Review of Economic Studies (1994) 61, 223-246 0034-6527/94/00120223$02.00

Irreversibility and Aggregate Investment

GUISEPPE BERTOLA
Università di Torino, Princeton University, CEPR, and NBER

and

RICARDO J. CABALLERO
Massachusetts Institute of Technology and NBER

First version received February 1991; *final version accepted September* 1993 (*Eds.*)

Investment is often irreversible: once installed, capital has little or no value unless used in production. This paper proposes and solves a model of sequential irreversible investment and characterizes the aggregate implications of microeconomic irreversibility and idiosyncratic uncertainty. If a large amount of idiosyncratic uncertainty is allowed for, the distributional dynamics induced by the nonlinear character of irreversible investment policies are capable of smoothing the dynamics of aggregate investment (relative to those of its forcing processes) to the extent required by U.S. data.

1. INTRODUCTION AND RELATED LITERATURE

Capital accumulation and investment decisions have an essential role both in the theory of production and in the study of macroeconomic fluctuations. Early models of individual firms' and aggregate investment were based on the static relationship between the marginal revenue product of capital and its user cost, as defined by Jorgenson (1963). Subsequent research recognized that technology and market structure make it costly for firms to adjust their capital stock, so that investment can only be studied in an explicitly dynamic framework. Standard investment models assume variations in capital input to entail convex adjustment costs, either internal to the firm and due to increasing costs of installing more capital in shorter intervals of time, or external to it and due to decreasing returns in the production of capital goods. Further assumptions are typically necessary to obtain analytically and empirically tractable investment models: firms may be assumed to be perfectly competitive and to operate under constant returns to scale (e.g. Lucas and Prescott (1971), Hayashi (1982)), or linear-quadratic functional forms may be assumed to obtain certainty equivalence (e.g. Sargent (1987)). Models of investment based on these assumptions do not provide a convincing interpretation of empirical evidence (Abel and Blanchard (1986)), and it is fair to doubt the realism of smooth adjustment costs as the source of investment dynamics. From a microeconomic point of view, in fact, the unit cost of additions to an individual firm's capital stock may well be constant in the rate of investment, or even decreasing if lump-sum adjustment costs are present. Rather, the forward-looking nature of investment decisions is often due to their (at least partly) irreversible character. Many facilities are specific to a particular production process;

conversion of industrial real estate is difficult; and sale of used machinery faces thin markets and heavy discounts. Installed capital is therefore valuable only to the extent that it is used in production, and firms' investment decisions must take into account *future* cost and demand conditions, as was recognized early on by Arrow (1968) and Nickell (1974) in non-stochastic partial-equilibrium models of investment.[1]

This paper discusses the relevance of uncertainty and irreversibilities to firms' investment decisions on the one hand and, on the other, to aggregate dynamic relationships between capital accumulation and its driving processes. At the microeconomic level, irreversible investment under uncertainty can be studied by dynamic programming or option pricing techniques (see e.g. McDonald and Siegel (1986), Demers (1991), and their references). Since an irreversible investment decision forsakes the option to wait for some of the uncertainty to be resolved, the project will be adopted only when the expected discounted payoff from investment exceeds the cost by an amount that can be impressively large for plausible parameter values. The insight has been developed by a burgeoning literature (see the surveys by Pindyck (1991) and Dixit (1992)). Pindyck (1988) applies option pricing techniques to marginal irreversible investment choices, and Bertola (1988) shows that the solution to problems of this type can be derived by dynamic programming as well as by option evaluation methods; formally similar problems of "singular" stochastic control have been studied in operations research by probabilistic and/or analytical methods (see e.g. Chow *et al.* (1985), Karatzas and Shreve (1984, 1988) and Kobila (1991)).

Section 2 below presents a simple microeconomic model of irreversible investment, where the irreversibility constraint generates an intermittent investment process at the firm level and drives a variable wedge between capital's marginal revenue product and Jorgensonian "user cost". We then proceed to study the relevance of irreversibility to empirical work on aggregate investment series.

Irreversibility is especially realistic at the aggregate level, as the direct consumption value of existing productive facilities is clearly very low; Sargent (1979) and Olson (1989) have implemented irreversibility constraints in the representative agent, single good framework of stochastic growth models. However, the dynamics of aggregate production and investment are not variable enough (at least in industrialized countries) for irreversibility constraints to be binding at the aggregate level. We argue instead that *microeconomic* irreversibilities are relevant to aggregate dynamics in the presence of important sources of idiosyncratic uncertainty.

In Section 3 we exploit the simplicity of our microeconomic model (and the similarity of its functional-form to those of previous empirical investment models) to study the implications of microeconomic irreversibilities for aggregate investment. We discuss a model of stochastic aggregation similar to those considered by Caballero and Engel (1991) and Bertola and Caballero (1990). In Section 4 we assess the empirical relevance of our modeling assumptions and results with an application to postwar U.S. investment series. Idiosyncratic uncertainty gives empirical relevance to irreversibility constraints in spite of the relatively low volatility of aggregate variables and, inasmuch as it implies that times of positive gross investment are imperfectly synchronized across firms, it also makes it possible to interpret the persistent (but smooth) dynamics of aggregate investment series.

1. Under certainty, the marginal revenue product of capital equals the neoclassical user cost of capital only when gross investment is strictly positive. Investment is not necessarily always positive, if it is irreversible: it ceases before a cyclical peak is reached, and starts again after the cyclical trough. At the individual firm's level, then, irreversibility drives a wedge (negative during booms, and positive during pronounced troughs) between the cost of capital and its marginal contribution to profits.

2. OPTIMAL SEQUENTIAL INVESTMENT UNDER UNCERTAINTY

We consider a firm whose operating cash flows are a constant elasticity function $Y(K, Z)$ of K, the installed capital stock, and Z, an index of business conditions:

$$Y(K(\tau), Z(\tau)) = K(\tau)^{\alpha} Z(\tau), \qquad 0 < \alpha < 1. \tag{1}$$

This functional form may be taken to represent a log-linear approximation to more general ones. Of course, the approximation is exact if the firm's production and demand functions have constant elasticity, and the accuracy of the approximation might in principle be evaluated empirically (in Section 4 below we find no evidence against this reduced form). In general, the business conditions process $\{Z(\tau)\}$ depends positively on the strength of demand for the firm's product and on productivity, and negatively on the cost of factors other than capital.

Let $\{Z(\tau)\}$ follow the process

$$dZ(\tau) = Z(\tau)(\vartheta_1 d\tau + \sigma_1' dW(\tau)) \tag{2}$$

where ϑ_1 is a constant scalar, σ_1 is a 2×1 constant vector, and $\{W(\tau)\}$ is a two-dimensional Wiener process (the presence of two sources of uncertainty allows for randomness in the purchase price of capital below, and the matrix notation makes it possible to do so at little or no cost in terms of complexity). As a univariate process, the business-conditions index Z follows a geometric Brownian motion. This modeling assumption may again be viewed as a simple and reasonably realistic representation of uncertainty, or it may be given a more structural interpretation. Under constant elasticity demand and production functions, the multiplicative disturbance $\{Z(\tau)\}$ follows the process in (2) if demand, productivity, and the cost of flexible factors of production grow at some constant mean rate which is perturbed in continuous time by normally distributed random variables, independent over time and possibly contemporaneously correlated.

Capital can be purchased and installed at unit price $P(\tau)$. Installed capital has no resale value, however. By equations (1) and (2), $Z(\tau) > 0$ for all τ and the marginal contribution of installed capital to operating profits is always positive. Thus, scrapping is never profitable, investment is irreversible, and the installed capital stock process $\{K(\tau)\}$ decreases only via depreciation, which we assume to take place at constant exponential rate δ. We shall take the purchase price of capital $\{P(\tau)\}$ to have dynamics described by

$$dP(\tau) = P(\tau)(\vartheta_2 d\tau + \sigma_2' dW(\tau)), \tag{3}$$

where ϑ_2 and σ_2 are conformable to ϑ_1 and σ_1.

The firm's managers choose the investment policy so as to maximise the market value of the firm, defined as the present discounted value at rate r of expected future cash flows. By our assumptions, the sample path of $\{W(\tau)\}$ contains all the information relevant to the firm's problem. By (2) and (3), the probability distribution of $\{P(\tau), Z(\tau)\}$ as of time t is uniquely determined by $P(t)$ and $Z(t)$ for all $\tau \geqq t$. The optimal investment process $\{G(\tau)\}$ is to be chosen among the class of non-decreasing processes which depend, at every time t, on the information contained in the sample path of $\{Z_\tau\}$ and $\{P_\tau\}$ up to time t: the investment process cannot predict the future.[2] We then define the value of an

2. Formally, what is required is that $\{G(\tau)\}$ be progressively measurable with respect to the filtration $F_t^W \equiv \sigma(W(s); 0 \leqq s \leqq t)$, the non-decreasing family of sigma-fields generated on the space of continuous functions $t \to R^2$ by observation of $W(\tau)$. By the accumulation constraint (5), the installed capital stock process $\{K(\tau)\}$ is also adapted to $\{F_t^W\}$.

investment programme as

$$V(K(t), Z(t), P(t)) \equiv \max_{\{G(\tau)\}} E_t\left\{\int_t^\infty e^{-r(\tau-t)}(K(\tau)^\alpha Z(\tau)d\tau - P(\tau)dG(\tau))\right\}, \tag{4}$$

$$\text{subject to } dK(\tau) = -\delta K(\tau)d\tau + dG(\tau), \qquad dG(\tau) \geqq 0, \tag{5}$$

where $\{G(\tau)\}$ is the cumulative gross investment process, restricted to have positive increments, and $E_t\{\cdot\}$ denotes conditional expectation taken, at time t, over the joint distribution of the $\{Z(\tau)\}$, $\{P(\tau)\}$, and $\{K(\tau)\}$ processes. While $\{Z(\tau)\}$ and $\{P(\tau)\}$ are exogenous to the firm's problem, the distribution of $\{K(\tau)\}$ is determined endogenously by the optimal investment rule.

Reversible investment

If capital could be purchased or sold at the same price $P(t)$, then $dG(t)$ would be unconstrained and the first-order condition for choice of capital stock at every point in time would be

$$\partial_K \Pi(K(t), Z(t)) = (r + \delta - \vartheta_2)P(t), \tag{6}$$

where $\partial_x f(y)$ denotes a partial derivative with respect to x evaluated at $x = y$. This is the Jorgenson (1963) optimality condition, equating the marginal revenue product of capital to its *user cost*. Intuitively, if the purchase and sale price of capital were always equal to each other (though random over time) the expected opportunity cost of carrying a stock of progressively depreciating capital should equal the flow operating cash flows from its use in production.

Under the assumed functional forms, (6) yields an expression for the *frictionless* capital stock,

$$K^f(Z(t), P(t)) = \left(\frac{(r+\delta-\vartheta_2)}{\alpha}\frac{P(t)}{Z(t)}\right)^{1/(\alpha-1)} \quad \forall t. \tag{7}$$

When investment is unconstrained, $K(t)$ is not a state variable and the value of the firm's investment strategy is given by

$$V^f(Z(t), P(t)) = E_t\left\{\int_t^\infty e^{-r(\tau-t)}(K^f(\tau)^\alpha Z(\tau)d\tau - P(\tau)(dK^f(\tau) + \delta K^f(\tau)d\tau))\right\} \tag{8}$$

Based on the information available at time t, $K^f(\tau)^\alpha Z(\tau)$ is log-normally distributed by (2), (3), and (7). It is then not difficult to show that the integral in (8) converges if

$$r > \left(\frac{\vartheta_1}{1-\alpha} - \frac{\alpha\vartheta_2}{1-\alpha} - \frac{1}{2}\sum{}^2 \frac{\alpha}{(1-\alpha)^2}\right) \tag{9}$$

where $\sum^2 \equiv (\sigma_1 - \sigma_2)'(\sigma_1 - \sigma_2)$ is the variance per unit time in the growth rate of $\{Z(t)/P(t)\}$. The right-hand side of (9) is the expected rate of increase of revenues. For the firm's infinite-horizon value to be finite, the required rate of return r must be large relative to ϑ_1, the growth rate of operating profits for given capital, and $-\vartheta_2$, the expected rate of deflation in the capital purchase price. We shall of course assume that the parameters are such that (9) holds.

It is worth remarking at this point that, insofar as the frictionless investment policies are concerned, the two sources of uncertainty represented by the elements of $W(t)$ are only relevant to the extent that they affect the *single* state variable $Z(t)/P(t)$. It will become apparent in what follows that the same univariate state variable is also a sufficient statistic for the derivation and characterization of irreversible investment decisions, and that $\sum$ as defined in (9) is a measure of the degree of uncertainty relevant to such decisions: if it were the case that $\sigma_1 = \sigma_2$, for example, the cost of capital would move one-for-one with its usefulness in production and, while such movements would in general affect the overall value of the firm, they would be irrelevant to marginal investment decisions. Our admittedly stylized functional forms afford useful flexibility in applications, by not constraining the ex-post cost of capital to be constant over time, without much increasing the algebraic complexity of the model.

Characterization of irreversible investment

If investment is irreversible, the installed capital stock $\{K(\tau)\}$ may depend on the whole past history of $\{Z(\tau)\}$, $P(\tau)\}$. Thus, $\{K(t)\}$ is a state variable at time t. However, history dependence does not extend past the last time of positive gross investment. It is convenient to define a *desired capital* process $K^d(P(t), Z(t))$ such that $K(t) \geqq K^d(P(t), Z(t))$ for all t, and $K(t) = K^d(P(t), Z(t))$ if $dG(t) > 0$. To see that an optimal irreversible investment programme can be characterized in terms of the "desired capital" construct, imagine momentarily lifting the irreversibility constraint at some time $\tilde{t}$. At $\tilde{t}$, the installed capital stock would be unconstrained, thus not a state variable. If an optimal choice of capital stock at $\tilde{t}$ exists, it must be a function of $Z(\tilde{t})$ and $P(\tilde{t})$. If $dG(\tilde{t}) > 0$ in the optimal irreversible process, the irreversibility constraint is not binding at $\tilde{t}$ and removing it has no effect: hence $dG(t) > 0$ implies $K(t) = K^d(P(t), Z(t))$. As the firm may choose to decrease its capital stock when given an opportunity to do so, $K(t) \geqq K^d(P(t), Z(t))$ at all times.

The desired-capital construct conveniently summarizes the firm's irreversible investment decision rule. If the currently installed capital stock $K(t)$ is smaller than $K^d(t)$, the firm should immediately invest so as to obtain $K(t) = K^d(t)$; otherwise $K(t)$ should be allowed to depreciate. The firm's managers need to form expectations for the distant future when deciding when and how much to invest, because irreversible investment decisions, unlike reversible ones, are relevant to future cash flows: the firm may find itself stuck with an excessive stock of capital. The desired capital stock, however, is by definition a function of current observables only.

The optimal investment rule can be computed explicitly under our functional form assumptions. Appendix A derives differential equations which are necessarily satisfied by the value function $V(\cdot)$ and by its derivative with respect to K, denoted $v(\cdot)$, along the optimal capital accumulation path. The investment policy that solves these functional relationships has a simple and intuitive form: the marginal revenue product of capital should never be allowed to exceed a *constant* proportion c of the purchase price of capital P,

$$\partial_K \Pi(K(t), Z(t), P(t)) \begin{cases} \leqq cP(t) & \forall t; \\ = cP(t) & \forall t \text{ such that } dG(t) > 0. \end{cases} \tag{10}$$

Under our functional form assumptions, the ratio of flow marginal profits to purchase price of capital which triggers investment is constant and equals

$$c \equiv r + \delta - \vartheta_2 + \tfrac{1}{2} \textstyle\sum^2 A, \tag{11}$$

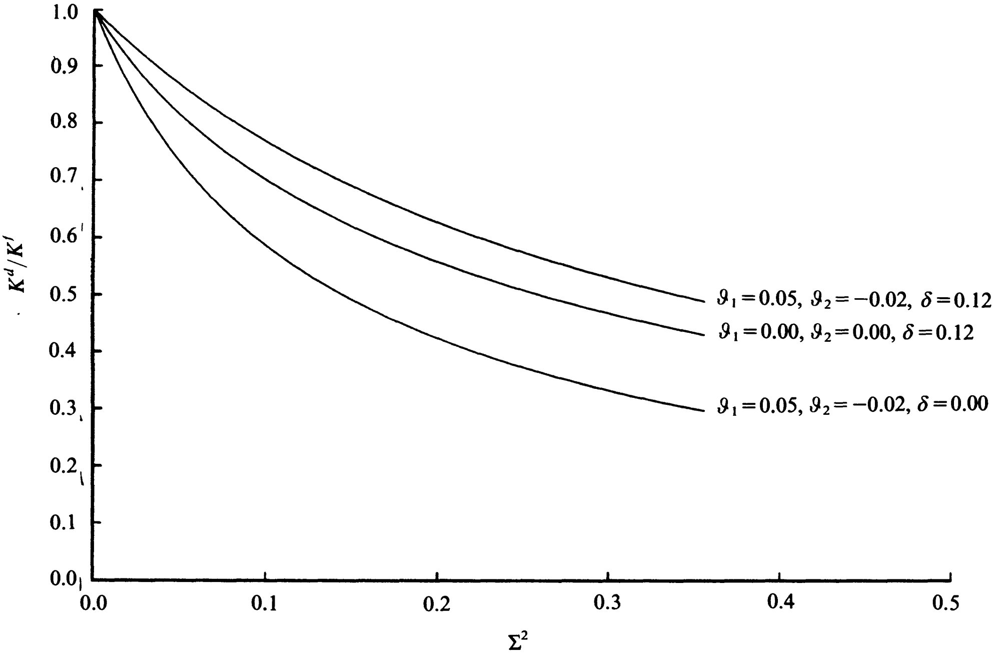


FIGURE 1

where the constant A (defined in Appendix A) is strictly positive under the parametric restriction (9). Thus, when $\sum^2>0$ the marginal revenue product of capital that triggers irreversible investment is larger than the neoclassical user cost of capital.

Inverting the marginal condition (10), we obtain an expression for the firm's desired capital stock as a function of the current values of $Z(t)$ and $P(t)$:

$$K^d(Z(t), P(t))=\left(\frac{c}{\alpha}\frac{P(t)}{Z(t)}\right)^{1/(\alpha-1)} \tag{12}$$

Under the assumed functional forms, the features of the $\{Z(t)\}$ and $\{P(t)\}$ stochastic processes which are relevant to the firm's problem can be summarized by the scalar constant c. It can be shown that $c/(r+\delta-\vartheta_2)$ is decreasing in ϑ_1, increasing in ϑ_2, δ, and $\sum^2$. Intuitively, current decisions are likely to make the irreversibility constraint binding over the relevant planning horizon, determined by the discount rate r, if the rate of increase of desired capital is expected to be lower (relative to the depreciation rate of installed capital), or if it is more volatile.

Figure 1 plots the ratio of desired to frictionless capital stocks against $\sum^2$ for several values of the other parameters. As long as $\sum^2>0$, the "desired" irreversible capital stock $K^d(Z(t), P(t))$ is smaller than the frictionless capital stock $K^f(Z(t), P(t))$ of equation (7). It does not follow, however, that the installed stock of capital should generally be smaller when investment is irreversible than when it is reversible. In fact, the risk-neutral firm under consideration is *ex-ante* reluctant to undertake irreversible investment only because adverse realizations in the process it takes as exogenous may *ex-post* leave it stuck with excess capital. Thus, although $K(t)<K^f(t)$ at times of positive gross investment, we should

observe $K(t) > K^f(t)$ when the realizations of $\{Z(t)\}$ and $\{P(t)\}$ are such as to make the firm regret having invested in the past.[3]

3. AGGREGATE INVESTMENT

No real firm's investment problem corresponds exactly to that solved in Section 2, of course. Still, our stylized model of homogeneous capital accumulation yields a closed-form investment rule under reasonable functional form assumptions, and the simplicity of the solution makes it possible to discuss the *aggregate* implications of firm-level irreversibility constraints. This is a difficult problem, and several simplifying assumptions are necessary. We focus on investment in *new* capital goods, and we model aggregate investment in terms of stochastic aggregation of a very large number (approximated by a continuum) of individual units indexed by $i \in [0, 1]$. Each "unit" may be taken to represent a specific type of homogeneous capital, owned by a specific economic agent. Real-life investment decisions are not taken in isolation: different firms' decisions to purchase similar equipment depend on each other through market interactions; capital equipment is in general heterogeneous even within the same productive process, with different types of capital being substitutable or complementary to each other; and it may be possible to reconvert capital on hand to new use, or to sell it to other users (at a price, of course, which reflects the equipment's current profitability and replacement cost). We shall model all linkages across investment decisions—those due to common ownership of heterogeneous types of capital, and those deriving from market interactions among distinct decision makers—in terms of cross-sectional correlation of innovations in the unit-specific stochastic processes. We assume each unit to provide its owner with net cash flows approximated by the constant-elasticity function of equation (1), disturbed by a process $Z_i(t)$, and we let the capital stock installed in it be (irreversibly) accumulated by paying a stochastic price $P_i(t)$. The process indexed by i are unit-specific, but their increments may be cross-sectionally correlated; the extent to which this is the case will be crucial to the character of aggregate investment dynamics.

As in the microeconomic section, we proceed in two steps. We first discuss the path of the aggregate stock of capital in the absence of irreversibility constraints, and we note that its dynamic properties are the same as those of the aggregate "desired" stock of capital. In the second step we describe aggregate investment in the presence of irreversibility constraints.

Aggregation of reversible investment

The parameters of the individual unit's problem could all be indexed by i without substantially affecting the results. For simplicity, however, let α, r, ϑ_1, ϑ_2, σ_1 and σ_2 be the same for all units. Given the log-linearity of the microeconomic model, *reversible* investment policies are easily aggregated, even allowing for cross-sectional heterogeneity in the realizations of unit-specific stochastic processes. Consider the logarithm of unit

3. On average, the capital intensity of production under investment irreversibility is actually *higher* than it would be if equation (7) applied at all times. This can be verified by the long-run average expressions derived in the next section, and is due to the discounting effects discussed in Bertola (1988) and Bentolila and Bertola (1990). We disregard these level effects in this paper, to focus on the dynamic implications of irreversibility for the investment series.

i's revenue-maximizing capital stock in the absence of irreversibility constraints, from equation (7):

$$k_i^f(t) \equiv \ln K_i^f(t) = \frac{1}{\alpha - 1} \ln \left(\frac{(r + \delta - \vartheta_2)}{\alpha} \frac{P_i(t)}{Z_i(t)} \right). \tag{13}$$

If $Z_i(t)$ and $P_i(t)$ follow geometric Brownian motion processes (equations (2) and (3)), then an application of Itô's lemma yields

$$dk_i^f(t) = \Theta dt + \sigma dW_i(t) \tag{14}$$

where $W_i(t)$ is a univariate Brownian motion process constructed as a combination of the processes driving $P_i(t)$ and $Z_i(t)$, and

$$\Theta = \frac{\vartheta_1 - \vartheta_2 - \frac{1}{2}(\sigma_1' \sigma_1 - \sigma_2' \sigma_2)}{1 - \alpha}, \qquad \sigma = \frac{\sqrt{(\sigma_1 - \sigma_2)'(\sigma_1 - \sigma_2)}}{1 - \alpha}.$$

Consider next the process followed by *aggregate* investment, which we define as a weighted integral on the interval [0, 1] (approximately a large number of finitely-sized units by a continuum of infinitesimal units): with $d\tilde{k}(t)$ denoting the rate of growth of aggregate capital, we have

$$d\tilde{k}(t) = \int_0^1 w_i(t) dk_i(t) di,$$

if unit i's share in aggregate capital is $w_i(t)$. Aggregating (14),

$$d\tilde{k}^f(t) = \Theta dt + \sigma_A d\tilde{W}(t) \tag{15}$$

where

$$\sigma_A d\tilde{W}(t) \equiv \int_0^1 w_i(t) \sigma dW_i(t) di$$

is the *aggregate* component of individual units' investment processes. This definition induces a decomposition $\sigma dW_i(t) \equiv \sigma_A d\tilde{W}(t) + \sigma_I dW_{Ii}(t)$ for unit-specific innovations, whose *idiosyncratic* component $dW_{Ii}(t)$ aggregates to zero identically. If investment were reversible, then idiosyncratic uncertainty would wash out in aggregate capital dynamics as well as in its aggregate forcing processes, and microeconomic relationships like (13) would translate exactly to aggregate data.

Irreversible investment

Investment functions derived from equations like (13), however, perform poorly when confronted with actual data, at all levels of aggregation. In particular, their error terms are strongly serially correlated, prompting researchers to include lags in their investment equations and to rationalize them by *ad hoc* adjustment cost functions. We prefer to interpret the empirical shortcomings of equations like (13) in terms of unit-level irreversibility. Each unit's irreversible capital accumulation path is determined by unit-specific $Z_i(t)$ and $P_i(t)$ processes through its desired capital stock process as defined in equation (12). Since the desired and reversible capital stocks differ only by a constant of proportionality, the dynamics of $k_i^d(t) \equiv \ln K_i^d(t)$ coincide with those of $k_i^f(t)$. Aggregating, we obtain

$$d\tilde{k}^d(t) = \Theta dt + \sigma_A d\tilde{W}(t). \tag{16}$$

Since binding irreversibility constraints destroy the convenient log-linearity of individual units' policies, aggregation of *actual* investment processes is not as easy. Roughly speaking, we have $dk_i(t)=dk_i^d(t)$ at times when unit i is investing, and $dk_i(t)=-\delta dt$ at all other times.[4] Investment by each individual unit responds fully to exogenous shocks if its capital stock is the "desired" one—but is completely insensitive to them if the irreversibility constraint is currently binding. To aggregate such nonlinear policy functions, we need to know *how many* of the microeconomic units are in the former or in the latter situation, and theory must then address *distributional* issues. It will be convenient to work with the cross-sectional distribution of $s_i(t)\equiv k_i(t)-k_i^d(t)$, the log-deviation of each unit's actual capital stock from its desired one. Clearly,

$$ds_i=\begin{cases}-\delta dt-dk_i^d(t) & \text{when } dG_i(t)=0,\\ 0 & \text{otherwise.}\end{cases}$$

The dynamic response of actual investment to aggregate shocks depends on the position of individual units in state space (the s_i's) and on the magnitude of idiosyncratic shocks. Integrating over i and defining $\tilde{s}(t)=\int_0^1 w_i(t)s_i(t)di$, we obtain the dynamics of the aggregate installed capital stock:

$$d\tilde{k}(t)=d\tilde{k}^d(t)+d\tilde{s}(t). \tag{17}$$

Noting that

$$d\tilde{k}(t)=d\ln\tilde{K}(t)=\frac{d\tilde{G}(t)-\delta\tilde{K}(t)dt}{\tilde{K}(t)},$$

where $d\tilde{G}(t)$ and $\tilde{K}(t)$ denote aggregate gross investment and aggregate capital, we can rewrite equation (17) in terms of gross investment ratios[5]

$$\Gamma(t)\equiv\frac{d\tilde{G}(t)}{\tilde{K}(t)},\qquad \Gamma^*(t)\equiv\frac{d\tilde{K}^f(t)}{\tilde{K}^f(t)}+\delta=\frac{d\tilde{K}^d(t)}{\tilde{K}^d(t)}+\delta,$$

to obtain

$$\Gamma(t)=\Gamma^*(t)+d\tilde{s}(t). \tag{18}$$

In our stochastic aggregation framework, the *actual* and *desired* gross investment/capital ratios differ by $d\tilde{s}(t)$: the change in the average difference between installed and desired capital stocks at the individual units' level.

While only the mean of the cross-sectional distribution of $k_i(t)-k_i^d(t)$ is directly relevant to aggregate phenomena at a point in time, the dynamics of this mean are determined by all moments of the distribution of s_i's in the recent past (see equation (23) below). To study aggregate investment, it is then necessary to track the whole cross-sectional density, which we denote by $f(s,t)$.[6]

4. This statement is formally correct since the processes under consideration have continuous sample paths. Note, however, that $dk_i(t)$ is (infinitesimally) positive only on a measure-zero set of time points, reflecting the *singular* character of the optimal investment policy.

5. For most purposes, we might equivalently work with capital stock (log) levels which, by (17), obey $\tilde{k}(t)=\tilde{k}^d(t)+\tilde{s}(t)$. We choose to work with first differences because we feel that the dynamics to be explained, at business-cycle frequencies, are those of the investment rate. We do make use of the (cointegrated) relationship in levels to check empirically our assumptions regarding the $k^d(t)$ series (see Section 4 below).

6. Since the identity of units at different points in the state space is irrelevant from the aggregate point of view, it is not necessary to study the *joint* probability distribution of individual units (a process of much higher dimensionality). Caballero and Engel (1991) discuss this point further.

In the case we are considering, microeconomic units are heterogeneous only in that they are affected by ongoing idiosyncratic uncertainty, and such cross-section heterogeneity is essential to the relationship between aggregate dynamics and the underlying microeconomic adjustment processes. In the absence of idiosyncratic uncertainty, individual units' investment problems could only possibly differ by their initial conditions; over time, however, whenever a firm invests its s_i would become closer to that of other firms, and in the limit $f(s, t)$ would converge to a spike. Individual units would thereafter be homogeneous in the levels as well as the dynamics of s_i, and the aggregate investment process would have the intermittent character implied by the microeconomic investment rule of Section 2.

In reality, of course, investment decisions are far from being perfectly synchronized, and idiosyncratic shocks prevent the cross-sectional density from degenerating into a spike. As a result, the aggregate investment process is a smoothed version of the nonlinear, discontinuous microeconomic policies. Before considering aggregate shocks, it is instructive to describe the limit case where all uncertainty faced by firms is idiosyncratic and the cross-sectional density has settled into a steady state. Since the number of firms is large, the steady state cross sectional density corresponds to the ergodic density of a single s_i (see Billingsley (1986)). As each s_i behaves as a Brownian motion regulated at 0, with standard deviation σ_I and drift $\vartheta \equiv -(\Theta + \delta)$, the steady-state density is exponential (see Appendix B below):

$$f(s) = \xi e^{-\xi s} \qquad s \geqq 0, \quad \text{where } \xi \equiv -2\vartheta/\sigma_I^2. \tag{19}$$

The two solid lines in Figure 2 plot cross-sectional densities for two positive values of ξ. With positive depreciation ($\delta > 0$) and a secular tendency for desired investment to be positive ($\Theta > 0$), we have $\xi > 0$ and every individual tends to drift towards the investment point, where $s_i = 0$. Hence, in steady state more units are found in the immediate neighbourhood of $s = 0$ than farther from it; further, the stronger is the drift in desired capital and rate of depreciation, the steeper is the slope of the cross-sectional density and the larger is the measure of units investing at any point in time.

Consider then the role of ongoing aggregate shocks in shaping the cross-sectional distribution of units. Accumulation over a finite time-period of abnormally positive aggregate shocks has roughly the same role as a larger mean rate of growth Θ: it implies a tendency for the cross-sectional distribution to become steeper, so that a larger measure of units will be investing at every point in time. Over an interval of time of given length, the effect of cumulative aggregate innovations on aggregate gross investment depends on the initial shape of the cross-sectional distribution. Formally, consider the evolution of the cross-sectional density in the time inverval $(t - \Delta t, t]$, starting at $f(s, t - \Delta t)$, and let the drift of desired capital growth during this time interval be Θ_t (accordingly, $\vartheta_t = -(\delta + \Theta_t)$ and $\xi_t = -2\vartheta_t/\sigma_I^2$). The path of the cross-sectional density during this time interval can be obtained from the solution of the forward Kolmogorov equation:

$$\partial_h f(s, h) = \tfrac{1}{2}\sigma^2 \partial_{ss} f(s, h) - \vartheta_t \partial_s f(s, h), \tag{20}$$

with boundary conditions

$$\tfrac{1}{2}\sigma^2 \partial_s f(0, h) = \vartheta_t f(0, h) \quad \forall h \in (t - \Delta t, t), \tag{21}$$

$$\lim_{s \to \infty} (\tfrac{1}{2}\sigma^2 \partial_s f(S, h) - \vartheta_t f(S, h)) = 0 \quad \forall h \in (t - \Delta t, t). \tag{22}$$

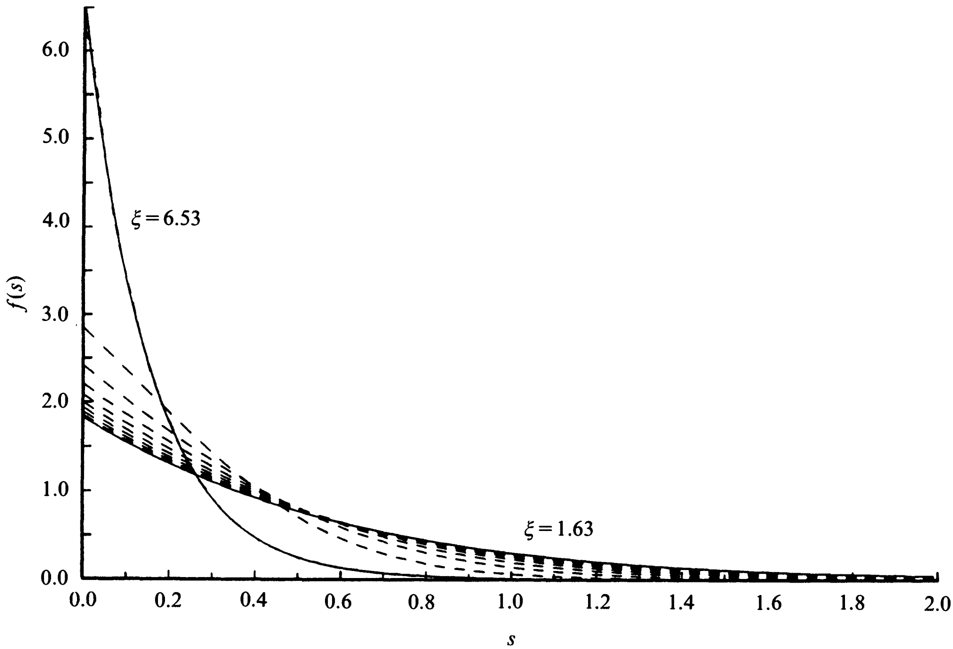


FIGURE 2

In Appendix B we show that the solution of the partial differential equation (20) and its boundary condition, for given $f(s, t-\Delta t)$, may be expressed in the form

$$f(s, h)=\xi_t e^{-\xi_t s}+\int_{0^+}^{\infty} A(\beta;t)e^{-\lambda(\beta)(h-(t-\Delta t))}e^{\frac{-\xi_t s}{2}}\left(\cos(\beta s)-\frac{\xi_t}{2\beta}\sin(\beta s)\right)d\beta, \qquad (23)$$

where $\lambda(\beta)>0$ for all $\beta \geqq 0$ and the functional constants of integration $A(\cdot\,;\cdot)$ depend on the entire initial cross-sectional distribution, $f(x, t-\Delta t)$ for $x\in[0, \infty)$. The dashed lines in Figure 2 plot some of the distributions encountered as the cross-sectional density converges from the steeper ergodic distribution to the flatter one.

This equation usefully highlights the *smoothness* and *inertia* which characterize aggregate dynamics in the presence of microeconomic irreversibility constraints. The first term on the right-hand side of (23) in the cross-sectional density that would result in steady state if the drift were constant at Θ_t forever. We know from Figure 2 that the cross-sectional density tends to concentrate a larger fraction of its mass near the investment barrier as this drift rises, thus lowering the limit value of cross-sectional density's mean, $\bar{s}_t=1/\xi_t$. Recalling from equation (18) that changes in $\bar{s}$ drive a wedge between the actual and "desired" investment ratios, we find that the model has *smoothing* properties: as the driving processes induce a stronger tendency to invest, the cross-sectional density's dynamics partially offset the change in the desired capital stocks. The integral term in equation (23), conversely, reflects the *inertial* role of history. The shape of the initial distribution (which determines the $A(\cdot\,;\cdot)$ functions) matters for subsequent investment dynamics, and its importance decreases over time since $\lambda(\beta)>0$ for all β. If the initial cross-sectional distribution was shaped by past Θ values which were larger than the current one, a large measure of firms is close to the investment barrier: and many of these firms are likely to be led to the investment point by idiosyncratic shocks and the current (smaller) drift.

The formal model of the previous sections assumes Brownian motion dynamics at both the idiosyncratic and the aggregate level. Roughly speaking, aggregate Brownian shocks induce infinitely frequent changes in the drift shaping the cross-sectional distribution. While a rigorous analysis of aggregate dynamics should take this into account, the formal results and economic insights obtained from the above discussion (of drift changes over time intervals of *finite* length Δt) will suffice for our purposes. The analytic framework outlined above provides a good approximation to the theoretical model if Δt is small, and no additional economic insights would be gained by explicitly considering infinite-variation components of aggregate shocks. Most importantly, the approximate solution readily lends itself to further analytical and empirical work, while a more rigorous approach would involve solution of stochastic partial differential equations whose solutions can be shown to exist (see e.g. Krylov and Rozovskii (1977)), but have no explicit analytical representation.

We consider the empirical implications of our model next. Only the mean of the cross-sectional distribution is directly observable in aggregate data. A mean, of course, is consistent with an infinite variety of shapes for the cross-sectional density $f(s, t)$; as noted above, each of these shapes has different implications for the mean's responsiveness to further aggregate shocks, and must in turn be consistent with the pattern of aggregate shocks observed in the past. The empirical problem is then one of inferring, from the observed dynamics of endogenous and exogenous variables, the shape of cross-sectional densities at every observation point—which depends on the history of aggregate shocks and, given the assumed probability structure, on the relative importance of aggregate and idiosyncratic sources of uncertainty. Appendix B below outlines how this can be done exploiting the discrete-time approximation introduced in this section, and the next section presents some results.

4. EMPIRICAL IMPLICATIONS AND EVIDENCE

This section interprets the behaviour of U.S. investment in light of our microeconomic and aggregation results. We work on the annual $\Gamma(t)$ series constructed as the ratio of gross investment in non-residential capital equipment to the relevant capital stock, over the 1954–1986 period, as reported by the Department of Commerce of the Bureau of Economic Analysis. The objective is to construct a predicted aggregate (gross) investment path $\hat{\Gamma}(t)$, which we decompose into the sum of the paths of aggregate frictionless investment, $\Gamma^*(t)$, and of the changes in the mean of the cross-sectional density, $\Delta\bar{s}(t)$. Once again, we proceed in two steps. We start by constructing a (hypothetical) frictionless investment series based on observables, and we proceed to match some moments of the predicted aggregate investment series to the corresponding ones of the actual investment rate $\Gamma(t)$.

"Desired" investment

The hypothetical frictionless investment series $\Gamma^*(t)$ depends only on the cross-sectional first moments of the process taken as given by the microeconomic unit considered in Section 2. These are not directly observable but, under the functional form assumptions of Section 2, can be inferred from capital stock and production data. By equation (1), in fact, $Z(t)=Y(t)K(t)^{-\alpha}$: replacing this in equation (13), adding the index i, and taking

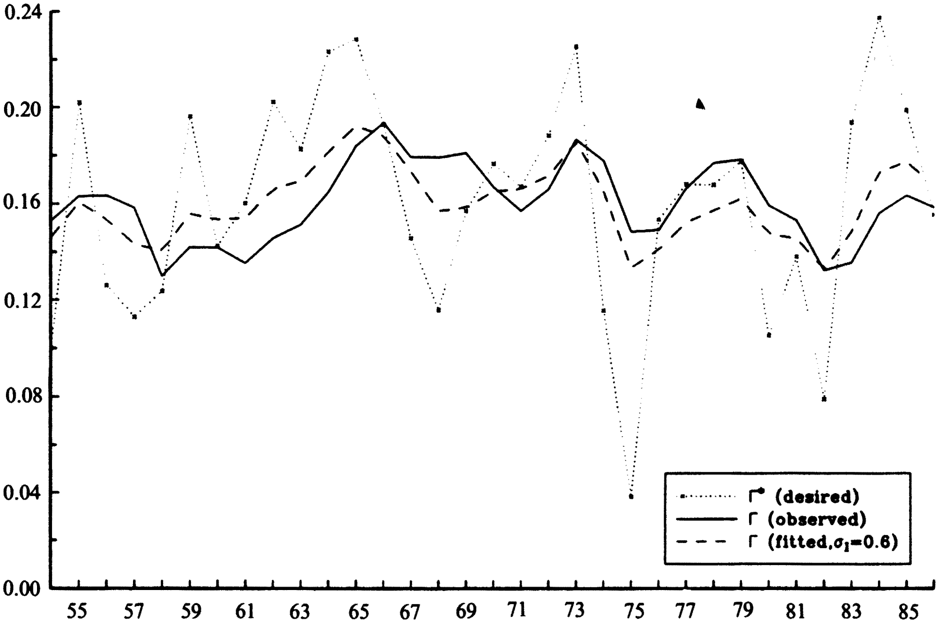


FIGURE 3

Desired, actual and fitted investment ratios

logarithms, we obtain

$$k_i^f(t) = \frac{1}{1-\alpha}(\ln Y_i(t) - \ln r_{k_i}(t)) - \frac{\alpha}{1-\alpha}k(t) + \frac{1}{1-\alpha}\ln(\alpha) \tag{24}$$

where $r_k \equiv (r + \delta - \theta_2)P$, the neoclassical user cost of capital. Taking first differences and accounting for depreciation, the dynamic "desired gross investment" expression may thus be written

$$\Gamma_i^*(t) = \frac{1}{1-\alpha}(\Delta \ln Y_i(t) - \Delta \ln r_{k_i}(t)) - \frac{\alpha}{1-\alpha}\Gamma_i(t) + \frac{1}{1-\alpha}\delta. \tag{25}$$

This relationship holds at the individual unit's as well as at the aggregate level (for the latter suppress the index i). Inasmuch as published data series are consistent with the theoretical aggregation process, they can be used in (24) to construct the frictionless capital stock around which the irreversible accumulation process oscillates. The line marked by crosses in Figure 3 plots a yearly gross investment rate series computed from (25) (the solid and dashed lines plot actual investment ratios and our model's fitted series, and are discussed below). We set $\alpha = 0{\cdot}10$, as is appropriate if the share of equipment capital in value added is 13% and the markup coefficient is 24%.[7] Recall that $Y(t)$ is the revenue of capital, thus it is proportional to Gross National Product (GNP) under Cobb–Douglas technology and isoelastic demand assumptions. We use the rate of GNP growth as a proxy for $\Delta \ln Y(t)$. We identify $K(t)$ with the stock non-residential private equipment, and we

7. The former corresponds to the average share of machinery in the U.S. during the period 1954–1986. The latter is in the low end of the estimates obtained with aggregate data (see e.g., Hall (1988)) and within the range of results obtained with more disaggregate data (see e.g. Domowitz *et al.* (1988)).

construct a neoclassical user-cost series $r_k(t)$ taking into account tax factors as well as fluctuations in the price of capital.[8] The empirical counterpart of equation (25) corresponds to the *hypothetical* investment rate that would be observed if disinvestment were possible at the individual units' level, *and* demand, prices and interest rates were those actually observed in the U.S. economy. In our partial equilibrium exercise, such a series simply summarizes the aggregate component of *partial equilibrium* investment problems of firms, and does not represent counterfactual general equilibrium experiments.

Before proceeding, we need to verify the realism of the constant drift and variance assumptions we made in our theoretical sections. The limited length of available time series make it difficult to test our assumption of homoskedastic, serially uncorrelated log-increments for the forcing processes. All we can say is that there is no evidence of dramatically different dynamics in the data: the estimated $\Gamma^*(t)$ series is statistically indistinguishable from white noise (the t-statistic of the first order serial correlation coefficient is only 1·3). The constant-elasticity functional form assumptions that yield (24) should of course be viewed as approximations to more general ones. To test the tightness of the approximation, we regress $k(t)=k^f(t)+\tilde{s}(t)$ on a constant, on $\ln Y(t)$, and on $\ln r_k(t)$. Under our assumptions, the coefficients of $\ln Y(t)$ and $\ln r_k(t)$ are, respectively, unity and minus unity. As $\tilde{s}(t)$ is stationary if our model is correctly specified (see Bertola and Caballero (1990)) and the observable series are all integrated of order one, this is a cointegrating regression. OLS estimates and test statistics are biased in small samples. However, as suggested by Stock and Watson (1989), these small-sample biases may be reduced if leads and lags of the first differences of the regressors are included so as to obtain residuals which are approximately orthogonal, at all leads and lags, to the integrated regressors. This testing procedure fails to reject the null hypothesis of interest: the coefficient of $Y(t)$ is 1·19 (standard error = 0·12), that of $r_k(t)$ is −0·87 (standard error = 0·39). These results indicate that the data we use are not inconsistent with our theoretical assumptions.

Idiosyncratic uncertainty and the character of actual investment

The actual investment/capital ratio $\Gamma(t)$, plotted in Figure 3 as a solid line, is clearly much less variable than the theoretical Jorgensonian construct. The standard deviation of the former is 0·017, that of the latter 0·046. The contemporaneous correlation between actual investment and desired investment is only 0·29. The first-order serial correlation coefficient of the $\Gamma^*(t)$ series is significantly different from zero (0·25 with a standard deviation of 0·19), while $\Gamma(t)$ exhibits substantially higher (0·68 with standard deviation 0·13) first-order serial correlation.

In previous research, these facts have been rationalized by postulating smooth and convex adjustment cost functions and have led researchers to estimate partial-adjustment equations of doubtful microeconomic realism. Irreversibility of investment decisions, like more familiar forms of adjustment costs, reduces the responsiveness of endogenous variables to exogenous shocks: at the individual unit's level gross investment is completely unresponsive to the forcing variables when the irreversibility constraint is binding. The extent to which microeconomic inaction affects aggregate dynamics depends on the degree

8. We measure cost of capital on the basis of the perfect-foresight present value of tax saving from investment credits series, $T(t)$, constructed by Auerbach and Hassett (1990). The ex-post/perfect foresight cost of capital measure is equal to $\ln[(r+\delta)(1-T(t))P_k(t)/((1-\tau(t))P(t))]$ where we have assumed that $(r+\delta)$ is constant and equal to 0·2 and $P_k(t)/P(t)$ is the ratio of National Account investment and GNP deflators. The corporate tax rate, $\tau(t)$, is also from Auerbach and Hassett. We construct an ex-ante counterpart to this cost-of-capital series by projecting it onto its currently observed components, $\ln[1-\tau(t)]$ and $\ln[P_k(t)/P(t)]$.

of synchronization of individual actions, which is in turn determined by the form of the adjustment policy on the one hand, and on the other by the importance of aggregate developments relative to that of idiosyncratic uncertainty (see Bertola and Caballero (1990) and Caballero and Engel (1991)).

In the problem we are considering, aggregate uncertainty is small relative to the drift: the sample mean of the gross investment/capital ratio, which approximates $\Theta + \delta$, is 0·16, almost four times as large as its standard deviation. At the *aggregate* level, then, the irreversibility constraint should almost never be binding if aggregate innovations are normally distributed as we assumed above. In fact, desired gross investment is always strictly positive in our sample. If idiosyncratic uncertainty were negligible, then, all units would be bunched in a spike at the $k_i(t) = k_i^d(t)$ point, and actual investment should track desired investment exactly. If σ_I is large, on the other hand, the irreversibility constraint is binding (for some of the units, in some of the periods), and changes in the cross-sectional distributions tend to smooth out the response of $\Gamma(t)$ to movements in $\Gamma^*(t)$.

It is indeed possible to match the volatility and serial correlation of the observed series by an appropriate choice of σ_I.[9] The dashed line in Figure 3 plots the aggregate investment path predicted by our model if $\sigma_I = 0{\cdot}6$, an admittedly rather extreme value for the variability of "desired capital" across microeconomic units and over time: in this simple implementation of our model, *all* of the time-series smoothing required by aggregate data must result from imperfect synchronization of microeconomic investments. In reality, of course, some forms of investment may be smoother than the relevant forcing variables because of convex adjustment costs, time-to-build distributed lags, and many other mechanisms emphasized by previous research. It is nevertheless interesting to find that the distributional mechanism we focus on is capable of smoothing the dynamics of the investment process, and of increasing the persistence of aggregate events' effects on capital accumulation, to the extent required by U.S. aggregate data. While the fit is far from perfect (the model explains 36% of the investment series's variability), raising σ_I from zero to 0·6 reduces the standard deviation of investment from 0·046 (the standard deviation of desired investment) to 0·015 (which is comparable to observed investment's standard deviation, 0·017); and raises its first-order serial correlation from 0·25, for desired investment, to 0·66 (comparable to the 0·68 correlation of observed investment).

On the results

These simple computations indicate that the smooth, persistent character of aggregate investment dynamics can be rationalized in terms of unit-level irreversibility constraints rather than in terms of ad hoc adjustment cost functions. Our model assigns to idiosyncratic uncertainty the crucial role of smoothing out the nonlinear, intermittent character of microeconomic irreversible investment decisions. The results indicate that volatility of desired capital at the individual unit's level would indeed have to be very high for the model to rationalize aggregate evidence without recourse to other sources of investment smoothness. With $\sigma_I = 0{\cdot}6$, the desired capital stock at the individual unit level should vary by as much as 60% about two-thirds of the cases, on a year-to-year basis. In

9. The parameter σ_I is reasonably well determined in this metric. Ratios of the predicted investment series's standard deviation to the observed one are 1·24, 1·06, 0·80 and 0·71, at $\sigma_I = 0{\cdot}4$, 0·5, 0·7 and 0·8. We pick $\sigma_I =$ 0.6 to match this moment, and the serial correlation coefficient as well. In principle, one could estimate σ_I by maximizing the fit of the $\hat{\Gamma}(t)$ series to the observed $\Gamma(t)$ series. When searching over large values of σ_I, the R^2 improves but the predicted series becomes too smooth. Monte Carlo experiments suggest that serially uncorrelated measurement errors may be responsible for this.

the model, symmetric shocks to other units' desired capital ensure that the degree of uncertainty is approximately ten times smaller at the aggregate level: in reality, investment does not become more intermittent and more volatile as more and more disaggregated data are considered. Given the rather abstract definition of "units" in our aggregative model, however, it is not easy to evaluate how close to 0·6 the proportional variability of desired capital might be in reality. Once again, the "desired capital" construct simply summarizes external influences on the firm's problem in our partial equilibrium setting, and its high volatility reflects the fact that much of these are specific to individual firms, product lines, and types of equipment. Of course, the volatility of *installed* capital would not necessarily be so high in a world without irreversibility constraints: if, counterfactually, capital were perfectly homogeneous and all investment decisions were reversible, the process followed by the price of new capital $P(t)$ would of course be quite different.

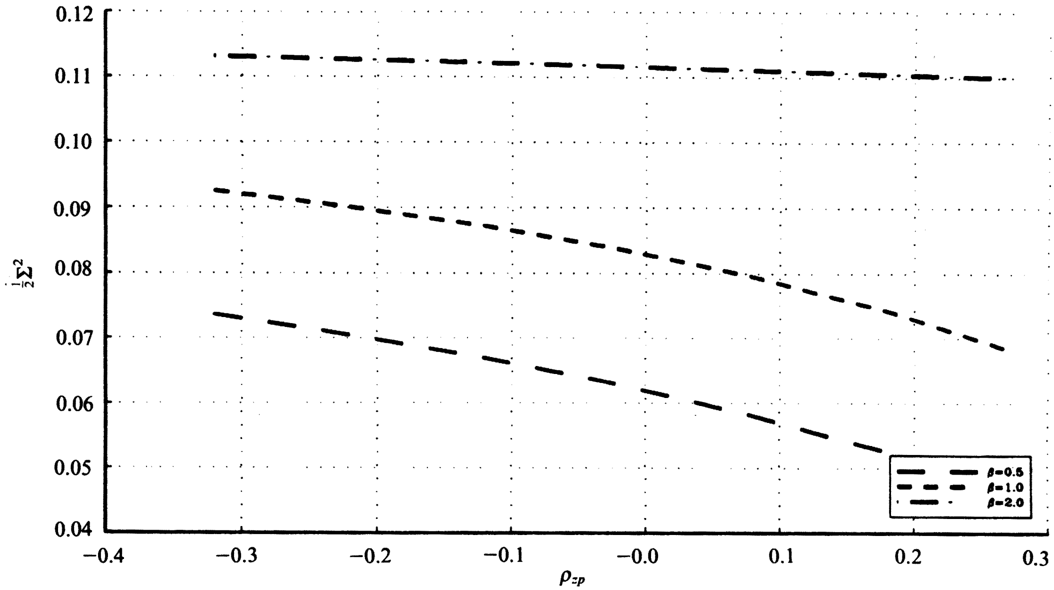
Since Section 2 provides a completely specified model of unit-level investment, we can characterize the microeconomic importance of investment irreversibility by specifying its parameters so as to match the observed rate of growth of capital and of capital costs as well as the degree of idiosyncratic uncertainty implied by the smoothness of aggregate investment. Aggregate data provide information for some, but not all of the parameters relevant to the microeconomic model of Section 2. The logarithmic rate of annual capital growth is 0·04 in the period we consider; with GNP (our proxy for Y) growing at an average rate of 3% per year and $\alpha=0{\cdot}1$, equation (1) implies that the "business conditions" index Z grows at the annual rate $\vartheta_1=2{\cdot}6\%$. We set $r+\delta=0{\cdot}20$ in our empirical work, reflecting the large 0·12 depreciation rate appropriate for the capital and investment series we use (U.S. business equipment investment) and an 8% required rate of return.

The mean rate of growth of the aggregate counterparts of Z and P are easily computed from the logarithmic variables used in our empirical work above. By equation (13), however, these estimates depend on the variance of the *idiosyncratic* P and Z processes. Absent convex adjustment costs (or time-to-build and information lags), the smoothness of aggregate investment series requires large idiosyncratic uncertainty:[10] with $\sigma_I=0{\cdot}6$, we have $\sum^2=((1-\alpha)\sigma_I)^2=0{\cdot}35$. While aggregate data provide information on $\sum\equiv(\sigma_1-\sigma_2)'(\sigma_1-\sigma_2)$, they do not identify σ_1 and σ_2 separately, as would be necessary for the purpose of obtaining estimates for ϑ_1 and ϑ_2 from (13) and the aggregate counterparts of P and Z. Since ϑ_1 and ϑ_2 have different roles in the microeconomic investment problem, two parameters remain to be chosen before aggregate evidence can be used to infer microeconomic investment policies: the correlation between the rates of growth of P_i and Z_i, and the relative variance of these two rates.

Figure 4 is constructed letting $\sigma_1=[\bar{\sigma}, \rho\bar{\sigma}]'$ and $\sigma_2=[\bar{\sigma}\beta\rho, \bar{\sigma}\beta]'$, where the scalar $\bar{\sigma}$ is pinned down by the estimated variance of dk_i^f, β indexes the volatility of P relative to that of Z, and ρ determines the correlation ρ_{zp} between the rates of growth of P_i and Z_i: $\rho_{zp}=2\rho/(1+\rho^2)$. In Figure 4(a) we plot numerical values from equation (12) for the excess of capital's marginal revenue product over its Jorgensonian user cost when investment is positive at the unit level; depending on what is assumed as to the relative variability of Z_i and P_i and as to the correlation of their increments, the excess required return varies between 4% and 12% per year. Figure 4(b) presents the corresponding ratio of desired

10. Note that we have assumed all units to have the same structural parameters, and the only source of comovement to be exogenous aggregate shocks. In a sense, structural heterogeneity plays a role very similar to that of exogenous idiosyncratic uncertainty (Caballero and Engel (1991)), while strategic interactions may exacerbate (strategic complementarities) or dampen (strategic substitutability) the relative importance of aggregate shocks (Caballero and Engel (1992)). These factors should be taken into account in order to interpret the magnitude of the sources of the uncertainty faced by individual units.

(a)



(b)

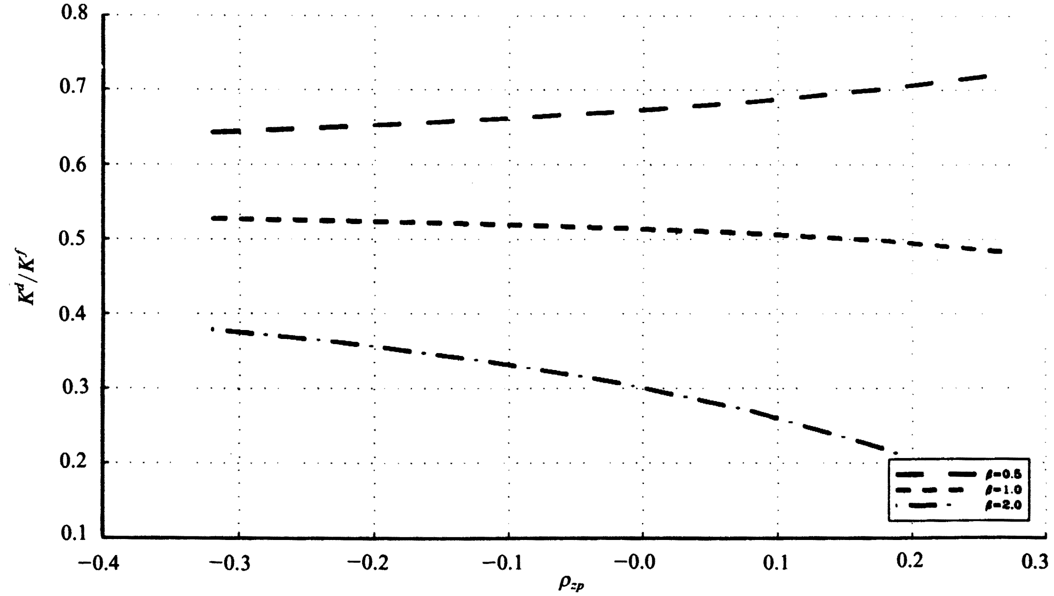


FIGURE 4

Required excess return at the investment point

(from equation (12)) to Jorgensonian stocks of capital at the investment points. This takes values between 0·2 and 0·7, to imply that if irreversibility is the only dynamic factor in real-life investment problems, as we assume, then the stock of capital ranges between 20% and 70% of the neoclassical "frictionless" construct at times when investment is positive at the unit level.

5. CONCLUSIONS

This paper proposes a closed-form sequential irreversible investment rule, studies its implications for aggregate investment behaviour when both idiosyncratic and aggregate sources of uncertainty are present, and provides some empirical evidence for the model using postwar U.S. private equipment investment data. We have shown that aggregate data are broadly consistent with a model in which microeconomic units rationally choose to install less capital than it would be implied by a frictionless neoclassical model when they invest, and allow the installed capital stock to depreciate when the irreversibility constraint becomes binding. In our model, microeconomic investment aims to keep installed capital close to a moving target which depends on the level of activity as well as on the cost of capital. The microeconomic irreversibility constraint, interacting with idiosyncratic sources of uncertainty, yields a smooth, highly persistent response of aggregate investment to innovations in activity and in the cost of capital.

Underlying such smooth aggregate processes are highly volatile, intermittent investment decisions by microeconomic "units". Such units may or may not correspond closely to observable sectors, firms, and types of capital: while such disaggregate empirical issues are beyond the scope of this paper, the notion of more nonlinear, infrequent, and volatile investment decisions at the microeconomic level accords well with casual empiricism and with the increasingly available hard evidence on plant-level data.[11]

Quite clearly, the results of this paper do not provide a complete interpretation of investment dynamics. The empirical section above shows that the combination of nonlinear investment policies and idiosyncratic uncertainty is capable of smoothing aggregate driving processes to the extent required by aggregate investment data. While the fit of our specification is satisfactory, a model where *only* this mechanism is present leaves unexplained a non-trivial and serially correlated error component. Future research should explore the role of time-to-build lags, of other non-convexities, and of quasi-fixed factors other than capital. Each of these realistic features complicates the three tasks tackled in this paper—microeconomic optimization, stochastic aggregation, and empirical inferences. Our model of aggregate investment addresses the three issues in a self-consistent way, with reasonably sound microeconomic foundations, and may form the basis for further applied work.

APPENDIX A

The Bellman equation for the irreversible investment problem takes the form

$$rV(K(t), Z(t), P(t))dt = \max_{dG(t)} [K(t)^{\alpha} Z(t)dt - P(t)dG(t) + E_t\{ V(K(t), Z(t), P(t))\}],$$
$$\text{subject to } dG(t) \geqq 0, \tag{A1}$$

at all times t. Under the usual regularity conditions, Itô's change-of-variable formula yields (omitting time indexes and the arguments of $V(\cdot)$):

$$\begin{aligned} E_t\{dV(\cdot)\} &= E_t\{\partial_K V(\cdot)dK + \partial_Z V(\cdot)dZ + \partial_P V(\cdot)dP \\ &\quad + \tfrac{1}{2}\partial_{zz} V(\cdot)(dZ)^2 + \tfrac{1}{2}\partial_{PP} V(\cdot)(dP)^2 + \partial_{PZ} V(\cdot)(dZ)(dP)\} \\ &= \partial_K V(\cdot)(-\delta K dt + dG) + \partial_Z V(\cdot) Z \vartheta_1 dt + \partial_P V(\cdot) P \vartheta_2 dt \\ &\quad + \tfrac{1}{2}(\partial_{ZZ} V(\cdot) Z^2 \sigma_1' \sigma_1 dt + \Sigma_{PP} V(\cdot) P^2 \sigma_2' \sigma_2 dt + \partial_{PZ} V(\cdot) PZ(\sigma_1' \sigma_2 + \sigma_2' \sigma_1) dt) \end{aligned} \tag{A2}$$

11. Ongoing work by Mark Doms and Timothy Dunne at the U.S. Census Bureau suggests that over 50 percent of a plants' cumulative equipment-investment over a period of 15 years is concentrated in a period of three (contiguous) years.

where $\partial_{XY}f(x, y)$ denotes the partial derivative of a function $f(\cdot)$ with respect to X and Y, evaluated at $X=x$, $Y=y$.

Using (A2) in (A1), optimal irreversible investment should satisfy the complementary slackness conditions

$$\begin{aligned} \partial_K V(K(t), Z(t), P(t)) &\leqq P(t) \quad \forall t; \\ \partial_K V(K(t), Z(t), P(t)) &= P(t) \quad \forall t : dG(t) > 0 \end{aligned} \tag{A3}$$

If $\partial_{KK} V(\cdot)$ exists and is not zero, the second line of (A3) implicitly defines $K^d(Z(t), P(t))$.

Using (A3) in (A1) we find that

$$\begin{aligned} rV(\cdot) = K^{\alpha}Z + \partial_K V(\cdot)(-\delta K) + \partial_Z V(\cdot)Z\vartheta_1 + \partial_P V(\cdot)P\vartheta_2 \\ + \tfrac{1}{2}(\partial_{ZZ} V(\cdot)Z^2\sigma_1'\sigma_1 + \partial_{PP} V(\cdot)P^2\sigma_2'\sigma_2 + \partial_{PZ} V(\cdot)PZ(\sigma_2'\sigma_2 + \sigma_2'\sigma_1)) \end{aligned} \tag{A4}$$

holds identically along the optimal path, and can be differentiated term-by-term with respect to K. Defining $v(K, Z, P) \equiv \partial_K V(K, Z, P)$, we obtain

$$\begin{aligned} (r+\delta)v(\cdot) = \alpha K^{\alpha-1}Z + \partial_K v(\cdot)(-\delta K) + \partial_Z v(\cdot)Z\vartheta_1 + \partial_P v(\cdot)P\vartheta_2 \\ + \tfrac{1}{2}\partial_{ZZ} v(\cdot)Z^2\sigma_1'\sigma_1 + \tfrac{1}{2}\partial_{PP} v(\cdot)P^2\sigma_2'\sigma_2 + \partial_{PZ} v(\cdot)PZ(\sigma_1'\sigma_2 + \sigma_2'\sigma_1) \end{aligned} \tag{A5}$$

A particular solution to (A5) is

$$v_0(K, Z, P) = \frac{\alpha K^{\alpha-1}Z}{r+\alpha\delta-\vartheta_1},$$

and $v(\cdot)$ solutions can be written as linear combinations of $v_0(\cdot)$ and terms in the form $(\alpha K^{\alpha-1}Z)^A P^{1-A}$, for A a solution to the characteristic equation

$$\tfrac{1}{2}\textstyle\sum^2 x^2 + (\vartheta_1 + (1-\alpha)\delta - \vartheta_2 - \tfrac{1}{2}\sum^2)x - (r+\delta-\vartheta_2) = 0 \tag{A6}$$

(recall that $\sum^2 \equiv (\sigma_1-\sigma_2)'(\sigma_1-\sigma_2)$.) If the parameters satisfy condition (9) in the main text, this quadratic equation has two roots of opposite sign. Since $Z=0$ is absorbing for the $\{Z(t)\}$ process, it must be the case that $\lim_{Z\to 0} v(K, Z, P) = 0$, to imply that only the positive root of (A6) need be considered. Accordingly, we have

$$v(K, Z, P) = \frac{\alpha K^{\alpha-1}Z}{r+\alpha\delta-\vartheta_1} + C(\alpha K^{\alpha-1}Z)^A P^{1-A} \tag{A7}$$

for $K \geqq K^d(Z, P)$, where A denotes the positive root of (A6), and C is a constant of integration that does not depend on K, Z, or P.

By (A3), the investment policy prevents $v(K, Z, P)$ from ever exceeding P. Thus, $v(\cdot, \cdot, \cdot)$ must satisfy the boundary condition

$$v(K^d(Z, P), Z, P) = P. \tag{A8}$$

For the differential relationship (A5) to be satisfied in the immediate neighbourhood of the locus where (A8) holds, the value function must be twice-differentiable there. This yields the "smooth pasting" conditions

$$\begin{aligned} \partial_K v(K^d(Z, P), Z, P) &= 0, \\ \partial_P v(K^d(Z, P), Z, P) &= 1, \\ \partial_Z v(K^d(Z, P), Z, P) &= 0. \end{aligned} \tag{A9}$$

Using (A7) in (A8) and (A9) we find

$$\frac{\alpha(K^d(Z, P))^{\alpha-1}Z}{P} = \frac{A}{A-1}(r+\alpha\delta-\vartheta_1).$$

Rearranging (A6), we may write

$$\frac{A}{A-1} = \frac{r+\delta-\vartheta_2+\sum^2 A}{r+\alpha\delta-\vartheta_1}$$

to imply that the desired-capital function has the form given as equations (11), (12) in the main text.

It can be verified that $A > 1/(1-\alpha)$ if condition (9) in the main text holds true, and that $c = r+\delta-\vartheta_2$ only if $\sum^2 = 0$ and $\vartheta_1 - (1-\alpha)\delta - \vartheta_2 > 0$. This would be a degenerate special case of Arrow's (1968) non-stochastic model, in which disinvestment is never desirable and the irreversibility constraint is completely irrelevent.

APPENDIX B

The purpose of this appendix is to describe the procedure we use to estimate the path of the cross-sectional density and its mean. Since continuous information is not available as to aggregate developments, we assume the realizations of aggregate uncertainty to be evenly spread within each observation period. Namely, if we can infer from aggregate data that the *average* desired stock of capital increases by $x\%$ between t and $t+h$, then we model aggregate dynamics *as if* the increase occured at a constant rate x/h in the continuous time interval between observations. This is only an approximation, of course. Investment being irreversible, the time-aggregated investment rate is path-dependent and the variability of desired capital at higher frequencies is, in principle, relevant for the observed path of installed capital. The approximation also neglects the infinite variation property of Brownian paths; we believe, however, that any empirical importance of these issues is overshadowed by the substantial simplification of the analytical and estimation problems: it reduces an intractable stochastic partial differential equation to a sequence of deterministic linear partial differential equations, whose solution is presented below.

Specifically, let observations on desired and actual capital stocks be available at the times $(t_0, t_1, t_2, \ldots)$, with $t_h - t_{h-1} = \Delta t$, $h = 0, 1, 2, \ldots$. The discrete counterpart of equation (18) is then

$$\Gamma(t_h) = \Gamma^*(t_h) + \Delta\tilde{s}(t_h)$$

where $\Gamma(t_h)$ and $\Gamma^*(t_h)$ denote, respectively, the observed and desired ratios of gross investment to capital between t_{h-1} and t_h.

We now find the (continuous) path of the cross-sectional density between t_{h-1} and t_h, momentarily omitting the time index for simplicity.

We first derive the dynamic density of a Brownian motion with reflecting barriers at zero and at $S>0$, and we then take the appropriate limit as $S \to \infty$.

Finite state space

Let $f(s, t)$ denote the probability density of a process $s(t)$ with stochastic differential

$$ds(t) = \vartheta dt + \sigma dW(t), \qquad \vartheta < 0, \sigma > 0$$

where $\{W(t)\}$ a standard Wiener process, and let $\{s\}$ be reflected at 0 and at $S>0$. The function $f(s, t)$ can be derived by solving the forward Kolmogorov equation

$$\partial_t f(s, t) = \tfrac{1}{2}\sigma^2 \partial_{ss} f(s, t) - \vartheta \partial_s f(s, t), \tag{B1}$$

with boundary conditions

$$\tfrac{1}{2}\sigma^2 \partial_s f(0, t) = \vartheta f(0, t) \quad \forall t, \tag{B2}$$

$$\tfrac{1}{2}\sigma^2 \partial_s f(S, t) = \vartheta f(S, t) \quad \forall t, \tag{B3}$$

and given initial condition

$$f(s, 0) = \bar{g}(s), \qquad \int_0^S \bar{g}(s) ds = 1. \tag{B4}$$

Separating the variables, we write $f(s, t) = g(s)h(t)$ and obtain a couple of ordinary differential equations. In the t direction,

$$h'(t) + \lambda h(t) = 0$$

has general solution $h(t) = Ae^{-\lambda t}$, A a constant of integration. In the s direction,

$$g''(s) + \xi g'(s) - \lambda \frac{\xi}{\vartheta} g(s) = 0 \tag{B5}$$

$$g'(0) = -\xi g(0) \tag{B6}$$

$$g'(S) = -\xi g(S) \tag{B7}$$

where $\xi \equiv -2\vartheta/\sigma^2$, $\xi > 0$.

Equations (B5–B7) define a Sturm–Liouville problem (see e.g. Churchill and Brown (1987)), with characteristic equation

$$\alpha^2+\xi\alpha-\frac{\lambda}{\vartheta}\xi=0.$$

If $\lambda\leqq-\xi\vartheta/4=\xi^2\sigma^2/8$, the roots are real and solutions taken the general form

$$g(s)=A_1e^{\alpha_1 s}+A_2e^{\alpha_2 s}. \tag{B8}$$

Solutions in this form need to be considered only if they can satisfy the boundary conditions with A_1 and/or A_2 different from zero: (B6) and (B7) require

$$A_1(\alpha_1+\xi)+A_2(\alpha_2+\xi)=0; \qquad A_1e^{\alpha_1 S}(\alpha_1+\xi)+A_2e^{\alpha_2 S}(\alpha_2+\xi)=0.$$

All solutions in the form (B8) are then identically zero except the one corresponding to $\lambda=0$, with $\alpha_1=-\xi$, $\alpha_2=0$, $A_2=0$:

$$g(s;\lambda=0)=Ae^{-\xi s}.$$

We then consider the solutions obtained for $\lambda>\xi^2\sigma^2/8$. The roots are complex, and the solution has the form

$$g(s;\lambda)=e^{\frac{-\xi s}{2}}(A\cos(\beta(\lambda)s)+B\sin(\beta(\lambda)s)) \tag{B9}$$

where

$$\beta(\lambda)=\sqrt{-\xi\left(\frac{\lambda}{\vartheta}+\frac{\xi}{4}\right)}.$$

Imposing (B6), we obtain

$$\frac{\xi}{2}A+\beta(\lambda)B=-\xi A$$

to imply that $B=-A\xi/2\beta$; for (B7) to be satisfied, we then need

$$A\left(\frac{\xi^2}{4\beta(\lambda)}-\beta(\lambda)-\frac{\xi^2}{2\beta(\lambda)}\right)\sin(\beta(\lambda)S)=0.$$

Using the definition of β and simplifying, we get

$$A\xi\frac{\lambda}{\vartheta}\sin(\beta(\lambda)S)=0.$$

Thus, all solutions in the form of (B9) are identically zero except those for which $\sin(\beta(\lambda)S)=0$, or

$$-\xi\left(\frac{\lambda}{\vartheta}+\frac{\xi}{4}\right)=\frac{n^2\pi^2}{S^2}, \qquad n=1,2,\ldots.$$

Combining the results, we find that the general solution to (B1–B3) can be written

$$f(s,t)=\sum_{n=0}^{\infty}A_nf_n(s)e^{-\lambda_n t}$$

$$\lambda_0=0; \qquad \lambda_n=\frac{\sigma^2}{2}\left(\frac{n^2\pi^2}{\bar{s}^2}+\frac{\vartheta^2}{\sigma^4}\right), \qquad n=1,2,\ldots$$

$$f_0(s)=e^{-\xi s}; \qquad f_n(s)=e^{\frac{-\xi s}{2}}\left(\cos\left(\frac{n\pi}{S}s\right)-\frac{S}{n\pi}\frac{\xi}{2}\sin\left(\frac{n\pi}{S}s\right)\right), \qquad n=1,2,\ldots.$$

The initial condition

$$\sum_{n=0}^{\infty}A_nf_n(s)=\bar{g}(s) \tag{B10}$$

determines the constants A_k, $k=0, 1, 2, \ldots$. Multiplying both sides of (B10) by $f_k(s)e^{\xi s}$, integrating between 0 and S, and exploiting the fact that $\int_0^S f_n(s)f_k(s)e^{\xi s}ds=0$ for $n\neq k$, we obtain

$$A_0=\xi/(1-e^{-\xi s})$$

$$A_k=\frac{\int_0^S \bar{g}(s)f_k(s)e^{\xi s}ds}{\int_0^S f_n(s)^2e^{\xi s}ds},\qquad k=0, 1, 2, \ldots. \tag{B11}$$

The integral in the numerator of (B11) can be computed numerically for general $\bar{g}(\cdot)$, and the denominator has the closed form

$$\int_0^S f_n(s)^2e^{\xi s}ds=\left(\frac{S}{2}+\frac{S^3\xi^2}{8n^2\pi^2}\right).$$

We note at this point that as $n\to\infty$ the constants converge to zero, and the λs diverge to positive infinity. Thus, truncation of the infinite series yields a satisfactory approximation to $f(s, t)$.

Unbounded state space

Given that $\xi>0$,

$$\lim_{S\to\infty} f(S, t)=0$$

and the boundary condition corresponding to (B6) is satisfied identically. All solutions in the form

$$f(s;\beta)=e^{\frac{-\xi s}{2}}\left(\cos(\beta s)-\frac{\xi}{2\beta}\sin(\beta s)\right)$$

and therefore admissible, as well as $f_0(S)$ defined above, and the solution of the PDE takes the integral form

$$f(s, t)=A_0f_0(s)+\int_{0^+}^{\infty} A(\beta)e^{-\lambda(\beta)t}f(s;\beta)d\beta,$$

where $\lambda(\beta)=\beta^{-1}(\lambda)$. In the limit as $S\to\infty$, the expressions for $A(\beta)$ and for the mean of the density take the form:

$$A(\beta)=\frac{2}{\pi(1+\xi^2/4\beta^2)}\int_0^{\infty} f(s, 0)e^{\frac{-\xi s}{2}}\left(\cos(\beta s)-\frac{\xi}{2\beta}\sin(\beta s)\right)ds,$$

and

$$\bar{s}(t)=\frac{1}{\xi}-\int_{0^+}^{\infty}\frac{A(\beta)e^{-\lambda(\beta)t}}{\beta^2(1+\xi^2/4\beta^2)}d\beta.$$

Dynamic recursion

We now bring back the time index. Defining

$$\xi_h=-\frac{2\Gamma^*(t_h)}{\sigma_I^2},$$

we can treat the cross-sectional density at the end of each period as the initial condition for the next period, and compute cross-sectional densities at each observation point by the following recursive relationships:

$$f(s, h)=\xi_he^{-\xi_h s}+e^{-(1/2)\xi_h s}\int_{0^+}^{\infty} A(\beta;h)e^{-\lambda(\beta)t}\left(\cos(\beta s)-\frac{\xi h}{2\beta}\sin(\beta s)\right)d\beta \tag{B12}$$

$$A(\beta;h)=\frac{2}{\pi(1+\xi_h^2/4\beta^2)}\int_0^{\infty} f(s, h-1)e^{(1/2)\xi_h s}\left(\cos(\beta s)-\frac{\xi_h}{2\beta}\sin(\beta s)\right)ds. \tag{B13}$$

If the investment/capital ratio were constant over time, as would be the case if $\sigma_A=0$, then we would have $\xi_h=\xi_{h-1}$ for all h, and the recursion would track at discrete times the convergent path of the cross-sectional density to its stable form. If aggregate investment fluctuates over time, however, the ξ_h values relevant to each observation are different, and the recursion generates a sequence of densities linked by initial and final conditions.

The change in the mean of the cross-sectional density, $\Delta\tilde{s}(t_h)$, is then readily computed from equation (B12):

$$\Delta\tilde{s}(t_h)=\frac{1}{\xi_h}-\int_{0^+}^{\infty}\frac{A(\beta,h)e^{-\lambda(\beta)t}}{\beta^2(1+\xi_h^2/4\beta^2)}d\beta-\left(\frac{1}{\xi_{h-1}}-\int_{0^+}^{\infty}\frac{A(\beta,h-1)e^{-\lambda(\beta)t}}{\beta^2(1+\xi_{h-1}^2/4\beta^2)}d\beta\right). \quad \text{(B14)}$$

This is the expression which, in conjunction with the estimates of Γ^*, yields an estimated path of the investment/capital ratio as a function of idiosyncratic uncertainty and other parameters of the model.

Acknowledgements. We are indebted to seminar participants, colleagues, and referees for helpful comments; to Kevin Hassett for lending us his data; and to Hyeng Keun Koo for very competent research assistance supported by the John M. Olin Program for the Study of Economic Organization and Public Policy. Financial support from the National Science Foundation is gratefully acknowledged. Ricardo Caballero also thanks the Alfred P. Sloan Foundation and the NBER (Olin Fellowship) for financial support.

REFERENCES

ABEL, A. B. and BLANCHARD, O. J. (1986), "The Present Value of Profits and Cyclical Movements in Investment", *Econometrica*, **54**, 249–273.

ARROW, K. J. (1968), "Optimal Capital Policy with Irreversible Investment", in J. N. Wolfe (ed.), *Value, Capital and Growth. Papers in Honour of Sir John Hicks* (Edinburgh: Edinburgh University Press).

AUERBACH, A. J. and HASSETT, K. (1990), "Tax Policy and Business Fixed Investment in the United States" (Working paper, University of Pennsylvania).

BENTOLILA, S. and BERTOLA, G. (1990), "Firing Costs and Labour Demand: How Bad Is Eurosclerosis?", *Review of Economic Studies*, **57**, 381–402.

BERTOLA, G. (1988), "Adjustment Costs and Dynamic Factor Demands: Investment and Employment under Uncertainty" (Ph.D. dissertation, M.I.T.).

BERTOLA, G. and CABALLERO, R. J. (1990), "Kinked Adjustment Costs and Aggregate Dynamics", in Blanchard, O. J. and Fischer, S. (eds.) *NBER Macroeconomic Annual 1990,* (Cambridge (MA):MIT Press).

BILLINGSLEY, P. (1986) *Probability and Measure*, *2nd ed.* (New York: John Wiley & Sons).

CABALLERO, R. J. and ENGEL, E. M. R. A. (1991), "Dynamic (*S*, *s*) Economies", *Econometrica*, **59**, 1659–1686.

CABALLERO, R. J. and ENGEL, E. M. R. A. (1993), "Heterogeneity and Output Fluctuations in a Dynamic Menu-Cost Economy", *Review of Economic Studies*, **60**, 95–119.

CHOW, P.-L., MENALDI, J. L. and ROBIN, M. (1985), "Additive Control of Stochastic Flow Systems with Finite Horizon", *SIAM Journal of Control and Optimization*, 858–899.

CHURCHILL, R. V. and BROWN, J. W. (1987) *Fourier Series and Boundary Value Problems, 4th ed.* (New York: McGraw-Hill).

DEMERS, M. (1991), "Investment under Uncertainty, Irreversibility and the Arrival of Information Over Time", *Review of Economic Studies*, **58**, 333–350.

DIXIT, A. (1992), "Investment and Hysteresis", *Journal of Economic Perspectives*, **6**, 107–132.

DOMOWITZ, I. R., HUBBARD, R. G. and PETERSEN, B. (1988), "Market Structure and Cyclical Fluctuations in U.S. Manufacturing", *Review of Economics and Statistics*, **70**, 55–66

HAYASHI, F. (1982), "Tobin's Marginal *q* and Average *q*: A Neoclassical Interpretation", *Econometrica*, **50**, 213–224.

JORGENSON, D. W. (1963), "Capital Theory and Investment Behavior", *American Economic Review, Papers and Proceedings*, **53,** 247–259.

KARATZAS, I. and SHREVE, S. E. (1984), "Connections between Optimal Stopping and Singular Stochastic Control I: Monotone Follower Problem", *SIAM Journal of Control and Optimization*, **22**, 856–877.

KARATZAS, I. and SHREVE, S. E. (1988) *Brownian Motion and Stochastic Calculus* (New York: Springer-Verlag).

KOBILA, T. O. (1991), "Partial Investment Under Uncertainty", in Lund and Oksendal (eds), *Stochastic Models and Option Values* (Amsterdam: North-Holland).

KRYLOV, N. V. and ROZOVSKII, B. L. (1977), "On the Cauchy Problem for Linear Stochastic Partial Differential Equations", *Mathematics USSR Izvestija*, **11**, 1267–1248.

LUCAS, R. E. Jr. and PRESCOTT, E. C. (1971), "Investment Under Uncertainty", *Econometrica*, **39**, 659–681.

McDONALD, R. and SIEGEL, D. (1986), "The Value of Waiting to Invest", *Quarterly Journal of Economics*, **101**, 707–727.

NICKELL, S. (1974), "On the Role of Expectations in the Pure Theory of Investment", *Review of Economic Studies,* **41**, 1–19.

OLSON, L. J. (1989), "Stochastic Growth with Irreversible Investment", *Journal of Economic Theory*, **47**, 101–129.

PINDYCK, R. S. (1988), "Irreversible Investment, Capacity Choice, and the Value of the Firm", *American Economic Review*, **78**, 969–985.

PINDYCK, R. S. (1991), "Irreversibility, Uncertainty, and Investment", *Journal of Economic Literature*, **29**, 1110–1149.

SARGENT, T. J. (1979), "Tobin's Q and the Rate of Investment in General Equilibrium", in K. Brunner and A. H. Meltzer (eds.), *On the State of Macroeconomics* (Amsterdam: North Holland).

SARGENT, T. J. (1987) *Macroeconomic Theory, Second Edition* (New York: Academic Press).

STOCK, J. H. and WATSON, M. W. (1989), "A Simple MLE of Cointegrating Vectors in Higher Order Integrated Systems" (NBER Technical Working Paper No.83, December).