = -F + \beta E_{A'}[V(1, A') - V(\rho k, A')].$$

Since A is i.i.d., the right side is independent of the current realization of the shock. Hence the gains to replacement are independent of A.

PROOF OF PROPOSITION 4:

Again, using $\Delta(k, A)$ as defined above, we have:

$$\Delta(k, A) \equiv Ak(\lambda - 1) + \beta[EV(1, A') - EV(\rho k, A')]$$

where the expectation is taken over A'. Since A

[32] In our preliminary exploration of these issues we are, in fact, estimating a model that includes both convex and nonconvex aspects of the adjustment process.

is assumed to be i.i.d., this expectation is independent of A.

To show that replacement is less likely when A is high, we need to show that $\Delta(k, A)$ is a decreasing function of A. With A i.i.d., this amounts to showing that $Ak(\lambda - 1)$ is a decreasing function of A; i.e., the costs of replacement rise in good times due to the opportunity cost aspect of the replacement process. This is clearly the case as $\lambda < 1$.

PROOF OF PROPOSITION 5:

For the case of $F > 0$, $\lambda = 1$, $\Delta(k, A)$ is given by

$$\Delta(k, A) = \beta E_{A'|A}[V(1, A') - V(\rho k, A')] - F.$$

Note that the expectation over A' is conditional on A so that the current state of productivity does influence the replacement choice even though $\lambda = 1$. Since high values of A put, by supposition, more weight on high values of A', it is sufficient to show that $V(1, A) - V(k, A)$ is increasing in A for any k. This is, in turn, equivalent to the condition that

$$\int_k^1 V_{kA}(z, A)\, dz > 0$$

for all k. This condition is satisfied if $V_{kA}(k, A) > 0$ for all (k, A). From (A1) and (A2) this positive cross-partial condition holds when $F > 0$ and $\lambda = 1$. To see this, note that by assumption, replacement will eventually occur so that (A1) is a sequence of current period returns with positive cross partials between k and A. From (A2), $V^r(k, A)$ has a positive cross partial since the second term is independent of k.

REFERENCES

Abel, Andrew B. "Consumption and Investment," in Benjamin M. Friedman and Frank H. Hahn, eds., *Handbook of monetary economics*, Vol. 2. New York: Elsevier Science, 1990, pp. 725–78.

Abel, Andrew B. and Eberly, Janice C. "A Unified Model of Investment Under Uncertainty." *American Economic Review*, December 1994, *84* (5), pp. 1369–84.

______. "Investment and q with Fixed Costs: An Empirical Analysis." Mimeo, University of Pennsylvania, 1996.

Bertola, Giuseppe and Caballero, Ricardo J. "Kinked Adjustment Costs and Aggregate Dynamics," in Olivier Jean Blanchard and Stanley Fischer, eds., *NBER macroeconomics annual 1990*. Cambridge, MA: MIT Press, 1990, pp. 237–87.

Caballero, Ricardo J. and Engel, Eduardo M. R. A. "Microeconomic Adjustment Hazards and Aggregate Dynamics." *Quarterly Journal of Economics*, May 1993, *108* (2), pp. 359–84.

______. "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S, s) Approach." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 4887, October 1994.

Caballero, Ricardo J.; Engel, Eduardo M. R. A. and Haltiwanger, John. "Plant-Level Adjustment and Aggregate Investment Dynamics." *Brookings Papers on Economic Activity*, 1995, (2), pp. 1–39.

______. "Aggregate Employment Dynamics: Building from Microeconomic Evidence." *American Economic Review*, March 1997, *87* (1), pp. 115–37.

Cooley, Thomas; Greenwood, Jeremy and Yorukglu, Mehmet. "The Replacement Problem." *Journal of Monetary Economics*, December 1997, *40* (3), pp. 457–99.

Cooper, Russell and Haltiwanger, John. "Macroeconomic Implications of Production Bunching: Factor Demand Linkages." *Journal of Monetary Economics*, October 1992, *30* (1), pp. 107–28.

______. "The Macroeconomic Implications of Machine Replacement: Theory and Evidence." *American Economic Review*, June 1993, *83* (3), pp. 360–82.

Cooper, Russell; Haltiwanger, John and Power, Laura. "Machine Replacement and the Business Cycle: Lumps and Bumps." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5260, September 1995.

Davis, Steve J. and Haltiwanger, John C. "Gross

Job Creation, Gross Job Destruction, and Employment Reallocation." *Quarterly Journal of Economics*, August 1992, *107* (3), pp. 819–63.

Dolton, Peter J. and Wilbert van der Klaauw. "Leaving Teaching in the U.K.: A Duration Analysis." *Economic Journal*, March 1995, *105* (429), pp. 431–44.

Doms, Mark and Dunne, Timothy. "Capital Adjustment Patterns in Manufacturing Plants." U.S. Bureau of the Census, Center for Economic Studies Discussion Paper No. 94/11, August 1994.

Han, Aaron and Hausman, Jerry A. "Flexible Parametric Estimation of Duration and Competing Risk Models." *Journal of Applied Econometrics*, January–March 1990, *5* (1), pp. 1–28.

Heckman, James and Singer, Burton. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*, March 1984, *52* (2), pp. 271–320.

King, Robert G.; Plosser, Charles I. and Rebelo, Sergio T. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics*, March/May 1988, *21* (2/3), pp. 195–232.

Matthews, R. C. O. *The business cycle*. Chicago: University of Chicago Press, 1959.

McClelland, John. "Permit Renewal and the Timing of Major Investment." Mimeo, University of Maryland, July 1997.

Meyer, Bruce D. "Unemployment Insurance and Unemployment Spells." *Econometrica*, July 1990, *58* (4), pp. 757–82.

Nilsen, Øivind and Schiantarelli, Fabio. "Zeroes and Lumps in Investment: Empirical Evidence on Irreversibilities and Non-convexities." Mimeo, Boston College, March 1998.

Power, Laura. "Causes and Consequences of Investment Spikes in U.S. Manufacturing Plants." Ph.D. dissertation, University of Maryland, 1994.

______. "The Missing Link: Technology, Investment, and Productivity." *Review of Economics and Statistics*, May 1998, *80* (2), pp. 300–13.

Rothschild, Michael. "On the Cost of Adjustment." *Quarterly Journal of Economics*, November 1971, *85* (4), pp. 605–22.

Rust, John. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, September 1987, *55* (5), pp. 999–1034.

Stokey, Nancy L. and Lucas, Robert E., Jr. *Recursive methods in economic dynamics*. Cambridge, MA: Harvard University Press, 1989.