Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S,s) Approach

Ricardo J. Caballero; Eduardo M. R. A. Engel

Econometrica, Vol. 67, No. 4 (Jul., 1999), 783-826.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199907%2967%3A4%3C783%3AEIDIUM%3E2.0.CO%3B2-S

Econometrica is currently published by The Econometric Society.

Econometrica, Vol. 67, No. 4 (July, 1999), 783–826

EXPLAINING INVESTMENT DYNAMICS IN U.S. MANUFACTURING: A GENERALIZED (S,s) APPROACH

BY RICARDO J. CABALLERO AND EDUARDO M. R. A. ENGEL[1]

In this paper we derive a model of aggregate investment that builds from the lumpy microeconomic behavior of firms facing stochastic fixed adjustment costs. Instead of the standard sharp (S,s) bands, firms' adjustment policies take the form of a probability of adjustment (*adjustment hazard*) that responds smoothly to changes in firms' capacity gap. The model has appealing aggregation properties, and yields nonlinear aggregate time series processes. The passivity of normal times is, occasionally, more than offset by the brisk response to large accumulated shocks. Using within and out-of-sample criteria, we find that the model performs substantially better than the standard linear models of investment for postwar sectoral U.S. manufacturing equipment and structures investment data.

KEYWORDS: Investment, adjustment costs, adjustment hazard, aggregation, heterogeneity, lumpiness, nonlinear time series.

1. INTRODUCTION

Minor upgrades and repairs aside, investment projects at the plant level are intermittent and lumpy rather than smooth. This is starkly documented in Doms and Dunne (1993). They use the Longitudinal Research Datafile to study the investment behavior of 12,000 continuing (and large) U.S. manufacturing establishments for the seventeen year period from 1972–1988, and find that: (i) more than half of the establishments exhibit capital growth close to 50 percent in a single year, and (ii) over 25 percent, and perhaps as much as 40 percent, of an average plant's gross investment over the seventeen year period is concentrated in a single year/project.[2,3]

Since this basic feature of microeconomic data is seldom considered in empirical investment equations, it perhaps should come as no surprise that success in estimating and testing investment equations is so rare.[4] At a broad level, our goal in this paper is to develop and test a framework to study the dynamic behavior of aggregate investment, subject to the constraint that it

[1] We are grateful to Olivier Blanchard, Whitney Newey, James Stock, an editor, three anonymous referees, and seminar participants at Brown, CEPR-Champoussin, Chicago, Columbia, Econometric Society Meetings (Caracas and Tokyo), EFCC, Harvard, IMPA, LSE, NBER, Princeton, Rochester, SITE, Toronto, U. de Chile, and Yale for their comments. Financial support (to Caballero) from the National Science and Sloan Foundations and (to Engel) from FONDECYT (Grant 195-510) and the Mellon Foundation (Grant 9608) is gratefully acknowledged.

[2] Since plants' entry is excluded from their sample, these statistics are likely to represent lower bounds on the degree of lumpiness in plants' investment patterns.

[3] We use the word "project" to emphasize the fact that the actual implementation of a project may cover more than a year-observation; realistic time-to-build aspects of investment are not in contradiction with the view that investment episodes are lumpy in nature.

[4] See Chirinko (1994) for a survey of the empirical investment literature.

builds up from microeconomic units generating the lumpy and intermittent pattern observed in microeconomic data.

Achieving such a goal requires three methodological ingredients: (i) a microeconomic model of lumpy adjustment; (ii) an aggregation procedure; and (iii) an estimation and testing method that is not only consistent with (i) and (ii), but also able to highlight the impact of the proposed microeconomic model on aggregate dynamics.

As in the standard (S,s) literature, our microeconomic model generates lumpy behavior through the presence of a fixed cost of adjusting the firm's capital stock. Unlike this literature, the fixed cost is random so the "inaction range" is no longer fixed over time (and across firms). The optimal policy still takes a simple form under standard assumptions about the stochastic process of exogenous variables: let z denote the log-difference between a firm's actual and frictionless (i.e., in the absence of adjustment cost) stock of capital, and let ω be a random variable indexing the adjustment cost faced by the firm at some point in time, with distribution $G(\omega)$. The solution to the firm's problem yields a function $\Omega(z)$ that represents the maximum realization of ω for which a firm with imbalance z chooses not to adjust. For any smaller ω, firms adjust fully. Removing conditioning on ω, on the other hand, yields an *adjustment hazard* function $\Lambda(z)$ that describes the *probability* that a firm with imbalance z adjusts. Since it varies smoothly with z, this probabilistic (S,s)-type rule is more amenable to aggregation than the standard fixed-bands (S,s) model and, more importantly, has the virtue of nesting a wide variety of models. At the extremes, when $G(\omega)$ degenerates into a spike we recover the (S,s) model, while when it becomes a distribution with plenty of mass at very low values of ω and the remaining mass at very high adjustment costs, we approximately recover a model with linear aggregate dynamics (the standard partial adjustment model).

Firms' actions are not perfectly synchronized. On one hand, at any point in time adjustment costs differ across firms. On the other, differences in initial conditions, idiosyncratic shocks, and previous actions, yield a nondegenerate cross-sectional density of capital imbalances, $f(z,t)$, at all times. *Aggregation* proceeds in two steps, both under the assumption of a large number of firms: First, within each z, the microeconomic adjustment hazard now represents the fraction of units with that imbalance that choose to adjust at any given moment in time. Second, to obtain aggregate investment we integrate these adjustments across z, using as measure the current cross-sectional density. In order to describe the dynamic path of aggregate investment we characterize the path of $f(z,t)$ which, under our assumptions, is Markovian with a transition operator that depends on the realization of aggregate shocks.

We make (fairly flexible) distributional assumptions about aggregate shocks and estimate the model by Maximum Likelihood using (aggregate) two-digit U.S. manufacturing investment/capital ratios for the period 1948–1992. We find clear evidence in favor of our generalized (S,s) model, both in terms of within sample criteria and out-of-sample predictive power. Our structural interpretation of these nonlinearities indicates that fixed adjustment costs faced by firms

are large. Although important for both, these features are more pronounced for structures than equipment. When compared with standard linear models, the forecasting accuracy of the model we postulate is substantially improved.

One of the main mechanisms by which aggregate dynamics generated by the (S, s) type model differ from their linear counterparts, is that the number of active firms changes over the cycle—a point emphasized by Bar-Ilan and Blinder (1992). Doms and Dunne (1993) confirm the importance of this mechanism by showing that the number of plants going through their primary investment spikes, rather than the average size of these spikes, tracks closely aggregate manufacturing investment over time. Consistently, and depending on the specific sequence of preceding events, the nonlinear model we estimate has the potential to generate brisker expansions than its linear counterparts. It is also this feature that largely explains its enhanced forecasting properties.

Beyond the empirical findings on investment and its integrative nature, this paper has two specific methodological contributions to the new literature on nonconvex adjustment costs and lumpy actions.

On the microeconomic side, there have been several developments on models of lumpy and intermittent adjustment (the (S, s) literature).[5] As we discussed above, here we extend these models so the adjustment trigger barriers vary randomly across firms and for a firm over time. This modification is a first step toward introducing the realistic and empirically important feature that units do not always wait for the same stock disequilibrium to adjust, and that adjustments are not always of the same size across firms and for the same firm over time, while preserving a fairly parsimonious aggregation setup.

More recently, there have also been developments of empirical models of aggregate dynamics with heterogeneous microeconomic units adjusting intermittently.[6] Econometric implementation of these models, however, has required observing (or estimating separately in an often debatable first stage) a measure of the exogenous component of the aggregate driving force. Our nonlinear time series procedure does not require the first stage; it only requires information on the aggregate investment series itself and on the generating process of the driving force (but not its realization). Somewhat analogously with the standard

[5] See Harrison, Selke, and Taylor (1983) for a technical discussion of impulse control problems. For a good survey of the economics literature—although with an emphasis on models where investment is infrequent but not lumpy—see Dixit and Pindyck (1994). More closely related to a special case of ours is Grossman and Laroque's (1990) model of consumer durable purchases.

[6] Blinder (1981), Bar-Ilan and Blinder (1992), and Lam (1991) look at data on inventories (the first one) and consumer durables (the other two) under the organizing principles of (S, s) models. Bertola and Caballero (1990) and Caballero (1993) provide a structural empirical framework and estimate (S, s) models for consumer durable goods. Bertola and Caballero (1994) implement empirically an irreversible investment model where microeconomic investment is intermittent but not lumpy. Caballero and Engel (1992a, 1993a, 1993b) estimate aggregate models of employment and price adjustments when microeconomic units follow more general (probabilistic) microeconomic adjustment rules but, contrary to the current paper, they do not derive these rules from a microeconomic optimization problem.

procedure of estimating convex adjustment cost parameters from the first (or higher) order serial correlation of investment, we learn about more complex lumpy adjustment cost functions from the structure of aggregate investment lags and their changes over time.

The next section presents the basic model. It is followed by Section 3, which describes the econometric method and presents our main empirical results. Conclusions and extensions are discussed in Section 4. Several technical appendices follow.

2. THE BASIC MODEL

2.1. *Overview*

We model a sector composed of a large but fixed number of monopolistically competitive firms. Each firm faces an isoelastic demand for its differentiated product, which is produced with a Cobb-Douglas constant returns technology in labor and capital. Both demand and technology are affected by multiplicative shocks described by a joint geometric random walk. These shocks have firm specific and sectoral (aggregate) components that we specify later. We work in discrete time.

The sector faces infinitely elastic supplies of labor and capital. We choose the price of the latter as numeraire and let the wage (relative to the price of capital) follow a geometric random walk process, possibly correlated with demand and technology shocks. Firms can adjust their labor input at will but suffer a loss when resizing their stock of capital. Since our aim is to capture firms' infrequent and lumpy investment, we assume this loss takes the form of a fixed cost, which can be interpreted either as an index of the degree of specificity of firms' capital, or as a secondary market imperfection if machines or structures are replaced, or as a reorganization cost associated with putting new capital to work. In order to capture some of the time series and cross-sectional heterogeneity in these fixed costs, we let the extent of the loss due to adjustment vary randomly over time as firms may, for example, find better or worse matches or uses for their old machines, or may face reorganizations of different degrees of difficulty.

As in standard (S, s) models, the resulting microeconomic policy is one of inaction interspersed with periods of large investment or disinvestment. As in standard search models, at each point in time the firm decides whether to "accept" the currently offered fixed adjustment cost or to postpone adjustment and draw a new adjustment cost next period. The interaction between these two mechanisms implies that, more realistically than in standard (S, s) models, the size of adjustments varies both across firms and over time for the same firm. During a given time period, firms with identical shortages or excesses of capital act differently. Over time, the same firm reacts differently to similar disequilibria in its stock of capital.

Intuitively, the largest adjustment cost a firm is prepared to tolerate without adjusting its capital stock decreases with the extent of its capital stock imbalance. If the distribution of adjustment costs is nondegenerate, this implies that the probability that a firm adjusts for a given disequilibrium—a concept we describe as the firm's *adjustment hazard*—increases smoothly and monotonically with the firm's disequilibrium in its stock of capital.[7]

Since we assume the number of firms is large and adjustment costs are independent across firms, the *adjustment hazard* described above characterizes actual sectoral investment at each point in time. Given firms' capital imbalances at the beginning of a period, the fraction of units resizing their stock of capital is determined by the adjustment hazard. Sectoral investment is the sum of the products of the adjustment hazard and the size of the investment undertaken by those firms that decide to adjust. Equivalently, it is the sum of the expected investment by firms, conditional on their capital stock imbalances before adjusting their capital stock.

Sectoral investment depends critically on the number of firms at each position in the space of capital imbalances, thereby motivating our focus on the *cross-sectional density of disequilibria*. The dynamics of sectoral investment are determined by the evolution of this density. The path of this density is driven by the interaction of sectoral, firm specific, and adjustment cost shocks with the history of shocks and actions contained in previous cross-sectional densities of disequilibria.

2.2. *The Firm*

Net Profits

When the firm is not investing, its flow of net profits is

$$\Pi(K,\theta) = K^{\beta}\theta - (r+\delta)K, \tag{1}$$

where K is the firm's stock of capital, θ is a geometric random walk shock to the profit function that combines demand, productivity, and wage shocks, r and δ are the discount and depreciation rates, and β is a parameter that is less than one, capturing our assumption of decreasing marginal profitability of capital, either due to decreasing returns in the technology or the presence of some degree of monopoly power.[8] For mathematical convenience, we have written the

[7] This should be contrasted with its two limiting cases: the standard (S,s) model, where the probability of a firm adjusting jumps from zero to one at the trigger points, and the standard linear partial adjustment models, where this probability is independent of the size of the firm's disequilibrium.

[8] For concreteness, let the production function be Cobb-Douglas and homogeneous of degree one with respect to capital and labor, with capital share $\alpha < 1$. Let the demand faced by the firm be isoelastic, with price elasticity minus η, $1 < \eta < \infty$. It follows from these assumptions that $\beta \equiv \alpha(\eta-1)/(1+\alpha(\eta-1)) < 1$.

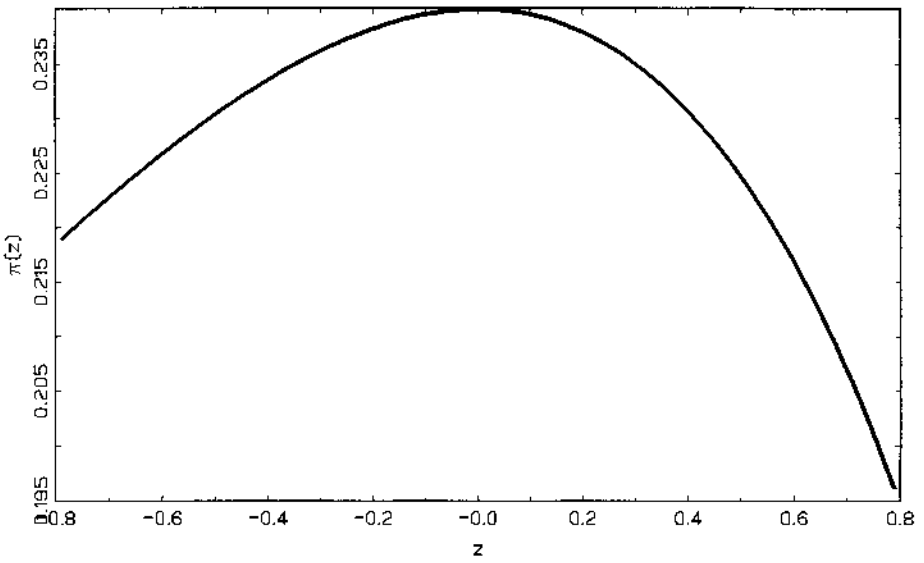

FIGURE 1

profit function net of flow payment on capital, $(r+\delta)K$, where the latter represents the irrevocable commitments associated to purchases of capital.[9]

It is useful to replace θ in the profit function by a variable with more economic content. We do this by defining the frictionless stock of capital of the firm, K^*, as the solution of the maximization of (1) with respect to capital, so that

$$\theta = \xi K^{*(1-\beta)},$$

where $\xi \equiv (r+\delta)/\beta$. Substituting this expression into (1), and defining the disequilibrium variable

$$z \equiv \ln(K/K^*),$$

allows us to rewrite the profit function as

$$\Pi(z, K^*) \equiv \pi(z)K^* = \xi(e^{\beta z} - \beta e^z)K^*. \tag{2}$$

Figure 1 illustrates, and equation (2) implicitly defines, profits *per unit of frictionless capital*, $\pi(z)$.[10]

[9] Since there are neither borrowing constraints nor bankruptcy options, the solution to the firm's problem is unchanged by replacing flow payments for a lump sum payment at the time of purchase. That is, conditional on buying new capital, all that matters to the firm is the present discounted value of payments, not when these payments take place.

[10] The parameters used to generate this figure are $\beta = 0.4$, $r = 0.06$, and $\delta = 0.1$.

Adjustment Costs

When investing, a firm not only commits to pay for the capital acquired, but also incurs adjustment costs. Since we wish to capture the intermittent and lumpy nature of firms' investments, we require these costs to exhibit some form of increasing returns. There are many ways to do so. One possibility is to follow Grossman and Laroque (1990), and assume the firm sells its old stock of capital at a discount when replacing it by a new one. An alternative, with similar implications for our purposes, is to assume firms must shut down operations for a fixed period of time when replacing capital. In the latter case, which is the one we pursue, the firm incurs an adjustment cost proportional to foregone profits due to reorganization:[11]

$$\text{Adjustment Cost} = \omega\{\Pi(K,\theta) + (r+\delta)K\} = \omega K^{\beta}\theta,$$

where ω represents the fraction of profits foregone due to the capital stock adjustment. A derivation similar to the one that led to (2) allows us to rewrite the adjustment cost in terms of z and K^*:

$$\text{Adjustment Cost} = \omega\xi e^{\beta z^-}K^*,$$

where z^- denotes the capital imbalance immediately before adjustment.

Rather than treating ω as fixed—as in standard (S,s) type models—we let it be a random variable with a distribution function, $G(\omega)$, independent across firms and over time, whose realization is observed at the beginning of each period. With this slight generalization of the standard fixed cost framework we capture—in an admittedly stylized form—two realistic features: heterogeneity in adjustment costs at any point in time, and time variation in these costs for any given firm.[12] More importantly, it will be apparent in Section 3 that this extension gives us an important degree of flexibility when estimating aggregate investment equations.

Microeconomic Adjustment

Given the increasing returns nature of the adjustment cost technology, the optimal policy is obviously not one of continuous and small investments but rather one of periods of inaction followed by occasional lumpy investment. Therefore, the firm's problem can be characterized in terms of two regimes: action and inaction. Finding a solution to the firm's problem is equivalent to characterizing the partition of (ω, z)-space into these two regions and specifying firms' actions when located in the region where they act. In what follows we

[11] See, e.g., Cooper and Haltiwanger (1993) for a model where the main cost of reorganization is its opportunity cost.

[12] A more realistic formulation would let adjustment costs exhibit some persistence at the individual level. It would also allow for a distribution of adjustment costs that depends on aggregate conditions. We do not incorporate these features into our model because they complicate substantially both the microeconomic and aggregation problems.

present the basic steps involved in finding this solution. In Appendix A we discuss the technical aspects and intermediate steps of the solution in more detail.

Inheriting the stochastic properties of θ, K^* follows a geometric random walk (with drift):

$$K_t^* = K_{t-1}^* e^{\mu_t},$$

where μ_t is i.i.d. and, throughout most of the paper, Normal. This implies that when there is no adjustment, z_t follows a random walk (with drift and Normal innovations). Together with the i.i.d. nature of ω, this assumption ensures that the firm's decision on whether to adjust its capital stock in period t, and if so by how much, is fully determined by the vector (z_t, K_t^*, ω_t), which we refer to as the "state of the firm." The value of a firm with before-adjustment-disequilibrium z, frictionless stock of capital K^*, and (current) adjustment cost parameter ω—which we denote by $V^*(z, K^*, \omega)$—is the maximum of the value of the firm if it does not adjust, $V(z, K^*)$, and the value if it does adjust, $V(c, K^*) - \omega\xi e^{\beta z} K^*$, where c is the optimally determined return point (see below). In short:

$$V^*(z_t, K_t^*, \omega_t) = \max\{V(z_t, K_t^*), V(c, K_t^*) - \omega_t \xi e^{\beta z_t} K_t^*\}. \tag{3}$$

The evolution of the value of a firm that does not adjust in the current period is described by

$$V(z_t, K_t^*) = \pi(z_t) K_t^* + (1+r)^{-1} \mathrm{E}_t[V^*(z_{t+1}, K_{t+1}^*, \omega_{t+1})]. \tag{4}$$

Since the profit and adjustment cost functions are homogeneous of degree one with respect to K^*, given z, so are the value functions $V(z, K^*)$ and $V^*(z, K^*, \omega)$. This allows us to reduce the number of state variables by relating the problem in terms of the value per unit of frictionless capital. Let $v(z) \equiv V(z, K^*)/K^*$ and $v^*(z, \omega) \equiv V^*(z, K^*, \omega)/K^*$. Dividing both sides of equations (3) and (4) by K^*, and noting that

$$\frac{K_{t+1}^*}{K_t^*} = (1-\delta) e^{-\Delta z_{t+1}},$$

yields

$$v^*(z_t, \omega_t) = \max\{v(z_t), v(c) - \omega_t \xi e^{\beta z_t}\}, \tag{5}$$

$$v(z_t) = \pi(z_t) + \psi \mathrm{E}_t\left[v^*(z_{t+1}, \omega_{t+1}) e^{-\Delta z_{t+1}}\right], \tag{6}$$

with $\psi \equiv (1-\delta)/(1+r)$. Figure 2 depicts in an example the basic setup developed up to now.[13] This figure shows how $v(z)$, $v(c) - \omega\xi e^{\beta z}$, and $v^*(z, \omega)$

[13] Parameters: $\beta = 0.4$, $r = 0.06$, $\delta = 0.1$; the mean and standard deviation of the logarithm of K_t^* are 0 and 0.1; the distribution of adjustment costs is Gamma with mean 0.17 and coefficient of variation 0.16. All the numbers are broadly consistent with our estimates and assumptions in the empirical part of the paper.

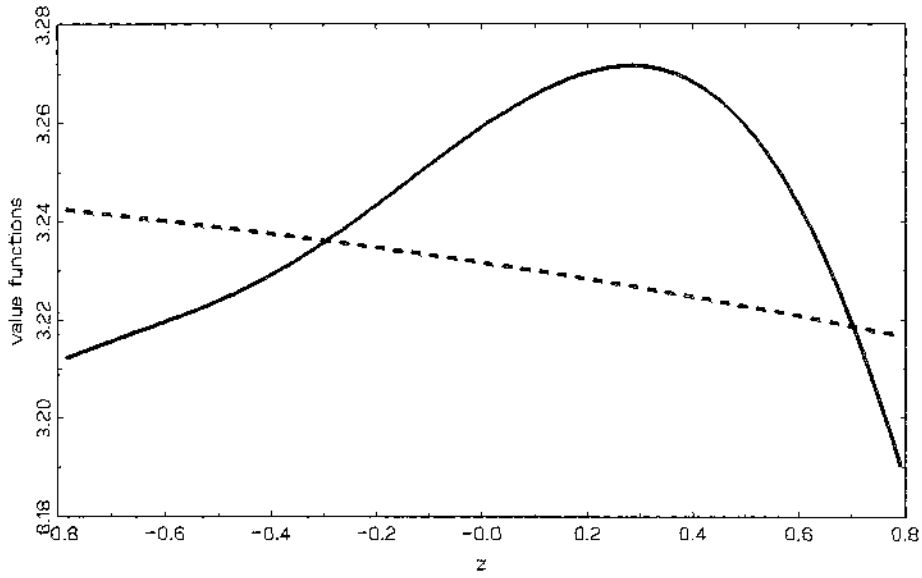

FIGURE 2

determine the trigger points, given a particular realization of the adjustment cost. The solid line illustrates the value of a firm that does not adjust in the current period. The dashed line represents the value of a firm that decides to adjust, given a realization of ω. The maximum between both lines describes $v^*(z, \omega)$, and the inaction range—for a given ω—corresponds to the interval between the intersection of the two lines.

It follows directly from maximization of the value of a firm that decides to adjust, $v(c) - \omega\xi e^{\beta z}$, with respect to the return point c, that the maximum of $v(z)$ and $v^*(z, \omega)$ is obtained at $z = c$ and that this return point is independent of the initial disequilibrium.[14]

The solution also can be characterized by the policy function, $\Omega(z)$, defined as the largest adjustment cost factor for which the firm finds it advantageous to adjust given a capital imbalance z. From the value matching condition that equates the two terms on the right-hand side of equation (5), it follows that

$$\Omega(z) = \xi^{-1} e^{-\beta z}(v(c) - v(z)), \tag{7}$$

[14] Proposition A1 in Appendix A shows that the Bellman equation obtained by substituting $v(z_t)$ from (6) into (5) has a unique solution, which is continuous and bounded. Even though the functions $v(z)$ we obtained via value iteration when performing estimation always had a unique maximum, we have been unable to show this generally. It follows that, strictly speaking, the return point c should be interpreted as one of the points where $v(z)$ attains its maximum, say the smallest value. In Proposition A5 we show that the set of maxima, and hence of possible return points, is finite.

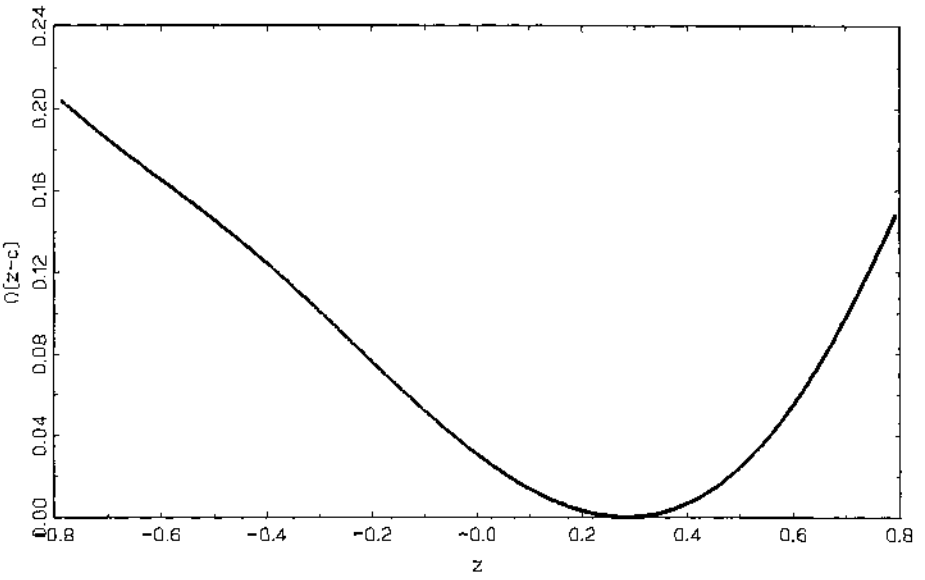

FIGURE 3A

which implies $\Omega(c) = 0$. Differentiating (7) with respect to z, evaluating the result at $z = c$, and using the first order condition $v'(c) = 0$, yields the additional "smooth pasting" condition $\Omega'(c) = 0$.

Figure 3a illustrates the function $\Omega(z)$ for the example in Figure 2, where the distribution of adjustment costs is a Gamma. As follows from equation (7), if the firm's disequilibrium is close enough to $z = c$, it will only adjust for arbitrarily small adjustment costs. From then on, $\Omega(z)$ increases with $|z - c|$.

Figure 3b depicts the inverse of the function $\Omega(z)$. We label $L(\omega)$ and $U(\omega)$ the segments of the curve below and above c, respectively. These functions correspond to the maximum shortage and excess of capital tolerated by the firm for any given realization of the adjustment cost factor ω. That is, for any fixed ω, they describe a standard (L, c, U) policy.[15] The area enclosed by the two curves corresponds to the combinations of disequilibria and adjustment cost factors for which the firm chooses not to adjust.[16]

The shape and location of the function $\Omega(z)$ and its inverse, $(L(\omega), c, U(\omega))$, depend on the entire distribution of adjustment cost factors, $G(\omega)$. A given realization of the adjustment cost factor will not generate the same inaction

[15] An (L, c, U) policy corresponds to a two-sided (S, s) model. The notation L, U, and c stands for lower bound, upper bound, and "center," respectively. See Harrison et al. (1983).

[16] Proposition A2 in Appendix A derives formally the existence of $\Omega(z)$, and Proposition A3 shows that it is an analytic function, and therefore has derivatives of all orders. Proposition A4 shows that $\Omega(z)$ tends to infinity as $|z|$ tends to infinity. Yet we have not been able to show that $\Omega(z)$ is unimodal, and therefore have no formal proof that, conditional on ω, the optimal policy is of the (L, c, U) type. It should be noted, though, that all the policies obtained numerically when estimating the parameters in Section 3 were unimodal.

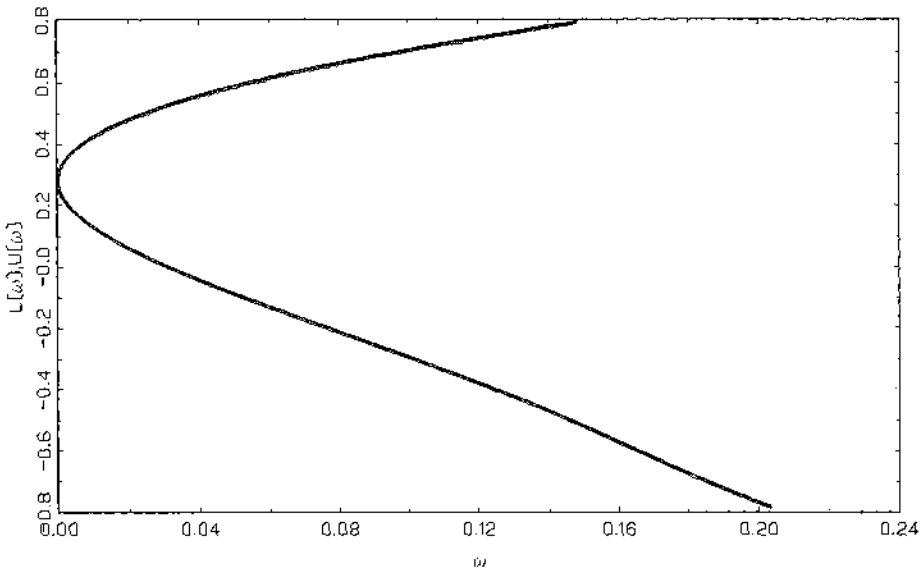

FIGURE 3B

range for different distribution functions $G(\omega)$. In particular, a low value of ω is more likely to lead to action when it comes from a distribution of adjustment cost factors with a high rather than a low average value.

Adjustment Hazard, Expected Investment, and Ergodic Density

Adjustment Hazard: Above we showed that for any given ω the firm follows a simple deterministic policy with respect to z: actions are taken only when z lies outside the $(L(\omega), U(\omega))$ interval, in which case investment occurs so as to bring z back to c. With aggregation in mind, here we reduce the amount of information contained in the policy. Rather than conditioning on ω, we only use information on its distribution and ask the question: what is the probability that a firm with disequilibrium z adjusts?

The answer to this question is contained in what we call the *adjustment hazard*. Let $x \equiv z - c$ denote a firm's imbalance with respect to its target point. A firm with deviation x adjusts only if the current adjustment cost is small enough to make adjusting profitable (i.e., if $\omega < \Omega(x + c)$), which means that the probability of a firm adjusting, conditional on its disequilibrium being equal to x (the *adjustment hazard*), is given by

$$\Lambda(x) = G(\Omega(x + c)), \tag{8}$$

where $G(\omega)$ denotes the cumulative distribution function for the adjustment cost factor ω. For example, if $G(\omega)$ is a *Gamma* distribution with mean $p\phi$ and

variance $p\phi^2$, the adjustment hazard is

$$\Lambda(x) = \frac{1}{\phi^p \Gamma(p)} \int_0^{\Omega(x+c)} \omega^{p-1} e^{-\omega/\phi}\, d\omega.$$

Figure 4a shows the adjustment hazard function for three different gamma distributions of adjustment cost factors. These distributions differ in their mean and variances: the solid line corresponds to an exponential distribution (mean and standard deviation of 0.1); the long dashes correspond to a high variance and mean distribution (mean: 80; standard deviation: 282), while the short dashes describe a low variance and mean distribution (mean: 0.14; standard deviation: 0.044). These examples illustrate the range of cases covered by our setup. Figure 4a shows that when the variance of adjustment costs is low, there is a range of adjustment costs where the firm (almost) never adjusts since adjustment costs are (almost) never small enough to justify it; the standard (S, s) —or (L, c, U)—case is an extreme version of this. Conversely, when the variance of adjustment costs is high, and so is their mean, the decision of adjustment is largely motivated by the adjustment cost draw rather than by the firm's disequilibrium; in the limit, adjustment costs are independent of the firm's disequilibrium, yielding the standard linear partial adjustment model.

In Proposition B2 in Appendix B we show that $\Lambda(x)$ is differentiable, with $\Lambda(0) = 0$ and $\lim_{|x|\to\infty} \Lambda(x) = 1$.

Expected Investment: A firm with disequilibrium x has a probability $\Lambda(x)$ of adjusting its stock of capital and, if it does so, it invests

$$(e^c - e^z)K_t^* = (e^{-x} - 1)e^z K_t^* = (e^{-x} - 1)K_t(x).$$

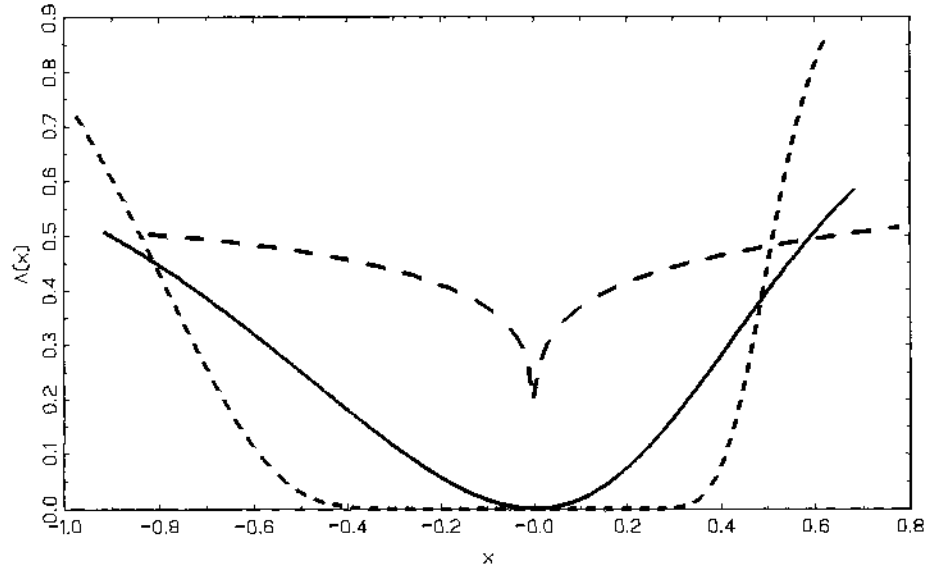

FIGURE 4A

Thus the average investment of firms with disequilibrium x immediately before adjusting their capital stock in period t is

$$\text{(9)} \qquad \mathrm{E}_{\omega}[I_t(x)|x] = \Lambda(x)(e^{-x} - 1)K_t(x).$$

Figure 4b depicts expected investment corresponding to the hazards in Figure 4a (with $K_t(x) \equiv 1$). The nonlinear-convex nature of expected investment is an important feature of the model, playing a key role in shaping aggregate investment dynamics. It says that incentives to invest rise more than proportionally with a firm's disequilibrium.[17]

Ergodic density: We conclude our characterization of microeconomic behavior by stating that the disequilibria through which a firm goes over its lifetime, have an invariant density. The formal proof, which also shows that convergence takes place at an exponential rate, is given in Appendix B. The nonlinear nature of microeconomic adjustment, with relatively small adjustment for small imbalances, imprints the opposite pattern on the ergodic density: steeper hazards translate into relatively less mass in the tails of the corresponding invariant density in exchange for more mass in the regions of low values of the adjustment hazard (i.e., relatively platokurtic).[18]

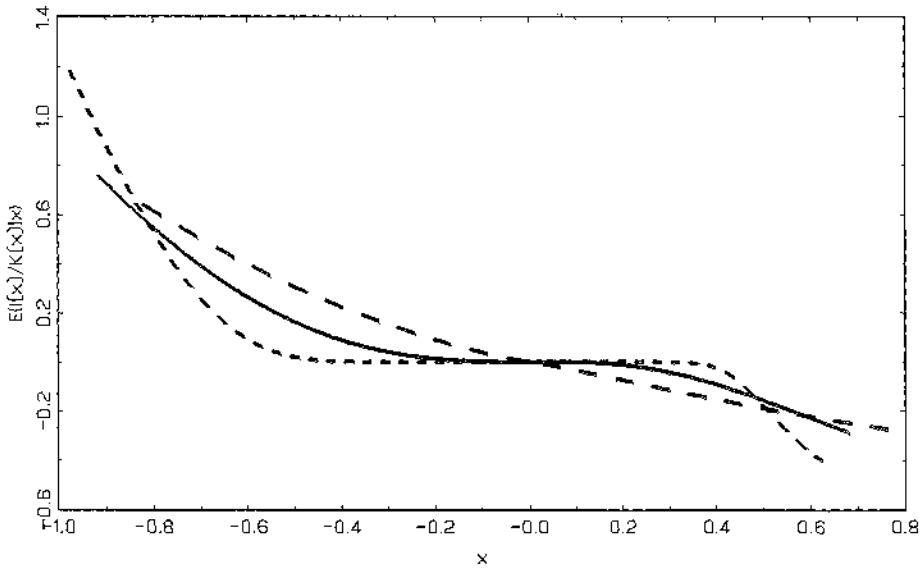

FIGURE 4B

[17] This is a feature not shared by the standard quadratic adjustment cost model but it is certainly not exclusive of models with fixed, or even nonconvex, adjustment costs.

[18] The U.S. manufacturing plant level data studied in Caballero, Engel, and Haltiwanger (1995) revealed that the average observed distribution of disequilibria is considerably more platokurtic than the average distribution of shocks affecting these plants.

2.3. *Sectoral Investment*

Sectoral Investment and the Cross-Sectional Density

Let K_t^A, I_t^A, $K_t(x)$, and $I_t(x)$ denote the aggregate (sectoral) stock of capital and gross investment, and the stock of capital and gross investment held by firms with disequilibrium x at time t (before adjustment).

Since adjustment cost shocks are i.i.d. across firms, it follows directly from (9) that

$$I_t(x) = (e^{-x} - 1)\,\Lambda(x)\overline{K_t(x)}, \tag{10}$$

where $\overline{K_t(x)}$ denotes the average stock of capital of firms with imbalance x.

Letting $\tilde{f}(x,t)$ denote the cross-sectional density of disequilibria just before adjustments take place, we can obtain an expression for aggregate investment:

$$I_t^A = \int (e^{-x} - 1)\,\Lambda(x)\overline{K_t(x)}\tilde{f}(x,t)\,dx.$$

Dividing through by K_t^A and rearranging terms, we obtain an expression for the aggregate investment/capital ratio:

$$\frac{I_t^A}{K_t^A} = \int (e^{-x} - 1)\,\Lambda(x)\tilde{f}(x,t)\,dx + \int (e^{-x} - 1)\,\Lambda(x)\left(\frac{\overline{K_t(x)}}{K_t^A} - 1\right)\tilde{f}(x,t)\,dx. \tag{11}$$

The second term on the right-hand side of (11) drops out if $(\overline{K_t(x)} - K_t^A)$ and $(e^{-x} - 1)\,\Lambda(x)$ are uncorrelated. Since such an assumption simplifies computations substantially, we make it and obtain an approximate expression for the aggregate investment/capital ratio:

$$\frac{I_t^A}{K_t^A} \simeq \int (e^{-x} - 1)\,\Lambda(x)\tilde{f}(x,t)\,dx. \tag{12}$$

In Appendix C we describe in detail the additional computational burden of using (11) instead of (12), and present the results of Monte Carlo simulations showing that the cost of the approximation is only minor.[19]

It is apparent from (12) that since $(e^{-x} - 1)\,\Lambda(x)$ is generally nonlinear in x, aggregate investment depends not only on the first but also on higher moments of the cross-sectional distribution of disequilibria.

[19] The results in Caballero, Engel, and Haltiwanger (1995) are reassuring on this respect. There we used comprehensive establishment level data for U.S. manufacturing during the 70s and 80s, and documented a very close empirical fit between the actual and approximate series during that period.

The Linear/Partial Adjustment Extreme

An exception to the statement in the previous paragraph occurs when the adjustment hazard does not depend on x, *and* $\tilde{f}$ has most of its mass near $x = 0$, so that $e^{-x} - 1 \simeq -x$. In that case

$$\frac{I_t^A}{K_t^A} \simeq -\Lambda_0 \bar{X}_t, \tag{13}$$

where Λ_0 denotes the constant hazard, and $\bar{X}_t$ denotes the average disequilibrium before adjustment. Equation (13) corresponds to the well known partial adjustment model (PAM), and also coincides with the standard linear equation arising from the quadratic adjustment costs model. This is the only adjustment hazard that does not require cross-sectional information on the right-hand side of the aggregate investment equation. Indeed, a few steps of algebra allow us to go from equation (13) to the standard expression:[20]

$$\frac{I_t^A}{K_t^A} \simeq \Lambda_0(\delta + v_t) + (1 - \Lambda_0)\frac{I_{t-1}^A}{K_{t-1}^A}, \tag{14}$$

where v_t represents an aggregate shock, to be defined in the next paragraph.

Sectoral Equilibrium and Cross-sectional Dynamics

Shocks to wages, demand, and productivity drive the dynamics of frictionless capital. We decompose these shocks into sectoral shocks, v_t, and firm specific (idiosyncratic) shocks, ϵ_t:

$$K_t^* = K_{t-1}^* e^{v_t + \epsilon_t},$$

which implies that when the firm does not adjust, the disequilibrium measure x evolves according to

$$\Delta x_t = -(\delta + v_t) - \epsilon_t,$$

where capital depreciates at a rate δ from one period to the next. We assume these shocks are exogenous to the firm and the sector.

Between two consecutive periods, the cross-sectional distribution of disequilibria changes as a result of firms' adjustments, depreciation, sectoral, and idiosyncratic shocks. Since we are working in discrete time, it is important to describe the timing convention we adopt for events within each period. We

[20] Let k_t and k_t^* denote the average of the logarithm of the pre-adjustment stock of capital and the frictionless stock of capital, respectively. We define $v_t \equiv \Delta k_t^*$, and note that $\Delta k_t = (I_{t-1}^A/K_{t-1}^A) - \delta$. Combining these two expressions with the fact that

$$\bar{X}_t = k_t - k_t^* = \Delta k_t - \Delta k_t^* + \bar{X}_{t-1}$$

yields (14).

denote the cross-section density at the end of period $t-1$ by $f(x,t-1)$. Depreciation and the aggregate shock corresponding to period t follow, resulting in the density $\tilde{f}(x,t)$. Next come adjustments, as determined by the hazard function $\Lambda(x)$. Period t concludes with the idiosyncratic shocks. The final density is $f(x,t)$, and the cycle starts again. Recalling that a positive shock leads to a decrease in x, we can summarize this chain of events as follows:

$$\tilde{f}(x,t) = f(x+\delta+\upsilon_t, t-1), \tag{15}$$

$$f(x,t) = \left[\int \Lambda(y)\tilde{f}(y,t)\,dy\right] g_\epsilon(-x) + \int [1-\Lambda(x+\epsilon)]\tilde{f}(x+\epsilon,t) g_\epsilon(-\epsilon)\,d\epsilon, \tag{16}$$

where $g_\epsilon(\epsilon)$ is the probability density for the idiosyncratic shocks. The integro-difference equation describing the evolution of the cross-sectional distribution from one period to the next follows directly from equations (15) and (16):

$$f(x,t) = \left[\int \Lambda(y) f(y+\delta+\upsilon_t, t-1)\,dy\right] g_\epsilon(-x) + \int [1-\Lambda(x+\epsilon)] f(x+\epsilon+\delta+\upsilon_t, t-1) g_\epsilon(-\epsilon)\,d\epsilon. \tag{17}$$

From equations (12) and (15) we obtain the following *aggregate* investment equation:

$$\frac{I_t^A}{K_t^A} = \int (e^{-x}-1)\Lambda(x) f(x+\delta+\upsilon_t, t-1)\,dx. \tag{18}$$

Combining equations (18) and (17) we can determine the sequence of aggregate investment determined by an initial cross-section distribution, $f(x,0)$, and a sequence of aggregate shocks, $\{\upsilon_t\}$. For details see Appendix C. We turn to estimation issues next.

3. EMPIRICAL EVIDENCE: U.S. MANUFACTURING INVESTMENT

3.1. *Data*

Our data are constructed from annual gross investment and capital series for 21 two-digit manufacturing industries from 1947 to 1992.[21] All series are in 1987 dollars, and the stocks of capital correspond to the series used by the Bureau of Labor Statistics for their productivity studies.[22] Since capital stocks are end-of-year, our measures of the investment/capital ratio used in estimation start in

[21] We have 21 rather than 20 sectors because Motor Vehicles is separated from Transportation equipment.

[22] This is one of the three capital stock series reported by the Bureau of Economic Activity.

1948. We report separate results for equipment and structures panels; each has 945 observations.

3.2. *Econometrics*

The econometric problem consists of estimating the parameters that characterize (a) firms' profit functions, (b) the initial distribution of disequilibria, (c) the distribution of adjustment costs, (d) the distribution of idiosyncratic shocks, and (e) the process generating aggregate shocks. In this subsection we outline the main features of the estimation procedure; a detailed description is presented in Appendix C.

For tractability, we limit the number of parameters being estimated to those characterizing the distribution of adjustment costs or, equivalently, the hazard function. As far as the remaining parameters, we either fix them (profit function and distribution of idiosyncratic shocks), show that their role is limited within a reasonable range (initial distribution of disequilibria), or concentrate them out of the likelihood function (process generating sectoral shocks: individual effects and cross-sectoral variance-covariance matrix). Since identifying nonlinearities from a purely time series (as opposed to regressions) dimension requires a large number of observations, we impose a hazard function or a distribution of adjustment costs that is common across sectors (depending on whether we estimate a semi-structural or structural model—see below).

The sources of randomness in our estimation procedure are the sectoral shocks, which we assume are multivariate Normal and independent over time, for most of our empirical analysis.

We approximate the initial sectoral cross-sections by the invariant cross-section of an individual plant, and proceed to use the Markovian nature of the process generating cross-sectional distributions to generate these distributions. For each sector, and at each date, the cross-sectional distribution is updated as a function of the sectoral shock, using an implicit law of large numbers at the microeconomic (firm) level. The observed sectoral investment rate is a nonlinear function of the current shock and the distribution prior to this shock—this function is one-to-one in the shock. Conditional on the initial distribution, the sequence of sectoral shocks and the cross-section distributions can be recovered from the time series of sectoral investment rates. The likelihood is calculated using these sectoral shocks. We also need to calculate the corresponding Jacobian terms, which correspond to the elasticities of sectoral investment rates with respect to sectoral shocks. These elasticities are a byproduct of the calculation of sectoral shocks.

3.3. *Semi-Structural and Structural Models*

We estimate two basic models. In the first one (semi-structural), we estimate directly the parameters of an ad-hoc adjustment hazard. While in the second one (structural), we estimate the adjustment cost parameters and obtain the

implied hazard via dynamic programming. Both of these models yield increasing hazards,[23] but include a constant hazard as a limit case.

Semi-structural: Although our main goal in the paper is to estimate and assess the structural model, there are good reasons to start by estimating a less structured version. It allows us to search and test for the presence of an *increasing hazard* more directly, and it facilitates comparisons with standard linear models. Furthermore, by assuming that the adjustment hazard is an inverted Normal:

$$\Lambda(x) = 1 - e^{-\lambda_0 - \lambda_2 x^2}, \tag{19}$$

with $\lambda_0 \geq 0$ and $\lambda_2 \geq 0$, we are able to obtain accurate computations and to reduce estimation time significantly (by a factor of 12) by keeping the cross-section distributions within a closed family of mixture of Normals (see Appendix C).

Structural: Rather than estimating the adjustment hazard directly, in this case we estimate the parameters of the adjustment cost function and obtain the hazard from the solution of the dynamic optimization problem presented in Section 2. Adjustment costs are drawn from a Gamma distribution:

$$G(\omega) = \frac{1}{\phi^p \Gamma(p)} \int_0^{\omega} \eta^{p-1} e^{-\eta/\phi} d\eta,$$

which has mean $\mu_\omega = p\phi$ and a coefficient of variation $\mathrm{cv}_\omega \approx 1/\sqrt{p}$. We estimate μ_ω and cv_ω.

As for the other structural parameters, we assume an interest rate, share of each type of capital, and markup of 6, 15, and 20 percent, respectively,[24] as well as depreciation rates for equipment and structures of 10 and 5 percent per year, respectively.[25] We estimated the standard deviation of idiosyncratic shocks, σ_ϵ, obtaining estimates in the range of 5 to 15 percent. Since these were not estimated very precisely, and comparisons across models are easier if idiosyncratic variances are similar, we only report results where we have imposed $\sigma_\epsilon = 0.1$ (both, in semi-structural as well as structural estimation).

[23] By increasing hazard we mean a hazard that increases with $|x|$, i.e., that is decreasing for $x < 0$ and increasing for $x > 0$.

[24] These parameters imply a value of β around 0.45 if the production function is constant returns and all other factors of production (including the other type of capital) are fully flexible, around 0.3 if all factors but the other form of capital are flexible, and around 0.15 if all other factors are fixed. Our conclusions are robust to reasonable variations of β, but we do not have enough power to identify this parameter in conjunction with those that we estimate. The results we report assume $\beta = 0.4$.

[25] Although there is a slight upward trend in the sample, these depreciation rates are consistent with the average depreciation rate computed from the ratio of actual depreciation to net capital stocks reported in Fixed Reproducible Tangible Wealth in the United States, 1925–89. In any event, it follows from our description in Appendix C that, conditional on the initial cross-section, the depreciation rate is confounded with the mean of aggregate shocks. Thus our choice of depreciation rate only affects the initial cross-section. And since we discard the first three periods when calculating the likelihood, this effect is minor.

3.4. *Main Results*

Table I contains our main results. The first two columns present semi-structural and structural results for equipment investment, while the last two do the same for investment in structures. All estimated models allow for individual effects on the sectoral shocks, and include a free additive constant that is common across sectors.

Semi-structural: The semi-structural results allow us to reject the *constant* hazard model, in favor of an *increasing* hazard one. The increasing hazard parameter, λ_2, is significant at the one percent level in both cases. The estimated hazard function suggests that the probability that a firm adjusts its capital stock of equipment increases from about 14% for small imbalances to 45% when its imbalance is 40%, while it goes from close to zero for small imbalances to about 32% for a 40% imbalance, in the case of structures. The sharp nonlinearity can also be captured through the expected investment/capital ratio (conditional on the imbalance); for equipment, it goes from close to 0.05 at a 20 percent imbalance to 0.23 at a 40 percent imbalance, while for structures it goes from 0.02 to 0.16, for the same imbalances.

Structural: The results of the structural models confirm the semi-structural increasing hazard findings. Moreover, the likelihoods rise, especially so for structures.

The estimates of the mean of the distribution of adjustment costs (the μ_ωs) indicate that the average adjustment cost drawn is the equivalent of 16.7 percent of a year's operational profits for equipment and 22.8 percent for structures. Since firms can "search" for a low realization of adjustment costs, these are upper bounds for the average costs effectively paid by firms when going through a major adjustment episode. Indeed, the average costs *paid* are 11.1 percent for

TABLE I

MAIN RESULTS

	Equipment		Structures	
Parameters	Semi-structural	Structural	Semi-structural	Structural
λ_0	0.155		0.000	
	(0.067)		(0.054)	
λ_2	2.804		2.437	
	(1.165)		(0.878)	
constant	0.057	−0.006	0.013	0.019
	(0.013)	(0.016)	(0.006)	(0.002)
μ_ω		0.166		0.228
		(0.029)		(0.046)
cv_ω		0.327		0.066
		(0.109)		(0.009)
LLK	2430.2	2431.4	2612.4	2637.2
LLK-NADJ	2315.2	2315.2	2497.0	2497.0

Notes: Standard deviations in parenthesis. LLK: log-likelihood. LLK-NADJ: log-likelihood without adjustment costs/dynamics.

equipment and 21.4 percent for structures. The difference between the unconditional and conditional (on adjustment) means rises with the coefficient of variation of these costs, which explains why adjustment costs actually paid for equipment are one third less than the mean adjustment cost faced by firms (see cv_ω) while in the case of structures both means are very similar.

Comparing the last two rows of Table I illustrates the goodness-of-fit of the model. LLK represents the log-likelihood of the model while LLK-NADJ represents the log-likelihood of a model with no adjustment costs or dynamics (i.e., only a constant). The likelihood ratio test statistics for both equipment and structures are over 200.

3.5. *Simple Alternatives*

We view the microeconomic foundation of our approach as one of its main virtues; however in this section we look purely at the *statistical* advantage of our structural models over simple linear counterparts. We do not intend to conduct "horse races" against alternative investment models but want to provide a simple metric to assess the contribution of the nonlinearities we estimated to the time series properties of aggregate investment.

The first obvious step is to compare our model with the "almost" nested PAM. As we argued above, the constant hazard model ($\lambda_2 = 0$), together with the approximation $e^{-x} - 1 \simeq -x$, yields the standard PAM, which corresponds to estimating an AR(1) for each sector's aggregate investment series. For comparability with the structural model, we constrain the correlation coefficient to be the same across sectors:[26]

$$\frac{I_{it}^A}{K_{it}^A} = a_i + \rho \frac{I_{it-1}^A}{K_{it-1}^A} + v_{it}.$$

More generally, we also run an AR(2) with unconstrained autoregressive coefficients for each sector:

$$\frac{I_{it}^A}{K_{it}^A} = a_i + \rho_{1i} \frac{I_{it-1}^A}{K_{it-1}^A} + \rho_{2i} \frac{I_{it-2}^A}{K_{it-2}^A} + v_{it}.$$

In both cases we preserve the assumption of jointly Normal aggregate shocks, and allow for individual effects. We look at within and out-of-sample criteria, and find widespread evidence supporting the structural nonlinear model over the linear representation for both equipment and structures.

Within Sample Criteria

The likelihoods of the linear models are uniformly lower than those of the corresponding structural models, even for the AR(2)s—which have 39 parame-

[26] Recall that the structural model has the same distribution of adjustment costs across sectors.

ters more than the nonlinear structural models (the likelihoods for the linear models are shown in Table II). But comparing the likelihoods is not strictly correct, since the linear models (especially the AR(2)s) are not nested in our structural models. Instead, we use the test for nonnested models developed by Vuong (1989) and Rivers and Vuong (1991).

Let l_1 and l_2 denote the maximum value attained by the log-likelihood for models 1 and 2, T denote the number of periods considered when calculating the likelihood (42, in our case), n_1 and n_2 denote the number of parameters for models 1 and 2, respectively, and $\hat{S}$ denote the Newey-West estimate for the variance of the time series of likelihood differences. The null hypothesis is that both models are $\sqrt{T}$-asymptotically equivalent;[27] should this be the case the test statistic

$$V_{1,2} = \frac{l_1 - l_2 - \frac{1}{2}(n_1 - n_2)\log(T)}{\sqrt{\hat{S}T}}, \tag{20}$$

has a Standard Normal distribution. Positive values of $V_{1,2}$ indicate evidence in favor of model 1; negative values evidence in favor of model 2.[28]

Table II presents Vuong's statistics and the p-values for the test that both models (linear and nonlinear) are equally close to the "true" model, against the alternative that the (nonlinear) structural model is closer. It is apparent that the null hypothesis is rejected in favor of the alternative even at very low significance levels.

TABLE II

NONNESTED MODELS TEST: NORMAL SHOCKS

	Equipment		Structures	
	PAM	AR(2)-UNC	PAM	AR(2)-UNC
Vuong statistic	2.61	4.25	2.92	4.07
p-value	0.0045	< 0.0001	0.0018	< 0.0001
Log-Lik.	2387.2	2419.0	2533.2	2578.5

Notes: Vuong statistic calculated as in (20). All test statistics compare the structural model estimated in Table I with the model in the table. If both models are "equally good," the asymptotic distribution of the statistic is Standard Normal. Large positive values provide evidence in favor of the structural model.

[27] I.e., $\lim_{T \to \infty} \sqrt{T}(l_1 - l_2) = 0$.

[28] Note that: First, the numerator of (20) contains a penalty term—the Bayesian Information Criterion, BIC—that corrects for differences in degrees of freedom between both models. Second, denoting the sum of sectoral likelihoods for model i at time t by $l_{i,t}$, $i = 1, 2$, and $d_t \equiv l_{1t} - l_{2t}$, we have that $\hat{S}$ in the denominator of $V_{1,2}$ is

$$\hat{S}(T, q) = \gamma_0 + 2 \sum_{j=1}^{q} \left[1 - \frac{j}{q+1}\right] \gamma_j,$$

where γ_j denotes the sample autocorrelation of order j of the d_t time series. Since in all cases $\hat{S}(T, q)$ does not vary much for values of q larger than 7, we choose $q = 8$ when calculating $\hat{S}$ in (20).

One possible reason for the bad relative performance of linear models is that we have assumed that aggregate shocks are Normally distributed. Sectoral investment rates, on the other hand, are clearly not Normal; the skewness and (excess) kurtosis coefficient of (standardized, for every sector) investment rates are 0.61 and 0.74 for equipment and 0.76 and 0.87 for structures. Obviously, linear models with Normal errors cannot account for these departures from Normality. The innovations generated by the best partial adjustment model and best second-order autoregressive models also depart from Normal, as can be seen in Table III: Their skewness and kurtosis coefficients are 0.49 and 1.15 for equipment and 0.95 and 1.88 for structures in the partial adjustment case, and 0.38 and 1.00 for equipment, and 0.86 and 1.65 for structures in the AR(2) case. All these numbers are significantly different from zero (the Normal case) at the 0.001 level (estimated via bootstrap). The last two rows show that the increasing hazard model generates innovations that are closer to Normal than its linear and constant hazard counterparts. The estimated skewness and kurtosis coefficients are considerably smaller, and both skewness coefficients do not depart significantly (at the 0.05 level) from their values under the Normality assumption. The increasing hazard model does not need to introduce nearly as much skewness and kurtosis in aggregate shocks to account for investment behavior.

Normality is the natural assumption when aggregate shocks are conceived as the sum of a wide variety of small shocks with limited dependence (by the Central Limit Theorem). In spite of this, we momentarily relax this assumption in order to consider shocks that admit skewness and kurtosis properties similar to those observed in sectoral investment/capital ratios. For this purpose, we generalize the distribution of the residual to consider convex combinations of Normal and log-Normal distributions. The log-Normal component does not add significantly to the structural model, while the linear models assign most of the weight to the log-Normal component. For this reason we compare the structural models with Normal shocks versus the linear models with log-Normal shocks in Table IV. Although the likelihood in the linear model improves substantially with this modification, Table IV shows that the test for nonnested models still favors the (nonlinear) structural models by a wide margin. In fact, the reduction

TABLE III

SKEWNESS AND KURTOSIS FOR INNOVATIONS

	Equipment		Structures	
Model	Skewness	Kurtosis	Skewness	Kurtosis
Part. Adj.	0.49	1.15	0.95	1.88
	(0.13)	(0.36)	(0.15)	(0.65)
AR(2)	0.38	1.00	0.86	1.65
	(0.12)	(0.31)	(0.14)	(0.60)
Structural	−0.04	0.65	0.17	0.79
	(0.11)	(0.20)	(0.13)	(0.34)

Note: Standard deviations, obtained via bootstrap, shown in parentheses.

TABLE IV

NONNESTED MODELS TEST: LOG-NORMAL / NORMAL SHOCKS

	Equipment		Structures	
	PAM	AR(2)-UNC	PAM	AR(2)-UNC
test (Normal)	2.42	7.97	3.57	4.53
p-value	0.0078	< 0.0001	0.0002	< 0.0001
Log-Lik.	2409.3	2453.8	2611.3	2678.6

Notes: Vuong statistic calculated as in (20). All test statistics compare the structural model estimated in Table I with the model in the table. If both models are "equally good," the asymptotic distribution of the statistic is Standard Normal. Large positive values provide evidence in favor of the structural model.

in the denominator due to the increased precision of the test (the likelihoods become more correlated across models) more than outweighs the increase in the likelihood of the linear model.

Out-of-Sample Criteria

Next, we evaluate the out-of-sample forecasting performance of our model. For this purpose we reestimate the nonlinear structural and AR(2) models dropping ten percent of our observations (the last five years for each of our 21 sectors), and generate the one-step-ahead forecast distributions, for each sector and year out of the sample. We only evaluate the model's performance relative to that of an AR(2) using a standard Mean-Square-Error criterion, although this reduces the potential forecasting edge of nonlinear models.[29] We postpone further discussion of forecasts' higher moments until the conclusion.

Table V reports, for each investment type, the average (across sectors) MSE for each year (first two columns), and the percentage increase in MSE generated by the AR(2) model over the nonlinear one (third column). Except for 1988, the structural nonlinear model systematically outperforms the AR(2) representation. This is particularly true for structures, where the gain is over 35 percent during four out of the five years for which we generated out-of-sample forecasts.

TABLE V

AVERAGE MSE FOR 1-STEP-AHEAD FORECASTS: 1988–92

	Equipment			Structures		
Year	AR(2)	MODEL	ln((1)/(2))	AR(2)	MODEL	ln((1)/(2))
88	0.591	0.640	−0.075	0.306	0.340	−0.092
89	0.240	0.230	0.038	0.371	0.220	0.502
90	0.308	0.270	0.131	0.207	0.150	0.359
91	0.566	0.540	0.044	0.507	0.350	0.371
92	0.399	0.330	0.194	0.518	0.340	0.425

Note: The parameters were estimated using data up to 1987. MSEs are multiplied by 10^3.

[29] See Ramsey (1996) for arguments on the bias against nonlinear models inherent in MSE comparisons.

This favorable evidence for the nonlinear model is reinforced by Table VI. It reports, for each investment type, the average (across years) MSE for each sector for the AR(2) and nonlinear models (first two columns), and the percentage increase in MSE generated by the AR(2) model over the nonlinear one (third column). At the bottom of the table, we report the percentage increase in average (across sectors and time) MSE generated by the AR(2) model over the nonlinear one, as well as the median (across sectors) increase. The sectoral dimension is one along which we would have expected the nonlinear model to do relatively poorly, since in order to gain statistical power for the nonlinearities we were forced to impose the same distribution of adjustment costs across sectors, which is not likely to hold too closely in the data. The AR(2), on the other hand, has no constraints across sectors. Table VI shows that even under this unfavorable metric the structural model outperforms the unconstrained AR(2) representation. Again, this is particularly true for structures, where the gain in terms of MSE is over 15 percent for the median sector, and above 30 percent for the average MSE.

TABLE VI

SECTORAL AVERAGE MSE FOR 1-STEP-AHEAD FORECASTS: 1988–92

	Equipment			Structures		
Sector	AR(2)	MODEL	ln((1)/(2))	AR(2)	MODEL	ln((1)/(2))
20	0.591	0.780	−0.425	0.742	0.150	1.571
21	0.085	0.084	0.003	0.117	0.150	−0.233
22	0.343	0.310	0.091	0.221	0.120	0.612
23	0.243	0.110	0.781	0.278	0.094	1.088
24	0.160	0.240	−0.405	0.342	0.250	0.333
25	0.083	0.092	−0.105	0.560	0.260	0.760
26	0.034	0.068	−0.707	0.633	0.400	0.456
27	1.620	1.710	−0.055	0.386	0.720	−0.619
28	0.129	0.180	−0.341	0.313	0.400	−0.247
29	0.098	0.100	−0.036	0.628	0.570	0.097
30	0.272	0.260	0.035	0.417	0.088	1.560
31	0.082	0.089	−0.097	0.063	0.054	0.159
32	1.433	1.000	0.357	1.483	1.000	0.392
33	0.078	0.076	0.023	0.017	0.024	−0.352
34	1.739	1.410	0.211	0.201	0.270	−0.288
35	0.551	0.730	−0.276	0.353	0.400	−0.127
36	0.506	0.420	0.181	0.328	0.390	−0.170
37-1	0.310	0.220	0.338	0.099	0.150	−0.411
37	0.241	0.270	−0.124	0.494	0.075	1.888
38	0.221	0.200	0.120	0.184	0.210	−0.124
39	0.099	0.072	0.319	0.162	0.100	0.467
$\ln(\Sigma_i(1)/\Sigma_i(2))$	—	—	0.047	—	—	0.312
Median	0.241	0.220	0.003	0.328	0.210	0.159

Note: The parameters were estimated using data up to 1987. MSEs are multiplied by 10^3.

3.6. *Dynamic Implications of Increasing Hazard Models*

Perhaps the main distinctive feature for the model we have estimated, compared with its linear counterparts or a constant hazard model, is that not only average investment by those that are investing but also the number of firms that choose to invest at any point in time fluctuates over the business cycle. This is a realistic feature according to the establishment-level evidence in Doms and Dunne (1993). Among many interesting facts, they show that the number of plants going through their primary investment spikes (i.e., the single year with the largest investment for the establishment), rather than the average size of these spikes, tracks closely aggregate manufacturing investment over time.

In terms of our model, this flexibility in the number of firms investing implies that the extent of the response of aggregate investment to aggregate shocks fluctuates over the business cycle. Figure 5 depicts the paths of the median (across sectors) derivatives of aggregate investment with respect to aggregate shocks for equipment and structures.[30] It is apparent that this "index of responsiveness" fluctuates widely over the sample. Moreover, it is strongly procyclical: Its correlations with aggregate shocks and aggregate investment are, respectively, 0.79 and 0.89 for equipment, and 0.72 and 0.39 for structures.[31]

There are traces of the cyclical features of our nonlinear model in our out-of-sample forecasts as well. The MSE gains of our model over the linear AR(2) are particularly pronounced during periods of high activity. To show this, we proceed in three steps: First, we construct, for each sector, a standardized series of the difference of the absolute values of the forecast error of the AR(2) and the structural model, $\Delta|\epsilon^f|$:

$$\Delta|\epsilon_{it}^f| \equiv \frac{|\epsilon_{it}^{f_{ar}}| - |\epsilon_{it}^{f_{nl}}|}{\sigma_{i,\,\Delta\epsilon^f}},$$

[30] If we define $y(v)$ as the right-hand side of (18) evaluated at v instead of v_t, then this index is equal to the derivative of y evaluated at v_t. Making the change of variable $u = x - v$ in the integral that defines $y(v)$ and differentiating under the integral leads to

$$y'(v_t) = \int [e^{v_t - u} A(u - v_t) - (e^{v_t - u} - 1) A'(u - v_t)] f(u + \delta, t - 1)\, du.$$

Adding and subtracting $A(u - v_t)$ to the first term in the integral, using (18) and changing variables gives

$$y'(v_t) = y(v_t) + \int A(x) f(x + \delta + v_t, t - 1)\, dx - \int (e^{-x} - 1) A'(x) f(x + \delta + v_t, t - 1)\, dx.$$

Alternatively, making the approximation $(e^{-x} - 1) \approx -x$ before differentiating, we obtain an index of responsiveness equal to $y'(v_t) - y(v_t)$, which is constant for the constant hazard case (PAM, in that case). The figure obtained with this alternative index is qualitatively identical to Figure 5, although the standard deviation of the index is about 30 percent less than that of the index reported.

[31] The correlations with lagged investment/capital ratios are 0.60 and 0.13, for equipment and structures, respectively.

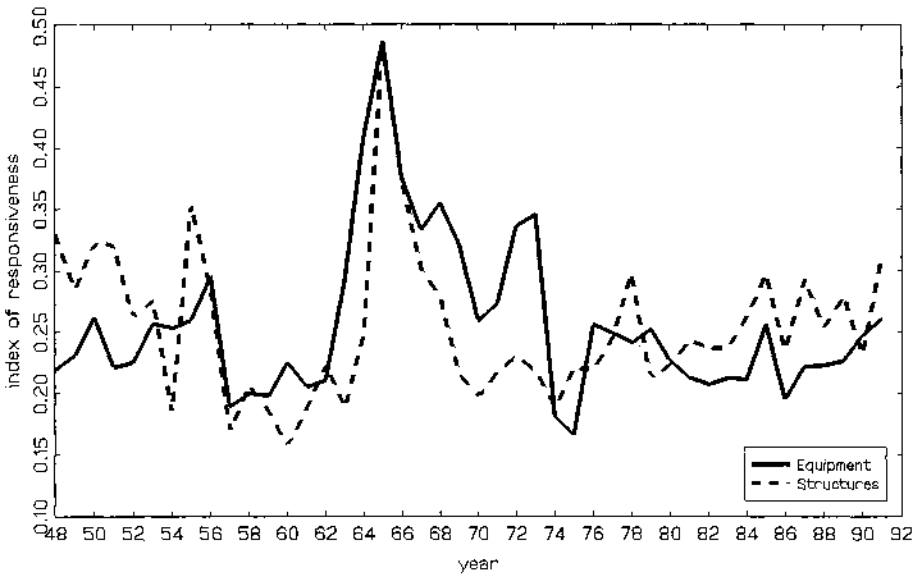

FIGURE 5

where ϵ^f denotes the forecast errors (ar and nl stand for AR(2) and nonlinear model, respectively), and $\sigma_{\Delta\epsilon^f}$ denotes the standard deviation of the difference in the absolute value of the forecast errors. Second, for each sector, we sort these standardized series by one of the following indicators of activity: the sectoral aggregate shock, v, the level of the sectoral investment/capital ratio, y, and the sectoral index of responsiveness, dy/dv. And third, we average across sectors the sorted standardized series.

Table VII reports these averages for times when the sorting variable was below and above its median. With only one exception, all the entries suggest that a substantial fraction of the better performance of the nonlinear model comes from periods when the sectoral indicators of activity (shock, investment, and sensitivity index) are high. For example, we find that periods when aggregate shocks are below their median, achieve a forecasting-performance improvement which is 0.286 standard deviations lower than the average MSE gain for equipment, while it is 0.056 standard deviations lower for structures. Conversely,

TABLE VII

STANDARDIZED DIFFERENCE IN ABSOLUTE VALUE OF FORECAST ERRORS (AR2-MODEL)

	Equipment			Structures		
	Sort by v	Sort by y	Sort by dy/dv	Sort by v	Sort by y	Sort by dy/dv
Below median	−0.286	−0.189	−0.043	−0.056	0.100	−0.013
Above median	0.255	0.185	0.155	0.158	−0.033	0.092

when aggregate shocks are above their median, the forecasting-performance improvement is 0.255 standard deviations higher than the average MSE gain for equipment, while it is 0.158 standard deviations higher for structures.

4. FINAL REMARKS

In this paper we derived and estimated a time series model of sectoral investment that builds from the realistic observation that lumpy adjustments play an important role in firms' investment behavior, but that allows for the empirically appealing feature that adjustments do not need to be of the same size across adjusting firms and for a firm over time.

Using a nonlinear *aggregate* time series procedure, we estimated the distributions of fixed adjustment costs faced by firms. The adjustment hazards implied by our estimates are nonconstant: they leave a significant range of inaction, and increase sharply thereafter. Depending upon the history of shocks, the estimated hazards have the potential to magnify or dampen the response of investment to aggregate shocks. The passivity of normal times is, occasionally, more than offset by the brisk response to large—current or accumulated—shocks. These nonlinearities clearly improve the aggregate performance of dynamic investment equations.

Both the microeconomic as well as the aggregate implications of the estimated model are largely consistent with the establishment level evidence presented by Caballero, Engel, and Haltiwanger (1995) for U.S. manufacturing (for the 1972–88 period). They found evidence of an increasing hazard for the range of disequilibria where establishments spent most of their time. More importantly, they also found an important role for the cross-sectional density of capital imbalances in explaining changes in the marginal response of aggregate investment to aggregate shocks.

In the process of assessing the contribution of the model, we found an important *forecasting* gain over simple linear models. In almost every year and sector, and particularly so for structures, the nonlinear model reduced the mean-squared-error by a substantial amount.

Beyond the current paper, there are three extensions and robustness issues worth mentioning at closing: First, the nonlinear model has nontrivial implications for forecasts' higher moments. Our preliminary exploration of this issue reveals that the standard deviation, skewness, and excess kurtosis of investment/capital ratios' forecasts, are highly correlated with the business cycle. These relations hint at a promising structural avenue to explore movements in forecasts' higher moments.[32]

Second, in the working paper version of this paper (Caballero and Engel (1994)) we allowed for serial correlation in the rate of growth of aggregate frictionless capital (the ν_{jt}'s) and found very little of it, which provides support for our i.i.d. assumption in the theory section. Interestingly, the linear (PAM)

[32] A good complement for nonstructural ARCH models, for example.

model we estimated left plenty of unexplained serial correlation, especially for equipment investment.

Finally, and also reported in Caballero and Engel (1994), we extended the theory and empirical sections to acknowledge the existence of continuous "maintenance" investment, which does not require paying sizeable adjustment costs. We found that while such an allowance was important to obtain a more realistic distribution of observed changes in capital and average investment rates (in particular, small changes account for a significant fraction of microeconomic investment changes), it did not diminish the role of microeconomic lumpiness in accounting for the dynamic aspects of aggregate investment.

Dept. of Economics, MIT, E52-252a, 50 Memorial Drive, Cambridge, MA 02139, U.S.A.; caball@mit.edu; http://web.mit.edu/caball/www

and

Center for Applied Economics (CEA), Dept. of Industrial Engineering, University of Chile, República 701, Santiago, Chile; eengel@dii.uchile.cl; http://www.dii.uchile.cl/cea/eengel.htm

Manuscript received October, 1994; final revision received June, 1998.

APPENDIX A.: DYNAMIC OPTIMIZATION

In this appendix we study the firm's stochastic dynamic optimization problem. In Section 1 we show existence and uniqueness of the solution to the firm's Bellman equation. In Section 2 we study the main properties of the optimal policy function.

1. EXISTENCE AND UNIQUENESS

From the main text it follows that the Bellman equation—for the firm's value function normalized by frictionless capital—is given by

$$v^*(z,\omega) = \max_i \left\{ \pi(z+i) - \omega\xi e^{\beta z}\{i \neq 0\} + \psi \iint e^{-\Delta z} v^*(z+i+\Delta z, \omega')\, dF(\Delta z)\, dG(\omega') \right\},$$

with $\{i \neq 0\}$ denoting an indicator function that takes the value 1 when the firm adjusts its capital stock and zero otherwise.[33]

The operator associated with the above equation is not bounded from below. For this reason we add a term to $v^*(z,\omega)$ that does not depend on the choice variable and therefore does not affect the firm's optimal choice, but does bound the corresponding operator:

$$\tilde{v}(z,\omega) = v^*(z,\omega) + \xi\omega e^{\beta z}. \tag{21}$$

Substituting (21) into the expression above, using the expression for $\pi(z)$ derived in Section 2 of the main text, and performing some straightforward (but tedious) calculations, leads to the Bellman

[33] All terms that are not explicitly defined in what follows were defined in Section 2 of the main text.

equation for $\tilde{v}(z, \omega)$:

$$(22)\qquad \tilde{v}(z,\omega) = \max\left\{ c_1(\omega)e^{\beta z} - c_2 e^{z} + \psi \iint e^{-\Delta z}\tilde{v}(z+\Delta z, \omega')\, dF(\Delta z)\, dG(\omega'),\right.$$
$$\left. \max_u \left[c_3 e^{\beta u} - c_2 e^{u} + \psi \iint e^{-\Delta z}\tilde{v}(u+\Delta z, \omega')\, dF(\Delta z)\, dG(\omega') \right] \right\},$$

with

$$c_1(\omega) = \xi(1 + \omega - \psi \mathrm{E}[e^{\beta \Delta z}]\,\mu_\omega),$$
$$c_2 = \xi\beta,$$
$$c_3 = \xi(1 - \psi \mathrm{E}[e^{\beta \Delta z}]\,\mu_\omega),$$

where μ_ω denotes the mean of the distribution of the adjustment cost factor $G(\omega)$.

To show that (22) has a unique solution we make the following three assumptions and prove the following two lemmas. The assumptions hold throughout the remainder of this appendix.

ASSUMPTION 1: *The adjustment cost factor, ω, is bounded from above by $\bar{\omega} < +\infty$ and from below by 0.*

ASSUMPTION 2: $\psi \mathrm{E}[e^{-\Delta z}] = \psi \int e^{-\Delta z} dF(\Delta z) < 1$.[34]

ASSUMPTION 3: $\psi \mu_\omega \mathrm{E}[e^{\beta \Delta z}] = \psi \mu_\omega \int e^{\beta \Delta z} dF(\Delta z) < 1$, *where μ_ω denotes the mean of the distribution of ω.*

LEMMA A1: *Let $f(z) = k_1 e^{\beta z} - k_2 e^{z}$, with $0 < \beta < 1$ and $k_1, k_2 > 0$. Denote $z_M = \log(k_1 \beta / k_2)/(1-\beta)$. Then $f(z)$ is increasing for $z < z_M$, decreasing for $z > z_M$, and attains its maximum value, which is equal to $(1-\beta)k_1^{1/(1-\beta)}(\beta/k_2)^{\beta/(1-\beta)}$, at $z = z_M$.*

PROOF: Elementary calculus. *Q.E.D.*

LEMMA A2: *Consider the operator T defined by posing $(T\tilde{v})(z, \omega)$ equal to the right-hand side of (22). This operator is defined on the set $\mathscr{B}$ of all real-valued, bounded, continuous functions with domain $\mathbb{R} \times [0, \bar{\omega}]$.*

Then T: (i) preserves boundedness; (ii) preserves continuity; and (iii) satisfies Blackwell's conditions.

PROOF: (i) Consider $u \in \mathscr{B}$, bounded from below by $\underline{u}$ and from above by $\bar{u}$. Then $(Tu)(z, \omega)$ is bounded from above, since

$$(Tu)(z,\omega) \le \psi \mathrm{E}[e^{-\Delta z}]\bar{u} + \max\left\{ c_1(\omega)e^{\beta z} - c_2 e^{z}, \max_w [c_3 e^{\beta w} - c_2 e^{w}] \right\}$$
$$\le \psi \mathrm{E}[e^{-\Delta z}]\bar{u} + \max_w [c_1(\omega)e^{\beta w} - c_2 e^{w}]$$
$$\le \psi \mathrm{E}[e^{-\Delta z}]\bar{u} + \xi(1-\beta)(1 + \bar{\omega} - \psi \mathrm{E}[e^{\beta \Delta z}]\,\mu_\omega)^{1/(1-\beta)},$$

where the second inequality follows from the fact that, for all $\omega \ge 0$, $c_3 < c_1(\omega)$, and the last inequality follows from Lemma A1.

[34] In the main text we assume that Δz follows a Normal distribution, with mean μ and variance σ^2. Then this condition is equivalent to $\mu > \frac{1}{2}\sigma^2 + \log(\psi)$, which for $r + \delta \ll 1$ corresponds approximately to $\mu > \frac{1}{2}\sigma^2 - r - \delta$. Thus, for the set of parameters we use in the empirical section, a sufficient condition is that $\mu > -0.15$ for equipment and $\mu > -0.10$ for structures. Both conditions can be expected to hold.

That $(Tu)(z, \omega)$ is bounded from below follows from

$$(Tu)(z, \omega) \geq \psi \mathrm{E}[e^{-\Delta z}]\underline{u} + \max\left\{c_1(\omega)e^{\beta z} - c_2 e^{z}, \max_{w}[c_3 e^{\beta w} - c_2 e^{w}]\right\}$$

$$\geq \psi \mathrm{E}[e^{-\Delta z}]\underline{u} + \max_{w}[c_3 e^{\beta w} - c_2 e^{w}]$$

$$= \psi \mathrm{E}[e^{-\Delta z}]\underline{u} + \xi(1-\beta)(1-\psi \mathrm{E}[e^{\beta \Delta z}]\mu_{\omega})^{1/(1-\beta)},$$

where we used Lemma A1 in the last step.

(ii) To show that the function (Tu) is continuous for $u \in \mathscr{B}$, we note that from (22) it follows that $(Tu)(z, \omega)$ is the maximum of two functions; the first inherits continuity from $u(z, \omega)$ and the second is constant. It follows that (Tu) is continuous.

(iii) To show that T satisfies Blackwell's conditions, we first note that if $u_1, u_2 \in \mathscr{B}$, and $u_1(z, \omega) \leq u_2(z, \omega)$ for all z and ω, then the expected value of any positive random variable, in particular $e^{-\Delta z}$, preserves the above inequality. Thus

$$(Tu_1)(z, \omega) \leq (Tu_2)(z, \omega).$$

A straightforward calculation shows that, for any $u \in \mathscr{B}$ and any constant a:

$$(T[u + a])(z, \omega) = (Tu)(z, \omega) + \psi \mathrm{E}[e^{-\Delta z}]a.$$

The second Blackwell condition follows from Assumption 2. *Q.E.D.*

PROPOSITION A1: *Equation* (22) *has exactly one solution* (*and this solution belongs to* $\mathscr{B}$).

PROOF: It follows from Lemma A2 that T defines a contraction mapping on the metric space $\mathscr{B}$ (normed with the sup-norm). The modulus of the contraction mapping is $\psi \mathrm{E}[e^{-\Delta z}]$. Existence and uniqueness of a solution to (22) now follows from the Continuous Mapping Theorem (see, e.g., Theorem 3.2 in Stokey, Lucas, and Prescott (1989)). *Q.E.D.*

2. PROPERTIES OF THE OPTIMAL POLICY

We define the following functions related to the solution of the Bellman equation, $\bar{v}(z, \omega)$, considered in the preceding section:

$$I(z) = \psi \iint e^{-\Delta z} \bar{v}(z + \Delta z, \omega)\, dF(\Delta z)\, dG(\omega), \tag{23}$$

$$J(z) = c_3 e^{\beta z} - c_2 e^{z} + I(z). \tag{24}$$

LEMMA A3: *The function* $J(z)$ *is bounded from above, i.e.,* $\sup_z J(z)$ *is finite. We denote this supremum by* $J_{\max}$.

PROOF: Since $\bar{v}(z, \omega)$ satisfies Bellman's equation, we have

$$\bar{v}(z, \omega) = \max\left\{\xi \omega e^{\beta z} + J(z), \max_{u} J(u)\right\}. \tag{25}$$

If $\max_z J(z)$ were not finite, we would have that $\bar{v}(z, \omega)$ is not bounded, contradicting Proposition A1. It follows that $\sup_z J(z)$ is finite. *Q.E.D.*

LEMMA A4: *The function* $J(z)$ *satisfies*

$$\lim_{z \to -\infty} J(z) = J_{\max}\psi \mathrm{E}[e^{-\Delta z}], \tag{26}$$

$$\lim_{z \to +\infty} J(z)e^{-z} = -c_2. \tag{27}$$

PROOF: From (25) it follows that

$$\lim_{z \to -\infty} \tilde{V}(z, \omega) = J_{\max},$$

which, from (23), implies that

$$\lim_{z \to -\infty} I(z) = J_{\max} \psi \mathrm{E}[e^{-\Delta z}].$$

Expression (26) now follows from (24).

Expression (27) holds because of (24) and the fact that $I(z)$ inherits boundedness from $\tilde{V}(z, \omega)$. *Q.E.D.*

The next result establishes the existence of a curve in (z, ω)-space partitioning this space into two regions: one where firms adjust their capital stock and another where they remain inactive.

PROPOSITION A2: *Define*

$$\Omega(z) = \xi^{-1}(J_{\max} - I(z))e^{-\beta z}. \tag{28}$$

Then firms adjust when their current adjustment cost factor, ω, is smaller than $\Omega(z)$, and remain inactive when $\omega > \Omega(z)$.[35]

PROOF: By equating both terms on the right-hand side of (25) we obtain

$$\xi\Omega(z) = (J_{\max} - I(z))e^{-\beta z}. \tag{29}$$

The inequalities that hold for ω larger and smaller than $\Omega(z)$ follow trivially. *Q.E.D.*

PROPOSITION A3: *The function $\Omega(z)$ is analytic on the real line, and therefore has derivatives of all order.*

PROOF: From the definition of $\Omega(z)$ and $I(z)$ (see (28) and (24)) we have that it suffices to show that $I(z)$ is analytic. To do this, we note that $I(z)$ may be written as the convolution of a normal density and a continuous, bounded function:

$$I(z) = \tilde{\psi} \int K(z + \Delta z)\, d\tilde{F}(\Delta z),$$

with

$$K(z) = \int \tilde{V}(z, \omega)\, dG(\omega),$$

$$\tilde{\psi} = \psi e^{\sigma^2/2 - \mu},$$

and $d\tilde{F}(\Delta z)$ is a normal density with mean $\tilde{\mu} \equiv \mu - \sigma^2$ and variance σ^2 (where $dF(\Delta z)$ is normal with mean μ and variance σ^2).

That the convolution of a normal density and an integrable (in particular, a bounded, continuous) function is analytic, follows from a well known property of the exponential family of distributions (see Theorem 9 on p. 59 in Lehmann (1986)). *Q.E.D.*

[35] When $\omega = \Omega(z)$ firms are indifferent between adjusting and not adjusting.

PROPOSITION A4: *As z tends to $-\infty$:*[36]

$$\xi\Omega(z) \sim J_{\max}\{1 - \psi \mathrm{E}[e^{-\Delta z}]\}e^{-\beta z}, \tag{30}$$

and as z tends to ∞:

$$\xi\Omega(z) \sim c_2 e^{(1-\beta)z}. \tag{31}$$

It follows that

$$\lim_{|z|\to\infty} \Omega(z) = \infty. \tag{32}$$

PROOF: Expression (30) follows from (28) and Lemma A4. Expression (31) follows from (27). Then $\lim_{|z|\to\infty} \Omega(z) = \infty$ follows from (30), (31), Assumption 2 and the fact that $\beta < 1$. *Q.E.D.*

PROPOSITION A5: *The set $\mathscr{C}$ of $z \in \mathbb{R}$ such that $J(z) = J_{\max}$ is a nonempty set with a finite number of points.*

PROOF: Continuity of $J(z)$ (it is analytic; see the proof of Proposition A3) and Lemma A4 combined with Assumption 2 ensure the existence of a bounded, closed set $\mathscr{C}$ within which $J(z)$ attains its maximum. Continuity of $J(z)$ on the compact set $\mathscr{C}$ ensures that the maximum is indeed attained and therefore $\mathscr{C}$ is nonempty. Finally, since $J(z)$ is analytic, we have that its maxima are isolated, thus showing that $\mathscr{C}$ contains a finite number of elements. *Q.E.D.*

PROPOSITION A6: *When adjusting its capital stock, a firm's optimal choice of z is any element in $\mathscr{C}$. Thus its disequilibrium after adjusting does not depend on its disequilibrium before adjusting.*

PROOF: The result follows from the fact that when the maximum between both terms in the right-hand side expression of (25) is attained at the second term, this expression does not depend on z. *Q.E.D.*

All calculations of $\Omega(z)$ performed while estimating the distribution of the adjustment cost factor (see Section 3.3 in the main text and Appendix C) led to a set $\mathscr{C}$ with a unique element, c, and a function $\Omega(z)$ that is decreasing to the left of $z = c$ and increasing to the right of $z = c$, therefore implying an optimal policy of the (L, c, U) type. Yet we have been unable to show formally that $\mathscr{C}$ has one element (i.e., a unique return point), and also have not shown formally that, conditional on ω, the firm's optimal policy is of the (L, c, U) type. As the following proposition shows, however, we can prove that the latter holds in a neighborhood of a return point.

PROPOSITION A7: *For (ω, z) in a neighborhood of $(0, c)$, the optimal policy is of the (L, c, U) type.*

PROOF: From Proposition A5 it follows that there exists a neighborhood $\mathscr{V}$ of $z = c$ such that $J(z)$ is decreasing to the left of $z = c$ and increasing to the right of $z = c$. It then follows from equation (28) that $\Omega(z)$ is decreasing to the left of $z = c$ and increasing to the right of $z = c$. Thus, for all $\omega \leq \max_{z \in \mathscr{C}}(J_{\max} - J(z))e^{-\beta z}$ and $z \in \mathscr{V}$ we have that the optimal policy, conditional on the current adjustment cost factor, is of the (L, c, U) type. *Q.E.D.*

APPENDIX B: AGGREGATION

This appendix is divided into three sections. In Section 1 we establish the exact expression for aggregate investment and present the results of simulations to assess the quality of the approximation we use. In Section 2 we study the main properties of the adjustment hazard. In Section 3 we characterize the average cross-section of firm deviations.

[36] We write $a(z) \sim b(z)$ as z tends to c if $\lim_{z\to c}(a(z)/b(z)) = 1$.

The following operators, defined on the set of probability measures on the real line, are used in Sections 1 and 3 of this Appendix and in Appendix C.

- SECTORAL (AGGREGATE) SHOCK, $\mathscr{A}(v)$: shifts a cross-section by v.
- ADJUSTMENT SHOCK, $\mathscr{H}(\theta)$: applies adjustments determined by the adjustment function characterized by the parameter vector θ.
- IDIOSYNCRATIC SHOCK, $\mathscr{I}(\sigma^2)$: convolves the probability measure with a Normal density with zero mean and variance σ^2.
- FULL CYCLE OF SHOCKS, $\mathscr{T}(v+\delta, \theta, \sigma)$: equal to the combination of the three shocks defined above,[37] i.e., to $\mathscr{I}(\sigma)\mathscr{H}(\theta)\mathscr{A}(v+\delta)$. Equation (17) in the main text gives an explicit expression for the cross-section that results from applying $\mathscr{T}$ to $f(x, t-1)$.
- AGGREGATE INVESTMENT FUNCTIONAL, $\mathscr{Y}$: Assigns to a cross-section the average investment rate that results after adjustments take place (see equation (12) in the main text).

1. AGGREGATE INVESTMENT

In this section we derive the exact expression for aggregate investment, of which equation (12) is an approximation. We also assess the quality of this approximation.

LEMMA B1: *We introduce the following notation:*

- $\xi_{i,t}$: *disequilibrium immediately before period t adjustment of firm i;*
- $r_{i,t}$: *number of periods, as of time t, since firm i last adjusted;*[38]
- $l_{i,t}$: *last time firm i adjusted (equal to* $t - r_{i,t}$*);*
- $k^*_{i,l_{i,t}}$: *desired level of (log) capital of firm i immediately before its last adjustment took place;*
- $k_{i,l_{i,t}}$: *firm i's (log) capital stock immediately after the last time it adjusted. Note that* $k_{i,l_{i,t}} = k^*_{i,l_{i,t}} + c$ *and that, as changes in capital since time* $l_{i,t}$ *have only reflected depreciation, this quantity completely determines the current capital stock.*

Then, conditional on $l_{i,t}$ *we have that* $\xi_{i,t}$ *and* $k_{i,l_{i,t}}$ *are independent. That is, conditional on when the firm last adjusted, its current disequilibrium and current capital stock are independent.*

PROOF: We have that, since $\xi_{i,t} = k^*_{i,t} - k^*_{i,l_{i,t}}$, it depends on shocks (aggregate and idiosyncratic) that took place during periods $t > l_{i,t}$. On the other hand, $k_{i,l_{i,t}}$ depends only on shocks that took place at $t \leq l_{i,t}$. Since shocks are i.i.d., it follows that, conditional on $l_{i,t}$, both quantities are independent. *Q.E.D.*

PROPOSITION B1: *Denote*

- $\pi_t(r)$: *fraction of plants, as of time t, that last adjusted r periods ago;*
- $\bar{K}_t(x|r)$: *average capital stock of plants with disequilibrium x at time t, that last adjusted r periods ago;*
- $\bar{K}_t(r)$: *average capital stock of all firms that last adjusted r periods ago; thus*

$$K_t^A = \sum_r \pi_t(r)\bar{K}_t(r);$$

- $\bar{I}_t(r)$: *average investment, at time t, of those that last adjusted r periods ago; thus*

$$I_t^A = \sum_r \pi_t(r)\bar{I}_t(r);$$

- $f_t^*(x|r)$: *cross-section, at time t, of plants that adjusted r periods ago; denoting*

$$\mathscr{T}_t = \mathscr{H}(\theta)\mathscr{A}(v_t+\delta)\mathscr{I}(\sigma_\varepsilon^2),$$

$$\mathscr{U} = \mathscr{A}(v_t+\delta)\mathscr{I}(\sigma_\varepsilon^2),$$

and a mass point at 0 by δ_0*, we have that*

$$f_t^*(x|r) = \mathscr{U}\mathscr{T}_{t-1}\cdots\mathscr{T}_{t-r}(\delta_0).$$

[37] Incorporating depreciation.

[38] Possible values are: $1, 2, 3, \ldots$.

Then

$$I_t^A = \sum_r \pi_t(r)\bar{K}_t(r)\int (e^{-x}-1)\,A(x)f_t(x|r)\,dx.$$

PROOF:

$$\begin{aligned} I_t^A &= \sum_r \pi_t(r)\bar{I}_t(r) \\ &= \sum_r \pi_t(r)\int (e^{-x}-1)\,A(x)\bar{K}_t(x|r)f_t^*(x|r)\,dx \\ &= \sum_r \pi_t(r)\bar{K}_t(r)\int (e^{-x}-1)\,A(x)f_t^*(x|r)\,dx, \end{aligned}$$

where we used Lemma B1 in the last step. *Q.E.D.*

It follows that, to calculate the exact expression for the aggregate investment/capital ratio, we need to keep track of a sufficiently large number of conditional cross-sections, $f_t(x|r)$, and the size distribution of cohorts, $\pi_t(r)$. Computationally, this is substantially more burdensome than the approximation we used. We show numerically, however, that this approximation mostly affects nuisance (secondary) parameters.

We note that if $\hat{\gamma}$ denotes the estimated values of the main parameters γ,[39] then a linear transformation of the shocks used when calculating the likelihood still leads to the same main parameters.[40] It follows that a good measure of the quality of the approximation we use is to determine the extent to which the exact expression comes close to our approximation when we allow for a linear change in the aggregate shocks that determine exact aggregate investment.

To implement this idea we consider 50,000 firms with initial capital stock equal to one and disequilibrium x equal to zero. All firms belong to the same sector. We simulate the evolution of these firms during 75 time periods, with parameters given by our estimated structural model. We keep track of the aggregate shocks (denoted by v_t) and our approximation to aggregate investment (denoted by y_t^A). Next we rerun the whole process with rescaled aggregate shocks ($w_t = a + bv_t$), this time keeping track of the exact expression for aggregate investment ($y_t^T(a,b)$). We find the values of a and b for which the series $y_t^T(a,b)$ is closest to y_t^A. To measure proximity between both series we consider two criteria, both of them applied to the last 45 observations of both series:

$$R_1^2 = 1 - \frac{\mathrm{MS}(y_t^T(a,b) - y_t^A)}{\mathrm{Var}(y_t^A)},$$

$$R^2 = 1 - \frac{\mathrm{Var}(y_t^T(a,b) - y_t^A)}{\mathrm{Var}(y_t^A)},$$

where $\mathrm{MS}(y_t)$ denotes the average of the squares of the corresponding series and $\mathrm{var}(y_t)$ its variance. These measures capture the fit of a regression of y_t^T on y_t^A, differing in whether they allow or not for an additive constant term.[41] Table VIII shows the results of our simulations. It is apparent that the excellent quality of the fit we obtained justifies approximating aggregate investment by (12).

[39] There are three of these parameters in both estimation approaches we use. In both cases we have a free constant. The two remaining parameters characterize the distribution of adjustment costs in the structural case and the adjustment hazard in the semistructural case.

[40] This follows from the expression derived for the likelihood (see Appendix C).

[41] It is arguable which criterion is more adequate in our case. On one hand, we allow for an additive constant term when estimating our models; on the other hand, it did not vary across sectors.

TABLE VIII

ASSESSING THE APPROXIMATION FOR AGGREGATE INVESTMENT

	Equipment	Structures
$\hat{a}$	0.015	0.003
$\hat{b}$	1.054	1.080
R_1^2	0.980	0.995
$\hat{a}$	−0.019	−0.005
$\hat{b}$	1.225	1.145
R^2	0.996	0.999

2. ADJUSTMENT HAZARD

In this section we study the properties of the adjustment hazard:

$$A(x) = G(\Omega(x + c)),$$

where c is a fixed element (say the smallest one) in the set $\mathscr{C}$ characterized in Proposition A5.

ASSUMPTION 4: *The distribution function G has a continuous density $g(\omega)$ with support $[0, \bar{\omega}]$, $\bar{\omega} < \infty$.*

PROPOSITION B2: *Under Assumptions* 1, 2, *and* 3 *made in Appendix A we have that the adjustment hazard satisfies*:

(a) $\lim_{|x| \to \infty} A(x) = 1$. *Furthermore, there exists a positive constant M such that for $|x| > M$ we have $A(x) = 1$.*

(b) *$A(x)$ is differentiable at all x and*

$$A'(x) = g(\Omega(x + c))\,\Omega'(x + c).$$

PROOF: (a) This follows immediately from Proposition A4 and Assumption 1.

(b) It follows from the fact that $A(x)$ is the composition of two differentiable functions and therefore differentiable (see Proposition A3). *Q.E.D.*

3. INVARIANT DISTRIBUTION

Due to the presence of aggregate shocks, the distribution of disequilibria that determines aggregate investment (see Section 2.3 of the main text) has no invariant distribution. In Caballero and Engel (1992b) we establish that, in a well defined sense, the average over all possible trajectories of aggregate shocks of the cross-section of deviations is equal to the invariant distribution faced by an individual firm.[42] In this section we show that such a distribution exists and that convergence toward it takes place at an exponential rate.

The following operators, all of which are defined on the set $\mathscr{F}$ of probability densities on the real line, will be useful throughout this section.

• We let $\mathscr{T}_1 \equiv \mathscr{N}(\mu + \delta)\mathscr{T}(\sigma^2)\mathscr{R}(\theta)$ and $\mathscr{T}_2 \equiv \mathscr{R}(\theta)\mathscr{N}(\mu + \delta)\mathscr{T}(\sigma^2)$, where μ denotes the mean of aggregate shocks, δ the depreciation rate, σ^2 the sum of the variance of aggregate and idiosyncratic shocks, and θ the set of parameters characterizing the function $A(x) = G(\Omega(x + c))$. The operators $\mathscr{N}$, $\mathscr{R}$, and $\mathscr{T}$ were defined earlier in this Appendix. For any integer $n > 1$ we denote by $\mathscr{T}_i^n$ the n-fold composition of $\mathscr{T}_i$, $i = 1, 2$.

[42] Thus the aggregate shock is constant and equal to μ and the idiosyncratic shock is Normal with zero mean and variance $\sigma^2 \equiv \sigma_A^2 + \sigma_\epsilon^2$.

• We denote by $\mathscr{V}(f)$ the function that associates to f the fraction of firms that adjust when applying $\mathscr{T}_2$.

• We denote by $\mathscr{G}_{n-1}$ the operator that associates to an initial density f the cross-section of firms that have not adjusted after $(n-1)$ shocks, normalized to one.

We assume throughout that shocks are Normal. The following lemma is needed to establish the main result of this section.

LEMMA B2: *Given a cross-section f_0, let $f_n \equiv \mathscr{T}_2^n(f_0)$. Denote by $\pi_n(f_0)$ the probability that a particular firm adjusts at time n, conditional on not having adjusted during the first $n-1$ periods, and denote by $\tau_n(f_0)$ the fraction of firms that do not adjust during the first n periods. Then there exists a constant $\alpha \in (0,1)$, common to all initial cross-sections, such that*:

$$\pi_n(f_0) \geq 1-\alpha, \tag{33}$$

$$\tau_n(f_0) \leq \alpha^n. \tag{34}$$

PROOF: Denote by $\mathscr{F}_{n-1}$ the set of all densities that may represent those firms that have not adjusted after $(n-1)$ shocks. Since $\mathscr{F}_{n-1} \subset \mathscr{F}$ and $\pi_n(f) = \pi_1(\mathscr{G}_{n-1}(f))$, we have that

$$\begin{aligned}\inf_{f\in\mathscr{F}_{n-1}} \pi_n(f) &= \inf_{g\in\mathscr{G}_{n-1}} \mathscr{V}(g)\\ &\geq \inf_{f\in\mathscr{F}} \mathscr{V}(f)\\ &= \inf_{f\in\mathscr{F}} \pi_1(f).\end{aligned}$$

Hence a lower bound for π_1 also is a lower bound for π_n, $n>1$.

Next note that it follows from Proposition B2 that there exists a constant M such that $\Lambda(x)=1$ for $|x|>M$. Denote by $\Lambda^*(x)$ the adjustment hazard that is equal to one when $|x|>M$ and equal to zero elsewhere. It is easy to see that both $\tau_n(f_0)$ and $1-\pi_1(f_0)$ corresponding to this adjustment hazard are larger than or equal to the corresponding quantities for the original hazard. Thus it suffices to prove (33) and (34) for $\Lambda^*(x)$ and, in the case of (33), for $n=1$.

Applying the operator $\mathscr{T}_2$ to a mass point at x, δ_x, we have that the value of x for which the fraction of firms that does not adjust is largest, is the value such that the distribution before adjustment is normal with zero mean and variance σ^2,[43] and this fraction is $\alpha = 2\Phi(M/\sigma)-1$, with $0<\alpha<1$ and Φ denoting the c.d.f. of a standard Normal. It follows that for any density $f(x)$ the fraction that does not adjust after applying $\mathscr{T}_2$ is bounded from above by α. Hence $\pi_1 \geq 1-\alpha$. Finally, given any cross-section $f_0(x)$, we have that $\tau_n(f_0) = \prod_{k=1}^n (1-\pi_k(f_0))$, which is bounded from above by α^n. *Q.E.D.*

PROPOSITION B3: *Given an arbitrary initial cross-section, f_0, let $f_k = \mathscr{T}_1^k(f_0)$. Also, define the sequence (f_k^o) as above, but for the particular case where f_0 is a mass point at $x=0$.*

Let π_k^0 denote the probability that a firm that starts off at $x=0$ adjusts at time n, conditional on not having adjusted during the first $n-1$ periods.

Denote

$$p_i = \frac{\prod_{l=0}^{i-1}(1-\pi_l^0)}{1+\sum_{j\geq 1}[\prod_{l=0}^{j-1}(1-\pi_l^0)]},$$

and define $f^e(x) = \sum_{i\geq 0} p_i f_i^o(x)$. Then f_n converges to f^e in the variation distance and convergence takes place at an exponential rate.

[43] Since $\Phi((M-x)/\sigma) - \Phi((-M-x)/\sigma)$ is maximized at $x=0$.

PROOF: We consider first the case where $f_0 = \delta_0$.

At any moment in time, f_n^o can be partitioned into groups of firms that last adjusted the same number of periods ago. Hence f_n^o is a convex combination of $f_0^o, f_1^o, \ldots, f_{n-1}^o$.

The weights on the above densities can be determined by the one-to-one correspondence with the Markov process with state space $\mathcal{S} = \{0, 1, 2, 3, \ldots\}$, and transition kernel:

$$P(s, t) = (1 - \pi_s)\delta_{s,t-1} + \pi_s \delta_{s,0},$$

where $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. State s leads to $s + 1$ (not adjusting) or to 0 (adjusting). The corresponding probabilities are $1 - \pi_s$ and π_s.

It follows from Lemma B2 that the above process satisfies Condition M in Stokey, Lucas, and Prescott (1989, p. 348). Indeed, with the notation of these authors we have that there exists $N = 1$ and $\epsilon = \max(\alpha, 1 - \alpha) > 0$, with α defined in the preceding lemma, such that for all subsets A of S:

$$\max(P_1(s, A), P_1(s, A^C)) \geq \epsilon,$$

since

$$\max(P_1(s, A), P_1(s, A^C)) = \begin{cases} 1 & \text{if } \{0, s+1\} \subset A \text{ or } \{0, s+1\} \subset A^C, \\ \max(\pi_s, 1 - \pi_s) & \text{otherwise,} \end{cases}$$

and from Lemma B2 we have $\max(\pi_s, 1 - \pi_s) = \epsilon$ for all s.

Hence Theorem 11.12 in Stokey, Lucas, and Prescott (1989, p. 350) implies that there exists an invariant distribution, f^e, and a constant $G \equiv (1 - \epsilon) < 1$, such that

$$\|f_n^o - f^e\| \leq G^n \|f_0^o - f^e\|. \tag{35}$$

Furthermore, f^e is the unique fixed point of the Markov operator. The latter and a straightforward calculation show that $f^e = \Sigma_i p_i f_i^o$.

Extending this results to the general case where f_0 can be arbitrary is straightforward. All firms eventually adjust and, once they adjust, the previous case applies (since they adjust to $x = 0$).

Given an arbitrary f_0 we may write

$$\mathcal{T}^n(f_0) = (1 - \tau_{n/2})\mathcal{T}^{n/2}(f_n^*) + \tau_{n/2}(g_n),$$

where f_n^* is a convex combination of $f_0^o, f_1^o, \ldots, f_{n/2}^o$, g_n could be any cross-section, and τ_n is defined in Lemma B2.

From Lemma B2 and the fact that the variation distance is bounded from above by 1 we have that there exists $\alpha \in (0, 1)$ such that

$$\begin{aligned} \|\mathcal{T}^n f_0 - f^e\| &\leq (1 - \tau_{n/2})\|\mathcal{T}^{n/2}(\delta_0) - f^e\| + \tau_{n/2} \\ &\leq G^{n/2}\|f_0 - f^e\| + \alpha^{n/2}. \end{aligned}$$

Convergence follows by letting n tend to infinity; the rate at which convergence takes place is geometric, being at least as fast as $\max(\sqrt{G}, \sqrt{\alpha})$. *Q.E.D.*

APPENDIX C.: ECONOMETRICS

This appendix is divided into two sections. In Section 1 we derive the likelihood function and sketch the general approach used for calculating this function at given parameter values. In Section 2 we describe implementation details for the semistructural (Section 2.1) and structural (Section 2.2) cases.

I. CALCULATING THE LIKELIHOOD FUNCTION

1.1. *An Expression for the Likelihood*

The sources of randomness ("error terms") are the sectoral (aggregate) shocks, i.e., the v_{it}'s, where $i = 1, \ldots, I$ and $t = 1, \ldots, T$.[44] We assume that sectoral shocks are Normal and independent over time, and denote the mean and variance of these shocks in sector i by μ_i and c_{ii}, respectively. The (column) vector of sector i's aggregate shocks is denoted by V_i; V denotes the column vector with V_1 followed by V_2 and so on, and $\mu_V \equiv \mathrm{E}[V]$. We allow for contemporaneous correlation among shocks from different sectors; the matrix $C = [c_{ij}]$ denotes the corresponding covariance matrix (which does not vary over time).

Standard change of variable calculations lead to the following expression for minus the log likelihood:

$$-\text{log. lik.} = \text{const} + \sum_{i,t} \log\left|\frac{\partial y_{it}}{\partial v_{it}}\right| + \frac{T}{2}\log|C| + \frac{1}{2}(V-\mu_V)'(C^{-1}\otimes I_T)(V-\mu_V),$$

where $A \otimes B$ denotes the Kronecker product of matrices A and B and $y_{it} = I_{it}/K_{it}$. Concentrating the likelihood with respect to C and μ_V leads to

$$-\text{log. lik} = \text{const} + \sum_{i,t} \log\left|\frac{\partial y_{it}}{\partial v_{it}}\right| + \frac{T}{2}\log\left|\frac{(V-\hat{\mu}_V)(V-\hat{\mu}_V)'}{T}\right|, \tag{36}$$

where $\hat{\mu}_V$ corresponds to the vector of sample means. That the Jacobian is well defined follows from Proposition B2.

Calculating the likelihood in (36) requires calculating the sectoral shocks (the v_{it}'s) and the corresponding partial derivatives (the $\partial y_{it}/\partial v_{it}$'s). Next we show that, conditional on the initial cross-section and the set of parameter values, the relation between sectoral shocks and sectoral investment rates is invertible. Our proof is constructive: it describes how the sectoral shocks are actually calculated for given parameter values.

1.2. *Calculating the Components of the Likelihood*

Suppose we know the cross-sections of disequilibria in every sector at time $t = 0$. It follows from the aggregate dynamics in our model that the observed capital-investment ratio in the ith sector during period t is determined by the aggregate shocks in the first t periods (see equations (15), (16), and (18)):

$$y_{it} = y_{it}(v_{i1}, v_{i2}, \ldots, v_{it}, f_i(\cdot, 0)) \qquad (i = 1, \ldots, I). \tag{37}$$

Furthermore, aggregate investment is a function of only the current aggregate shock and the cross-section prior to this shock:

$$y_{it} = y_{it}(f_i(\cdot, t-1), v_{it}) \tag{38}$$

$$= \int \phi(v_{it} + \delta - x)\, \Lambda(x - v_{it} - \delta) f_i(x, t-1)\, dx, \tag{39}$$

[44] By working with a continuum of firms we have that, despite the presence of idiosyncratic and adjustment shocks at the micro level, the only source of sectoral randomness are aggregate shocks. That is, the cross-section that results after adjustments is uniquely determined by the adjustment function and the cross-section prior to adjustments. Also, the cross-section that results after the idiosyncratic shocks is the convolution of the density (common across plants and sectors) from which these shocks are drawn with the cross-section prior to the shock. See equations (15) and (16) in the main text for the corresponding formulas.

where in our case $\phi(u) = e^u - 1$ but, more generally, in the derivation that follows $\phi(u)$ could be any smooth and strictly increasing function with $\phi(0) = 0$. The derivative of the above expression with respect to v_{it} evaluated at v is equal to

$$\frac{\partial y_{it}}{\partial v_{it}}(f_i(\cdot, t-1), v) = \int [\phi'(v+\delta-x)A(x-v-\delta) + \phi(v+\delta-x)A'(x-v-\delta)]f_i(x, t-1)\,dx.$$

Recalling that $\bar{f}(x,t)$ denotes the cross-section immediately after period t's sectoral (and depreciation) shocks, we have that

$$\frac{\partial y_{it}}{\partial v_{it}}(f_i(\cdot, t-1), v_{it}) = \int [\phi'(-x)A(x) - \phi(-x)A'(x)]\bar{f}_i(x, t-1)\,dx. \tag{40}$$

It follows from our assumptions on ϕ and the fact that $\Phi(x)$ is decreasing for negative x and increasing for positive x (the "increasing hazard" property) that the above derivative is strictly positive when $f_i(x, t-1)$ has support equal to the real line (as in our case). Thus v_{it} is uniquely determined from (38):

$$v_{it} = v_{it}(y_{it}, f_i(\cdot, t-1)) \tag{41}$$

$$= v_{it}(y_{it}, v_{i,t-1}, \ldots, v_{i,1}, f_i(\cdot, 0)), \tag{42}$$

and proceeding inductively we conclude that

$$v_{it} = v_{it}(y_{it}, y_{i,t-1}, \ldots, y_{i1}, f_i(\cdot, 0)).$$

It follows that, conditional on the initial cross-sections, the v_{it}'s are uniquely determined by the y_{it}'s.

1.3. *Initial Cross-sections*

The initial cross-section in sector i is set equal to the invariant probability measure of the unconditional process describing the evolution of disequilibria for an individual plant in that sector.[45] This is the cross-section obtained when averaging over all possible sample paths of aggregate shocks.[46]

Although this selection is arbitrary, we checked the robustness of our results by studying the convergence properties of the cross-sections distribution near our initial distribution. We compared the sequence of cross-sections used in our likelihood calculations with those obtained when we perturbed the mean of the invariant-initial distribution by one standard deviation of the average (across sectors) aggregate shocks. The Markov structure of our problem, combined with the contractionary features derived in Appendix A, ensure that for any given sequence of aggregate shocks, the distance between both cross-sections tends to zero over time with probability one; the issue is how fast one distribution converges to the other. Simulations showed that, for the parameter values considered, the distance between both sequences of cross-sections becomes negligible (variation distance less than 0.01) sometime between the second and third cross-section after the initial one.[47] For this reason, we discarded the first three observations for all series when calculating the likelihood.

[45] By unconditional, we mean that we do not condition on actual sectoral shocks. For this reason the variance of shocks relevant for this distribution is the sum of the variances of sectoral and idiosyncratic shocks. In Appendix B.1 we show that if F_0 denotes the probability measure describing a particular plant's deviation at time $t = 0$, and F_t the corresponding probability measure t periods later, then F_t converges in the variation distance to a distribution F^* which does not depend on the initial distribution F_0.

[46] See Caballero and Engel (1992b) for a proof.

[47] The parameter on which convergence depends most is the variance of idiosyncratic shocks; convergence is faster as this parameter becomes larger.

1.4. *Summary*

Given a set of parameter values, we calculate the likelihood in (36) as follows:

1. The initial cross-section of firms' disequilibria (one for each of the 21 two-digit manufacturing sectors considered) are set equal to the invariant distribution faced by an individual plant. These cross-sections are denoted $f_i(x,0)$; $i = 1,\ldots,21$.

2. For $t = 1$ to T:

(a) Solve (38) to find v_{it}, $i = 1,\ldots,21$.

(b) Calculate $\partial y_{it}/\partial v_{it}$ from (40).

(c) Determine the next set of cross-sections of disequilibria (the $f_i(\cdot,t)$'s) based upon (15) and (16).

The next section provides the details on exactly how every one of the steps above is conducted in both estimation approaches (semi-structural and structural).

2. Implementation

2.1. *Semi-structural Approach*

This approach estimates the adjustment rate function directly. We assume that the adjustment rate function is common across sectors and of the form:

$$A(x) = 1 - e^{-\lambda_0 - \lambda_2 x^2}, \tag{43}$$

with $\lambda_0 \geq 0$ and $\lambda_2 \geq 0$. We estimate three parameters (besides the mean and variance-covariance matrix of aggregate shocks): λ_0, λ_2, and an additive constant (common across sectors).

Estimating the Initial Cross-sections

In what follows, we do not make any assumptions about the mean of the aggregate shock. If we knew this mean, or could estimate it directly from the observed data, then we could determine the invariant density by calculating the invariant probability function of a standard-fixed Markov chain (see Caballero and Engel (1994) for details).

To compute the initial cross-section we proceed as follows. For sector i we let $g_i(x,0)$ denote a Normal density with zero mean and variance $\sigma_0^2 \equiv \sigma_\varepsilon^2 + c_{ii}$. We set c_{ii} equal to 0.035 for all sectors.[48] Given $g_i(x,\tau-1)$ we calculate $g_i(x,\tau)$ by first solving for v in

$$\mathscr{Y}[\mathscr{K}(\lambda_0,\lambda_2)\mathscr{A}(v_\tau+\delta)]g_i(x,\tau-1) = \bar{y}_i,$$

where $\bar{y}_i$ denotes the average capital-investment ratio of sector i. The solution is denoted by $v_{i\tau}$. Then we set

$$g_i(x,\tau) = \mathscr{F}(v_{i\tau}+\delta,\lambda_0,\lambda_2,\sigma_0)g_i(x,\tau-1).$$

As τ grows, $g_i(x,\tau)$ approaches the unconditional invariant density for an individual plant, and $v_{i\tau}$ approaches a constant consistent with the mean of sectoral investment/capital ratio. We use this density as the initial cross-section when calculating the likelihood. Simulations showed that using 30 iterations (for each sector) was sufficient for all practical purposes.

[48] There are two reasons for fixing c_{ii}: First, it avoids estimating an additional nonlinear parameter. Second, since we may expect that the variance of idiosyncratic shocks is significantly larger than the variance of aggregate shocks, the value of the latter is of little relevance when determining the adjustment function and the invariant density.

Calculating the Likelihood

The family of adjustment functions with which we work has the attractive property that the evolution of the cross-sections can be tracked efficiently using a convex combination of a small number of Normal densities, thus reducing computational time substantially. To see this, we show next that if we assume that $f_i(x, t-1)$ is a convex combination of N Normal densities, then $f_i(x,t)$ is a convex combination of $N+1$ densities. We also derive simple expressions to update the means, variances, and weights assigned to the Normal densities characterizing $f_i(x,t)$.

Consider first the case $N=1$ and assume $f_i(x, t-1)$ is Normal, with mean μ and variance σ^2. A simple but tedious calculation shows that solving (38) reduces to solving for v_{it} in

$$y_{it} = e^{c(v_{it}+\delta)} - 1 - \frac{\tau}{\sigma} e^{-d(v_{it}+\delta)+(\tau^2/\sigma^2)c(v_{it}+\delta)}, \tag{44}$$

where

$$\tau^2 = \frac{\sigma^2}{1+2\lambda_2\sigma^2},$$

$$c(v) = v - \mu + \frac{1}{2}\sigma^2,$$

$$d(v) = \lambda_0 + \frac{\tau^2}{\sigma^2}\lambda_2(v-\mu)^2.$$

The partial derivative in (40) is equal to

$$\frac{\partial y_{it}}{\partial v_{it}} = e^{c(v_{it}+\delta)} + \frac{\tau^3}{\sigma^3} e^{-d(v_{it}+\delta)} - 2\lambda_2(v_{it}+\delta-\mu)\frac{\tau^2}{\sigma^2}[e^{c(v_{it}+\delta)} - y_{it} - 1]. \tag{45}$$

It follows from equation (17) in the main text that the cross-section density after the tth period's sectoral (aggregate), hazard and idiosyncratic shocks, $f_i(\cdot, t)$, is a convex combination of two Normal densities, one of them with mean $\eta = (\mu - v_{it} - \delta)/(1+2\lambda_2\sigma^2)$ and variance $\tau^2 + \sigma_\epsilon^2$, and the other with zero mean and variance σ_ϵ^2. The former corresponds to those firms that did not adjust, the latter to those that adjusted their capital stock. The fraction of firms in the group that does not adjust is

$$\kappa = \frac{\tau}{\sigma}\exp\left(-\lambda_0 + \frac{(\mu - v_{it} - \delta)^2}{2\sigma^2}\left[\frac{\tau^2}{\sigma^2} - 1\right]\right). \tag{46}$$

In the more general case, where $f_i(x, t-1) = \sum_{k=1}^{N} \alpha_k f^k(x, t-1)$ is a convex combination of N Normal densities, v_{it} is obtained by solving an equation analogous to (44) with a linear combination of terms like the one on the right-hand side of that equation:

$$y_{it} = \sum_k \alpha_k \int \phi(v_{it} + \delta - x)\, A(x - v_{it} - \delta) f^k(x, t-1)\, dx.$$

The partial derivative is equal to a convex combination of terms like those in (45). We also have that $f_i(x,t)$ will be a convex combination of $N+1$ Normal densities. Each of these cross-sections corresponds to a specific cohort, grouping plants that have not adjusted for the same number of periods. The "older" cross-sections are more spread out than the "younger" ones and have lost mass monotonically due to the adjustment of their members. Simulations showed that keeping track of 30 densities is extremely conservative: the impact on aggregate investment of cohorts much older than 30 years is negligible. For this reason, in every period we merge the two oldest cohorts into one Normal density with mean and variance equal to those of the convex combination of the densities being merged.

2.2. *Structural Approach*

Instead of estimating the adjustment function directly, as in the semistructural case, here we estimate the parameters of the distribution of adjustment costs and obtain the adjustment function from the solution of the dynamic optimization problem described in Section 2.

The initial distribution is calculated in a way analogous to the semistructural case. Adjustment costs are drawn from a Gamma distribution:

$$G(\omega) = \frac{1}{\phi^p \Gamma(p)} \int_0^{\omega} \eta^{p-1} e^{-\eta/\phi} d\eta,$$

which has mean $\mu_\omega = p\phi$ and coefficient of variation $\mathrm{cv}_\omega = 1/\sqrt{p}$. Again, we estimate three parameters (besides the means and variance-covariance matrix of aggregate shocks): μ_ω, cv_ω, and an additive constant (common across sectors).

Adjustment Function

The adjustment function for a given set of parameters is obtained by solving numerically the stochastic dynamic optimization problem described in Section 2. For this purpose—but not when evaluating the likelihood from the shocks and Jacobian terms—we disregard sectoral differences in μ_A and σ_A, and assume the parameters that determine the adjustment function (μ_ω, cv_ω, σ_ε, r, δ, β, in addition to μ_A and σ_A) are common across sectors. This allows us to calculate only one adjustment function and use it for all sectors.

We use a grid of 800 equally spaced points on the interval $[-3.5, 3.5]$ to determine the value function via value iteration.[49] The corresponding steps, for which extensive simulations showed that 30 iterations were sufficient, are:[50]

$$v_n(z) = \pi(z) + \psi \mathrm{E}\left[e^{-\Delta z} \left\{ v_{n-1}(z + \Delta z) + \xi e^{\beta(z+\Delta z)} \int_0^{\Omega_{n-1}(z+\Delta z)} G(\omega)\, d\omega \right\} \right],$$

$$c_n = \mathrm{argmax}(v_n(z)),$$

$$\Omega_n(z) = \xi^{-1} e^{-\beta z} [v_n(c_n) - v_n(z)].$$

The distribution of Δz is Normal with mean $\ln(1-\delta)$ and variance equal to the total variance faced by an individual firm ($\sigma_0^2 \equiv \sigma_\varepsilon^2 + \sigma_A^2$). When calculating c_n we interpolate with a quadratic polynomial the value function $v_n(z)$ at the three points on the grid where the function is largest, and set c_n equal to the argument of the maximum value of this polynomial. By doing this maximization over a smoothed function, we avoid having to work with a discontinuous likelihood function.

We set the mean of the aggregate shocks equal to the mean estimated with the semi-structural approach.[51]

[49] Using 200 points makes no significant difference; we used 800 because the additional time involved was small. The reason why we need at least 200 points is that we fix the grid of possible values of x (between -3.5 and 3.5) in advance, so that often a significant part of this interval becomes irrelevant (the hazard is almost equal to one on it). Also, the finer the grid, the closer we can get to the case where the adjustment hazard looks like that of an (S, s) policy.

[50] See Section 2.2 for the derivations.

[51] We set this value ex-ante to avoid having to estimate additional nonlinear parameters. Because of the nonlinear adjustment term $\phi(v_{it} + \delta - x)$ in equation (38), there is no simple way to obtain an estimate of this drift from the data. Also note that, as described earlier, when calculating the invariant density of firm deviations, we allow for a firm-specific mean that is approximately equal to the observed mean of the corresponding sector.

Family of Adjustment Functions

The adjustment function estimated via dynamic optimization is evaluated on a grid of 800 points. This makes it computationally infeasible to solve the 945 nonlinear equations needed to calculate the v_{it}'s in every evaluation of the likelihood. For this reason we work with a family of adjustment functions characterized by only a few parameters and such that the derivatives needed for the Jacobian terms do not need to be calculated numerically.

Experimentation with a variety of distributions of adjustment costs showed that the family of continuous, piecewise inverted Normal adjustment functions approximates well the adjustment functions obtained via value iteration. Three pieces suffice for most practical purposes, with the middle piece equal to zero. A representative member of this four parameter family is of the form

$$\Lambda(x) = \begin{cases} 1 - e^{-\lambda^-(x - x^-)^2} & \text{if } x < x^-, \\ 0 & \text{if } x^- \le x \le x^+, \\ 1 - e^{-\lambda^+(x - x^+)^2} & \text{if } x > x^+. \end{cases} \tag{47}$$

We approximated the positive ($x > 0$) and negative ($x < 0$) arms of the adjustment function obtained via value iteration separately. We determined x^+ and λ^+ by imposing that the approximation matches the function obtained via value iteration at the (positive) points where the hazard equals 0.25 and 0.75. We obtained x^- and λ^- imposing an analogous condition for negative values of x.

Calculating the v_{it}'s and the Corresponding Derivatives

When keeping track of the cross-section of deviations, we approximate $f_i(\cdot, t-1)$ by 33 mass points on a grid of equally spaced points (we discuss why we chose 33 points shortly). We solve for v in

$$\mathscr{Y}[\mathscr{R}(\lambda^-, \lambda^+, x^-, x^+)\mathscr{A}(v + \delta)]f_i(x, t-1) = y_{it}.$$

The partial derivatives are calculated from (40). Next $f_i(x, t)$ is obtained from

$$f_i(x, t) = \mathscr{T}(v_{it} + \delta, \lambda^-, \lambda^+, x^-, x^+, \sigma_\epsilon)f_i(x, t-1).$$

The operator $\mathscr{A}(v_{it} + \delta)$ is implemented by shifting the 33 mass points describing $f_i(\cdot, t-1)$ by $v_{it} + \delta$. The adjustment operator, $\mathscr{R}$, is applied next, leading to 34 mass points (one at each point where there was mass before adjustments and a new mass at zero): if x is a point with mass $m(x)$ on the pre-shock grid, then after the adjustment shock we have mass $(1 - \Lambda(x))m(x)$ at x and mass $\Lambda(x)m(x)$ stemming from this point at zero.

Finally the idiosyncratic shock takes place: each of the 34 mass points becomes a Normal density with mean equal to the point where the mass was located and standard deviation σ_ϵ. The resulting density is computed at 33 equally spaced points on $[\mu_{it} - 4\sigma_{it}, \mu_{it} + 4\sigma_{it}]$, where μ_{it} and σ_{it}^2 denote the mean and variance of the cross-section obtained after the idiosyncratic shock.[52] We work with a dynamic grid to reduce the number of points needed to track the cross-section. Simulations showed that 33 points on a grid of width equal to 8 standard deviations, centered around the mean, suffice to obtain accurate estimates.

REFERENCES

BAR-ILAN, A., AND A. S. BLINDER (1992): "Consumer Durables: Evidence on the Optimality of Doing Nothing," *Journal of Money, Credit and Banking*, 24, 253–272.

BERTOLA, G., AND R. J. CABALLERO (1990): "Kinked Adjustment Costs and Aggregate Dynamics," in *NBER Macroeconomics Annual* 1990, ed. by O. J. Blanchard and S. Fischer. Cambridge, Mass.: The MIT Press, 237–288.

——— (1994): "Irreversibility and Aggregate Investment," *Review of Economic Studies*, 61, 223–246.

[52] Both the mean and variance are calculated analytically, thereby avoiding any circularity.

BLINDER, A. S. (1981): "Retail Inventory Behavior and Business Fluctuations," *Brookings Papers on Economic Activity*, 2, 443–505.

CABALLERO, R. J. (1993): "Durable Goods: An Explanation for their Slow Adjustment," *Journal of Political Economy*, 101, 351–384.

CABALLERO, R. J., AND E. M. R. A. ENGEL (1992a): "Beyond the Partial Adjustment Model," *American Economic Review, Papers and Proceedings*, 82, 360–364.

——— (1992b): "Price Rigidities, Asymmetries, and Output Fluctuations," NBER Working Paper 4091.

——— (1993a): "Microeconomic Adjustment Hazards and Aggregate Dynamics," *Quarterly Journal of Economics*, 108, 359–383.

——— (1993b): "Microeconomic Rigidities and Aggregate Price Dynamics," *European Economic Review*, 37, 697–711.

——— (1994): "Explaining Investment Dynamics in US Manufacturing: A Generalized (S, s) Approach," NBER Working Paper 4887.

CABALLERO, R. J., E. M. R. A. ENGEL, AND J. HALTIWANGER (1995): "Plant-Level Adjustment and Aggregate Investment Dynamics," *Brookings Papers on Economic Activity*, 2, 1–39.

CHIRINKO, R. S. (1993): "Business Fixed Investment Spending: A Critical Survey of Modelling Strategies, Empirical Results, and Policy Implications," *Journal of Economic Literature*, 31, 1875–1911.

COOPER, R., AND J. HALTIWANGER (1993): "The Aggregate Implications of Machine Replacement: Theory and Evidence," *American Economic Review*, 83, 360–382.

DIXIT, A., AND R. S. PINDYCK (1994): *Investment Under Uncertainty*. Princeton University Press.

DOMS, M., AND T. DUNNE (1993): "An Investigation into Capital and Labor Adjustment at the Plant Level," mimeo, Center for Economic Studies, Census Bureau.

GROSSMAN, S. J., AND G. LAROQUE (1990): "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58, 25–51.

HARRISON, J. M., T. M. SELLKE, AND A. J. TAYLOR (1983): "Impulse Control of Brownian Motion," *Mathematics of Operations Research*, 8, 454–466.

LAM, P. S. (1991): "Permanent Incoming, Liquidity, and Adjustment of Automobile Stocks: Evidence from Panel Data," *Quarterly Journal of Economics*, 106, 203–230.

LEHMANN, E. L. (1986): *Testing Statistical Hypothesis*, Second Edition. New York: John Wiley.

RAMSEY, J. (1996): "If Nonlinear Models Cannot Forecast, What Use Are They?" *Studies in Nonlinear Dynamics and Econometrics*, 1, 65–86.

RIVERS, D., AND Q. H. VUONG (1991): "Model Selection Tests for Nonlinear Dynamic Models," I.N.R.A., Toulouse Working Paper #91-08.

STOKEY, N., R. LUCAS, AND E. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*. Cambridge, Mass.: Harvard University Press.

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 257–306.