ADJUSTMENT COSTS AND DYNAMIC FACTOR DEMANDS: INVESTMENT AND EMPLOYMENT UNDER UNCERTAINTY

by

GIUSEPPE BERTOLA

Laurea in Economia e Commercio,
Università di Torino (1983)

Submitted to the Department of Economics
in Partial Fulfillment of the Requirements
of the Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 1988

Signature of Author ______________________________
Department of Economics
April, 1988

Certified by ______________________________
Rudiger Dornbusch
Professor of Economics

Accepted by ______________________________
Professor Peter Temin, Chairman
Departmental Graduate Committee
Department of Economics

ADJUSTMENT COSTS AND DYNAMIC FACTOR DEMANDS:
INVESTMENT AND EMPLOYMENT UNDER UNCERTAINTY

by

GIUSEPPE BERTOLA

Submitted to the Department of Economics on April 28, 1988
in partial fulfillment of the requirements
of the Degree of Doctor of Philosophy.

ABSTRACT

Chapter 1 studies solution techniques for problems of dynamic control under uncertainty with linear costs of control. Necessary conditions for optimality of control policies are derived from a feasible perturbation argument, and it is shown that under restrictive conditions the optimal policy can be based on current events only. A solution is explicitly derived under the assumption of constant-elasticity functional forms and of uncertainty described by geometric Brownian motion processes. Under these assumption, an alternative approach to the solution is proposed, based on optimal stopping arguments: using well-known financial techniques, the solution can be found via valuation of options to exercise control at the margin.

Chapter 2 applies the control technique to the problem of irreversible capital accumulation. Under the realistic assumption that capital equipment has no value unless used in production, the optimal investment rule is derived in closed form. It is found that the degree of uncertainty facing the firm is an important determinant of the irreversible investment decision: the more uncertain are future business conditions and the more variable is the purchase price of capital, the more cautious firms should be in their investment decisions. The dynamics of the firm's value and the ergodic distribution of several observable variables are derived, and a preliminary discussion is offered of the results' implications for the empirical study of investment.

Chapter 3 studies the effect on labor demand of European severance pay legislation. The form of firms' employment policies is derived using the techniques developes in Chapter 1; firms are more reluctant to hire in the presence of firing costs, but are more reluctant to fire as well. It is found that employment is, on average, higher when firing costs are large. The parameters of stochastic processes taken as exogenous by firms are shown to affect the employment policy in intuitive ways, and the European employment experience of the last fifteen years is interpreted under the assumption that a regime switch in the stochastic environment of European firms occurred after the first oil shock.

Thesis Supervisor: Rudiger Dornbusch
Title: Professor of economics

ACKNOWLEDGMENTS

In these four years at MIT I have learned a lot about economics, about myself and about life. I am very grateful to my primary advisors: Rudiger Dornbusch saw me through the confusing initial attempts at original research, and during the past two years gave me much needed support and encouragement, as well as always insightful advice and sense of direction; Olivier Blanchard was always available and interested at what I did: he helped me a great deal to focus on substance and intuition, while his interest in the technical aspects of my work spurred major advances. I had much more interaction than is usually the case with the third reader of this thesis, Robert Pindyck; his work inspired most of what I have done, and he was always helpful and available for illuminating discussions.

I was fortunate to be Daniel McFadden's teaching assistant for two semesters; from him I learned most of what formal rigor I have. I consider a privilege to have been able to work a little with Franco Modigliani, whom I thank for his warm interest in my progress and for always insightful conversations.

Countless other individuals have contributed to the development of this thesis, both at MIT and at the institutions I visited while looking for a job. But the special debt I owe to my classmates and friends goes beyond it. Ricardo Caballero was an essential part of my life at MIT, both at work and at leisure. Anil Kashyap and Takeo Hoshi were wonderful officemates, generous with their time, entertaining and wise. I enjoyed discussing with Samuel Bentolila our joint work (which became the third chapter of this thesis) and the meaning of it all. I am sure we'll all remain in touch with each other and with the other friends we made at MIT.

Financial support for part of my years at MIT was provided by Istituto Bancario San Paolo di Torino and by Fondazione Luigi Einaudi di Torino, to whom I am most grateful.

I owe the greatest debt to Clara, my wife. She makes me a much better person and helps me in so many ways; without her, it would have been impossible to go through all this. Thanks.

INTRODUCTION

This Thesis studies the optimal dynamic factor demand policies of firms subject to exogenous uncertainty, under the assumption that changes in the use of factors of production incur first-order adjustment costs: the combination of uncertainty and first-order costs of adjustment is on the one hand quite realistic, and on the other has far-reaching implications for the study of many issues in dynamic economics.

If adjustment costs for the use of factors of production did not exist, the firm's dynamic problem would be uninteresting: all factors would continuously be adjusted so that their marginal contribution to profits would at all times be equal to their rental cost. In reality, of course, the use of factors of production cannot be costlessly adjusted: machinery and buildings have to be installed and uninstalled, workers have to be trained, and severance payments often have to be paid to dismissed workers; second-hand capital equipment has much lower value than new capital equipment, and is often so specialized that it can only be sold to other firms faced by the same exogenous uncertainty. In such a situation, capital has value only if used in production, and capital accumulation is irreversible.

Realistic adjustment costs are nonnegligible even for small adjustments in factors' use; the assumption that they are in fact linear does less violence to reality than the more usual (and more easily tractable) assumption of quadratic adjustment costs.

Linearity of adjustment costs enhances the importance of uncertainty in the firm's problem. Since adjustment entails first-order costs, the firm has to be careful in exercising control over the amount of factors of production it uses. If an exogenous shock is immediately followed by one of opposite sign, the firm will congratulate itself if it has not adjusted to the first shock, and will regret the previous decision if it has; conversely, if subsequent shocks have the same sign as the first one, the firm will regret not having adjusted right away. Every adjustment decision must then balance these possibilities, and has to take explicit account of risk. The dynamics produced by linear adjustment costs are very different from those produced by convex ones. Convex costs of adjustment make it optimal to adjust only partially to any shock; with linearity, adjustment is complete if it occurs, but may not occur at all.

Chapter 1 illustrates a set of new techniques that make it possible to solve fairly complicated and realistic models of this type, and discuss their relationship to earlier economic and financial literature and to abstract optimal control models in the mathematical and engineering literature.

The main assumptions needed to obtain the solution are that exogenous uncertainty be described in continuous time by a process with independent increments and continuous sample paths, and that all functions relating exogenous and endogenous variables have constant-elasticity form. The assumption of independent increments (random walk) is clearly a simplification of reality. But, at least, nonstationarity of the stochastic environment of economic agents can be defended on theoretical grounds (uncertainty about the future should realistically increase with the forecast horizon) and, on empirical grounds, cannot be refuted by the limited time-series data available: tests on most economic data fail to reject nonstationarity of the underlying stochastic processes. The assumption of constant elasticity (loglinear) functional forms is consistent with much empirical literature, and allows construction of fairly complicated models of the firm: problems with multiple state variables can be reduced to equivalent problems with only one state variable.

Chapter 1 also shows how optimal risk taking techniques (option valuation, optimal stopping) can be used to solve stochastic control problems under first-order adjustment costs. As noted above, every adjustment entails some risk of regret if there is ongoing uncertainty, and optimal stopping techniques indicate how such risky decisions should be taken.

The following two Chapters apply the control technique to economically interesting problems.

Chapter 2 studies the dynamics of irreversible capital accumulation, and the implications of irreversibility for empirical studies of investment: the assumption that the scrap value of capital be negligible is certainly realistic at the macroeconomic level, since production facilities have no direct consumption value; and is a very close approximation for an individual firm's problem, since capital equipment is usually specific to a firm's needs and has little (if any) resale value. The investment rule has a closed form under the assumption that the cost of adjusting capital use downward be prohibitive, and has intuitive comparative statics properties: firms will be more reluctant to invest if their environment is very uncertain. The implications of optimal irreversible investment decisions for observable quantities are also derived, and it is found (perhaps surprisingly) that higher uncertainty implies that on average more capital will ex-post be used if uncertainty is larger and investment is irreversible. Firms are more reluctant to invest in such a situation, but the impossibility of decumulating capital builds a ratchet in the accumulation process and increases the long-run capital intensity of production.

Chapter 2 then discusses the implications of investment irreversibility for the behavior of observable variables such as the value of the firm, investment and Tobin's Q: the dynamics of all variables are non-standard, and exogenous shocks can potentially have very long-lasting consequences. The results of the Chapter are arguably consistent with the empirical evidence based on more standard dynamic and pseudo-dynamic models of investment. Dynamic models based on convex costs of adjustment are mispecified if investment is irreversible in reality, and their poor empirical performance is therefore not surprising, but it is possible to interpreted their results under the assumption of investment irreversibility. Quasi-static models of investment, based on the asssumption of equality (on average) of capital's marginal profitability to its user cost, are also mispecified if capital accumulation is in fact irreversible: and irreversibility can help explain recent empirical results in this strand of literature.

Chapter 3 (joint work with Samuel Bentolila) applies the optimization model of Chapter 1 to dynamic labor demand in the presence of hiring, and especially firing, costs. The European unemployment experience has often been partly blamed on restrictive severance pay legislation, which appears to explain well some features of the dismal employment-creation record of most European countries. A formal model shows that firms will exercise more caution in their employment policies if labor-force adjustment costs are large and the environment is highly uncertain; the characteristics of the firm's employment policies are also related to other parameters, notably the attrition rate of the employed labor force and the growth rate of desired production. The implications of high firing costs for observed employment are also derived: it is found that the size of firing costs scarcely affects the average level of employment in the long run (and, via the same ratchet effect found for irreversible capital accumulation, larger firing costs <u>increase</u> average employment); but large firing costs have clear dynamic effects, inducing sluggish adjustment of employment to exogenous events. The implications of the model are quantitatively evaluated with realistic parameters values for the four large European economies: the change in firms' stochastic environment that followed the first oil shock, and the presence of high firing costs, are found to be by and large consistent with the observed behavior of employment in those economies, characterized by very little hiring and firing, but large reductions in employment via labor attrition.

CHAPTER 1

DYNAMIC STOCHASTIC PROGRAMMING UNDER LINEAR ADJUSTMENT COSTS

This Chapter discusses the solution of a class of discounted dynamic control problems, distinguished by the fact that the cost of adjusting endogenous state variables is linear in their displacement - i.e., the marginal cost of adjustment does not depend on the speed at which adjustment occurs. Problems of this type have been studied in the Operation Research literature; economic applications to the theory of the firm could include the study of inventory processes, pricing in the presence of menu costs, irreversible investment decisions, and dynamic labor demand. Two such applications are considered in the following Chapters, and the set of techniques proposed in this Chapters should prove useful in future research as well.

The techniques presented below are not totally new, although their application to economically interesting functional forms is. The treatment privileges economic intuition over abstract technical rigor: the thought experiments (feasible perturbations) underlying optimality conditions are described in economic terms, and the equivalence among the different approaches should eventually become clear to economically minded readers.

Section 1 presents the general form of the problem under study; Section 2 proves necessity of an economically intuitive first order condition for optimality of a control policy, based on a feasible perturbation argument. Section 3 proves that the problem under study can sometimes yield "myopic" policy rules, and refers the reader to previous work by Arrow[1968] and Nickell[1974]. Section 4 specializes to functional forms that allow explicit solution of the problem under uncertainty, and derives the solution making use of the necessary conditions discussed in Section 2. Section 5 proposes an alternative approach to the same problem, based on optimal stopping rather than on the feasible perturbation argument, and uses this result to prove existence of the solution. Section 6 discusses the characteristics of the solution and reviews related Operations Research literature on similar control problems.

1 - Statement of the problem

Consider the optimization problem of a firm (or some more general economic entity, for example Robinson Crusoe or the social planner whose optimal plan is mimicked by a competitive market) that tries to maximize a time-separable objective function. The discount factor is constant and equal to r, but instantaneous payoffs depend on exogenous factors W_t (the state of the affairs), whose probability law the firm knows and takes as given, and on endogenous state variables K_t which the firm can manipulate. For simplicity, K_t is assumed to be a scalar. To obtain an interesting problem, we assume that manipulating K_t is costly, and we ask what is the firm's optimal adjustment policy.

In order to state the firm's problem formally, some terms are used whose (rough) definition follows: a σ-field is a partition of the states of nature ω, which are elements of the sample space Ω; statements that are made "almost surely" are true in all states of the world ω, except that they can be false for a set of ωs (an event) which has zero probability measure. If there is no memory loss, observation over time of a stochastic process (a function of time and of the state of nature ω) generates a finer and finer sequence of σ-fields, i.e. provides ever more detailed information about the true state of the world; a probability law assigns probability measure to all sets (events) of the partition; and a random variable X is said to be "adapted" to a σ-field if observation of X's realization does not provide information further to that provided by the σ-field (if the σ-field is generated by a stochastic process, the adapted random variable X is "non-anticipatory" with respect to that stochastic process).

PROBLEM

Given:

- the probability law P for exogenous forcing factors $\{W_\tau; 0\le\tau\le\infty\}$, and the sequence of σ-fields $\mathcal{F}_\tau$ generated by $\{W_s; s\le\tau\}$; W is a possibly vector-valued stochastic process, i.e. a function $W(\omega,\tau):\Omega\otimes\mathbb{R}^+\to\mathbb{R}^n$ that maps every state of nature in ω into a complete sample path for the exogenous variables; we assume that sample paths are right-continuous.

- the functional form of instantaneous payoffs $\pi(K_t,W_t)$;

- a given instantaneous discount rate r;

- the adjustment cost $f\left[dX_\tau, W_\tau\right]$ for each unit displacement of the endogenous state variable; f(.,.) is assumed to have the following form:

$$[1.1]\quad f\left[x,W_\tau\right] = \begin{cases} P(W_\tau) & , \text{ if } x > 0 \\ 0 & , \text{ if } x = 0 \\ p(W_\tau) & , \text{ if } x < 0 \end{cases}$$

with $P(W_\tau)\ge p(W_\tau)$ almost surely for all τ;

- the spontaneous dynamics of the endogenous state variable, given by

[1.2] $dK_t = -\delta K_t dt + X_t$ for all t

where δ is an exponential depreciation rate, assumed constant for simplicity;

Solve for the value function V^* and the associated optimal feedback control process $\{K_\tau;\ K_\tau$ adapted to $\mathcal{F}_\tau,\ t\le\tau\le T\}$, defined by

[1.3] $$V^*(K_t,W_t) \equiv \sup_{\{K_\tau\}} V(K_t,W_t)$$

where $$V(K_t,W_t) \equiv E_t\left\{\int_t^\infty e^{-r(\tau-t)}\left[\pi(K_\tau,W_\tau)d\tau - f\left[dX_\tau,W_\tau\right]dX_\tau\right]\right\}$$

is the value function, and the conditional (on the information available at time t) expectation, $E_t\left\{..\right\}\equiv\int_\Omega ..dP(\omega;\mathcal{F}_t)$, is taken over the joint probability distribution of exogenous variables $\{W_t\}$ and endogenous variables $\{K_t\}$.

Some remarks on the characteristics of the problem considered are in order. Since the problem in described in continuous time, the firm can continuously monitor the state of affairs and act accordingly; as it is often the case, this turns out to simplify the solution. Note however that the degenerate stochastic process (calendar time) $X_\tau=\tau$ a.s. could well be one of the elements of W, and the adjustment cost function could then be specified so as to prevent adjustment at non-integer values of τ, for example, or on Sunday.

Given the assumed form of the adjustment cost function, larger displacements of K incur constant marginal adjustment costs; this produces solutions with "bang-bang" character: if adjustment is undertaken, the speed of adjustment can be infinitely large, so that the paths of the endogenous variable K fail to be differentiable functions of time (K "can jump"). The firm need not continuously adjust the factors that are costly to move, and can instantaneously displace them by a finite amount when it does act. Since X_τ can fail to be differentiable, the second integral in [1.3] is to be interpreted as a Stiltjes integral, with integrating function $dX_\tau \equiv X_{\tau+} - X_{\tau-}$. If the adjustment costs were strictly convex, then the sample paths of $\{X_\tau\}$ would be differentiable, i.e. $dX_\tau = x_\tau d\tau$ where x_τ is the rate of control per unit time.

Note from the form of the first integral above that the instantaneous payoff $\pi(.,.)$ is instead restricted to be a flow: apart from discounting, additions to total revenue are $\pi(K_t, W_t)dt$ in a small time interval dt. In other words, the problem is restricted by the assumption that total undiscounted revenues $\Pi \equiv \int d\Pi$ can be written as $\int \pi(t)dt$, or that the cumulation of revenues is differentiable with respect to time. Nondifferentiability of the total revenue function Π would realistically arise if the firm only sold its product and/or paid variable costs at distinct, and possibly random, times. This is ruled out in [1.3] for simplicity.

Note that, for now, it is assumed that the problem in [1] is well defined, i.e. that the integrals and the expectations exist. This needs to be verified for individual applications: further restrictions will be specified in what follows as necessary.

2 - Characterization of the optimal path

It is possible to characterize the optimal path to some extent (if it exists) without actually solving the problem [2] - in fact, without finding the value function.

A further assumption is necessary:

Assumption The instantaneous payoff function is twice differentiable, increasing and strictly concave in the endogenous state variables:

$$\frac{\partial \pi(K_t, W_t)}{\partial K_t} > 0 \text{ , for all t, almost surely;}$$

$$\frac{\partial^2 \pi(K_t, W_t)}{\partial K_t^2} < 0 \text{ , for all t, almost surely;}$$

Under these conditions, the following is true:

Proposition 1 (Feasible perturbation)

If the firm adopts the optimal control rule, then whenever control is taking place the following is true (almost surely):

$$[2.1] \quad f\left[dX_t, W_t\right] = E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} \frac{\partial \pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau\right\} \quad \text{if } dX_{it} \neq 0$$

And when no control is taking place, it must almost surely be the case that

$$[2.2] \quad p(W_t) \leq E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} \frac{\partial \pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau\right\} \leq P(W_t)$$

In other words, the conditions in [2.1] and [2.2] are __necessary__ for optimality of the firm's policy.

PROOF: By assumption, the firm is following the optimal dynamic program; let $\{K^*_\tau\ ;\ 0\le\tau\le T\}$ be the stochastic process for the endogenous variables corresponding to the optimal feedback rule. We now prove that [2.1,2.2] must almost surely be satisfied by $\{K^*_\tau\ ;\ 0\le\tau\le\infty\}$, or else the value function would not be attaining the maximum. Let A be a subset of Ω, with $p(A)>0$, such that neither [2.1] nor [2.2] are satisfied at some $T<\infty$ if $\omega\in A$; note that $A\in\mathcal{F}_T$, i.e. it is known at time T whether or not the true state of the world is in A ("A occurs"), since all expressions in [2.1] and [2.2] are observable at T. It is then legitimate to perturb the original investment policy $\{dX^*_t\}$ by a small amount Δ at time T if A occurs, __without otherwise modifying the feedback rule__ (so that the amount of control for the endogenous variables is the original one for every time except T and every state of the world except all $\omega\in A$: $d\tilde{X}_T=dX^*_T+\Delta$ if A occurs, and $d\tilde{X}_\tau=dX^*_\tau$ almost surely for $\tau\ne T$. Note that the new policy is legitimate in that it is still adapted to $\mathcal{F}_t$, i.e. depends on an event that is known when the perturbation occurs). By integrating equation [1.2] above, if $\{\tilde{K}_\tau;0\le\tau\le\infty)$ are the endogenous variable paths for the perturbed policy and $\{K^*_\tau;0\le\tau\le\infty\}$ are the paths corresponding to the optimal policy, the perturbed policy under consideration are such that for $0\le\tau\le T$ $\tilde{K}_\tau=K^*_\tau$ almost surely, and for $T\le\tau\le\infty$ $\tilde{K}_\tau=K^*_\tau+\Delta e^{-\delta(\tau-t)}$ if A occurs, $\tilde{K}_\tau=K^*_\tau$ otherwise.

As long as all integrals and expectations are well defined, we can exploit their linearity and additivity properties to write, for any $t\le T$,

$$[2.3]\quad \tilde{V}(K_t,W_t) - V^*(K_t,W_t) =$$

$$= E_t\left\{\int_t^T e^{-r(\tau-t)}\left[\pi(K_\tau^*,W_\tau)-\pi(\tilde{K}_\tau,W_\tau)\right]d\tau\right\} +$$

$$E_t\left\{\int_T^\infty e^{-r(\tau-t)}\left[\pi(K_\tau^* + \Delta e^{-\delta(\tau-T)},W_\tau)-\pi(K_\tau^*,W_\tau)\right]d\tau\right\} -$$

$$E_t\left\{ e^{-r(T-t)}\left[f\left(dX_T^*+\Delta,W_T\right)\ (dX_T^*+\Delta) - f\left(dX_T^*,W_T\right)\ dX_T^*\right]\right\}$$

$$= \int_A e^{-r(T-t)}\int_T^\infty e^{-r(\tau-T)}\left[\pi(K_\tau^* + \Delta e^{-\delta(\tau-T)},W_\tau)-\pi(K_\tau^*,W_\tau)\right]d\tau -$$

$$e^{-r(T-t)}\left[f\left(dX_T^*+\Delta,W_T\right)\Delta+\left[f\left(dX_T^*+\Delta,W_T\right)-f\left(dX_T^*,W_T\right)\right]dX_T^*\right]\ dP(\omega;\mathcal{F}_t)$$

$$> \int_A e^{-r(T-t)}\left[\int_T^\infty e^{-r(\tau-T)}\left[\frac{\partial\pi(K_\tau^*,W_\tau)}{\partial K_\tau^*}\right]\Delta e^{-\delta(\tau-T)}\ d\tau -\right.$$

$$\left. f\left(dX_T^*+\Delta,W_T\right)\ \Delta - \left[f\left(dX_T^*+\Delta,W_T\right)-f\left(dX_T^*,W_T\right)\right]\ dX_T^*\right]\ dP(\omega;\mathcal{F}_t)$$

$$= \int_A e^{-r(T-t)}\left[\int_A\left\{\int_T^\infty e^{-r(\tau-T)}\left[\frac{\partial\pi(K_\tau^*,W_\tau)}{\partial K_\tau^*}\right]\Delta e^{-\delta(\tau-T)}\ d\tau -\right.\right.$$

$$f\left[dX_T^*+\Delta, W_T\right]\Delta - \left[f\left[dX_T^*+\Delta, W_T\right] - f\left[dX_T^*, W_T\right]\right] \quad dX_t^*\Big\} dP(\omega;\mathcal{F}_T)\Big] dP(\omega;\mathcal{F}_t)$$

The inequality above follows from the assumed strict concavity of $\pi(.,.)$ in its first argument, and the last equality uses $\mathcal{F}_t \subseteq \mathcal{F}_T$ for $t \leq T$ (the law of iterated expectations).

It is now possible to show that Δ can be chosen to obtain $\bar{V}(K_t, W_t) > V^*(K_t, W_t)$ if the probability is not zero that neither [2.1] nor [2.2] will be true at some finite time.

Suppose first that $dX_T^* \neq 0$ but, if A occurs, then

$$[2.1'] \quad f\left[dX_T, W_T\right] - E_T\left\{\int_T^\infty e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau\right\} = -\xi \neq 0$$

contradicting [2.1]; note that, since by assumption it is known at T that the true state of the world is in A, $p(\omega;\mathcal{F}_T)=0$ for $\omega \in \{\Omega \backslash A\}$ (the complement of A) and therefore

$$E_T\left\{\int_T^\infty e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau\right\} =$$

$$= \int_\Omega \int_T^\infty e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau \; dp(\omega;\mathcal{F}_T) =$$

$$= \int_A \int_T^\infty e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau \; dp(\omega;\mathcal{F}_T) \quad ;$$

Choose a Δ with the same sign as ξ and smaller in absolute value than dX_T^*, so that (recalling the definition of $f(.,.)$ above) $f\left[dX_T^*+\Delta, W_T\right]=f\left[dX_T^*, W_T\right]$. This choice for Δ yields, after insertion of [2.1'] into [2.3],

$$[2.4] \quad \tilde{V}(K_t,W_t) - V^*(K_t,W_t) > \int_A e^{-r(T-t)}\Delta\ \xi\ dp(\omega;\mathcal{F}_t)$$

But the right hand side of [2.4] is strictly positive if A has positive probability, which contradicts the assumed optimality of the feedback rule which produces $\{K_\tau^*\}$.

Suppose instead that $dX_{it}=0$ but [2.2] fails, for example

$$[2.1'] \quad p(W_t) - E_t\left\{\int_t^T e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K_\tau,W_\tau)}{\partial K_\tau}\, d\tau\right\} = \xi < 0$$

Then set $\Delta<0$ and obtain from [2.3] the inequality

$$\tilde{V}(K_t,W_t) - V^*(K_t,W_t) > \int_A e^{-r(T-t)}\Delta\ \xi\ dp(\omega;\mathcal{F}_t) > 0$$

If it is the second inequality in [2.2] to fail, then choosing a positive Δ will yield a similar inequality.

We conclude that any failure of [2.1,2.2] <u>that occurs with positive probability</u> implies that the firm is not following the optimal feedback rule in its control policy, in that a feasible feedback policy would yield a strictly larger value function.
[end of proof].

This characterization of the optimal policy rule has a straightforward interpretation: whenever the firm is in fact acting, concavity of the payoff function implies that at the margin the action does not alter the value of the policy; the value of the last infinitesimal unit of K_i installed or uninstalled at the present time is simply the expected, present discounted value of the contribution to instantaneous payoffs by the infinitesimal unit that will be marginal at all future times (the shadow price of K). If the firm is not acting, that expected, present discounted value is (weakly) less than the certain, immediate cost of adjustment.

The firm should then, when deciding about the amount of control to be applied at the present time, view the currently marginal unit as the marginal one throughout the planning horizon, and take future investment decisions as given in probability distribution. The very fact that conditions [2.1] and [2.2] will be satisfied at all future times then <u>defines</u> the optimal control policy.

This can be interpreted as an implication of the envelope theorem: the firm is justified in taking future control as given when deciding on the amount of control to be applied today, because any effect of today's control on <u>future</u> value has to occur through a modification of future investment decisions; but these are assumed to be optimal, hence at the margin a small change in future control has no effect on the value function.

It should be noted that this characterization can be obtained imposing very little structure on the problem. Apart from the assumptions of linearity of the adjustment cost function and of concavity of the payoff function, it is only assumed that policies are non-anticipative with respect to the exogenous variables, that all the integrals and expectations are well-defined and therefore have the usual linearity and additivity properties. Once again, existence of the value function needs to be verified for specific applications.

3 - Myopic policy rules

Recall that it has been assumed above that the exogenous variables $\{W_t\}$ have right-continuous sample paths. An additional assumption will guarantee that p(.), P(.), $\pi(K,.)$ and $\partial\pi(K,.)/\partial K$ have right-continuous sample paths as well:

Assumption: p(W), P(W) and $\pi(K,W)$ are continuously differentiable in W.

This, together with the characterization obtained above for the optimal control rule, suffice to obtain an important result:

Proposition 2 (Euler equation) a) If an optimal feedback rule exists, and $dX_\tau \neq 0$ for $\tau \in [t_1,t_2)$, $t_1<t_2$ (i.e. control of the endogenous variable takes place with probability one between t_1 and t_2), then

$$[3.1] \quad \frac{\partial\pi(K_\tau,W_\tau)}{\partial K_\tau}\, d\tau = (r+\delta)\ f(dX_\tau,W_\tau)\ d\tau - E_\tau df(dX_\tau,W_\tau) \quad \text{for } \tau\in[t_1,t_2)$$

PROOF: Since control occurs continuously in $[t_1,t_2)$, we have from Proposition 1 that

$$[3.2] \qquad f\left[dX_t, W_t\right] = E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K_\tau, W_\tau)}{\partial K_\tau}\, d\tau\right\}$$

almost surely for $t\in[t_1, t_2)$

For a fixed ω, consider the path-by-path differential of the two sides of equation [3.2] with respect to t (the differential is well defined since W - and therefore K - have right-continuous sample paths) to obtain:

$$[3.3] \quad df\left[dX(\omega,t), W(\omega,t)\right] = \frac{\partial\pi(K(\omega,t), W(\omega,t))}{\partial K(\omega,\tau)}\, dt$$

$$+\left[\int_t^{\infty} (r+\delta)\; e^{-(r+\delta)(\tau-t)} \frac{\partial\pi(K(\omega,\tau), W(\omega,\tau))}{\partial K(\omega,\tau)}\, d\tau\right] dt$$

which is true for all $t\in[t_1, t_2)$ and all $\omega\in\{\Omega\backslash A\}$, where A is any subset of Ω such that $P(A)=0$.

Now integrate both sides of [3.3] over Ω with respect to $P(\omega;\mathcal{F}_t)$, noting that any ω such for which [3.3] is not satisfied belongs to sets that receive zero weight in the integration, and that K_t and W_t are known at time t so that integration over states of nature of the first term on the right-hand side of [3.3] returns its actual value:

$$[3.4] \qquad E_t\left\{df\left[dX_t, W_t\right]\right\} = \frac{\partial\pi(K_t, W_t)}{\partial K_t}\, dt \; +$$

$$+ (r+\delta)\ E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)}\ \frac{\partial\pi(K_\tau,W_\tau)}{\partial K_\tau}\ d\tau\right\}$$

from which the assertion follows noting from equation [3.2] that the last term is equal to $(r+\delta)\ f(dX_t,W_t)$.

[end of proof]

Proposition 2 states the conditions necessary for the firm to temporarily base its control policy only on current events, without looking forward and with no need to take the endogenous variable's process into account. Note that the proof does not go through as soon as there is any probability that it will be optimal to abstain from control in the next instant. In particular, of course, there is no presumption that a condition like [3.1] should hold when the firm is abstaining from control, i.e. when dX=0.

The economic interpretation of [3.1] is straightforward: if control is certainly occurring throughout $[t_1,t_2)$, then it must be distributed along that interval in such a way that [3.1] is true. Otherwise, a reallocation of control would increase the value of the firm; in other words, a version of the Euler equation would be violated.

Arrow[1967] and Nickell[1974] used the equivalent of [3.1] to characterize investment under linear adjustment costs, assuming that the exogenous variables in W follow <u>piecewise continuous</u> paths, about which the firm has <u>no uncertainty</u>: then when investment occurs, it occurs continuously over an interval, and consideration of [3.1] is sufficient to solve for the path of the endogenous variable K. Control (investment) will stop and resume at points in time where

$$[3.1'] \quad \frac{\partial \pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau = (r + \delta - \frac{dP_\tau}{P_\tau}) P_\tau$$

and knowledge of the exogenous variables' path will suffice to construct the optimal investment policy: Arrow and Nickell provide algorithms for this purpose, and show that in general investment will stop <u>before</u> a cyclical peak is attained and resume <u>after</u> the cyclical trough.

If control is certain to occur at <u>all</u> times, i.e. the equations in [2.1] always hold with equality, then [3.1] is always true and the firm can follow a "myopic" policy rule, varying the endogenous factor K so as to equate its <u>current</u> marginal payoff to the <u>current</u> value of the right-hand side of [3.1]. A sufficient condition for this to be the case is that $P(W_\tau)=p(W_\tau)$ for all τ, almost surely: then [2.1] and [2.2] collapse to the single equation

$$[3.5] \quad E_t\left\{\int_t^\infty e^{-(r+\delta)(\tau-t)} \frac{\partial \pi(K_\tau, W_\tau)}{\partial K_\tau} d\tau\right\} = P(W_t) \quad , \text{ all } t$$

In such a situation, of course, the firm does not need to solve a truly dynamic problem: use of K_t can be varied continuously to satisfy [3.1], whose right-hand side is the so called "user cost of capital" (see Jorgenson[1963]) if K is the installed capital stock.

4 - A class of solvable problems

Consider the following specialization of the general problem described in [1] and [2] above:

$\{W_t\}=\{W_{zt},W_{pt}\}$ is taken to be a two-dimensional Brownian motion process, i.e. each state of the world ω is associated with two continuous sample paths with increments $\{W_{z\tau 2}-W_{z\tau 1},W_{p\tau 2}-W_{p\tau 1};$ $\tau 2>\tau 1\}$ which are independent of $\{W_{z\tau 1},W_{p\tau 2}\}$ and have a bivariate normal distribution given the information available at $\tau 1$;

[4.1] $\pi(K_t,Z_t) = \dfrac{1}{1+\beta} K_t^{1+\beta} Z_t$, $-1<\beta<0$, a constant elasticity function of the endogenous and exogenous state variables, with

[4.2] $dZ_t= \vartheta_z Z_t dt + Z_t \sigma_z dW_{zt}$, a univariate geometric Brownian motion process;

P_t follows a geometric Brownian motion with stochastic differential

[4.4] $dP_t=\vartheta_p P_t dt + P_t \sigma_p dW_{pt}$ where dW_{pt} is the increment of another standard Wiener process with correlation ρ to dW_{zt};

[4.5] $p_t = \lambda P_t$, λ constant, $\lambda\le 1$ to satisfy the assumption that $p_t\le P_t$ always.

The problem data are β, ϑ_z, ϑ_p, σ_z, σ_p, ρ, λ, δ and r: all these are assumed to be constant over time.

Note that geometric Brownian motion processes are particularly convenient since their Markov state space is completely described by their level alone; and assuming a constant-elasticity functional form for the instantaneous payoff facilitates solution because constant-elasticity functions of geometric Brownian motion follow geometric Brownian motion.

The restriction $-1<\beta<0$ makes Proposition 1 in the previous section applicable; we then seek a non-anticipative control rule that satisfies [2.1] and [2.2] at all times: i.e., the optimal control rule must specify a nonanticipating control process $\{X_t\}$ such that

$$[4.6]\quad E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} K_\tau^{\beta} Z_\tau \, d\tau\right\} = P_t \quad \text{if } dX_t>0$$

$$[4.7]\quad p_t < E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} K_\tau^{\beta} Z_\tau \, d\tau\right\} < P_t \quad \text{if } dX_t=0$$

$$[4.8]\quad E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} K_\tau^{\beta} Z_\tau \, d\tau\right\} = p_t \quad \text{if } dX_t<0$$

With the assumptions made above, it is reasonable to guess that the optimal control process will have the following form: the endogenous variable should be displaced only as necessary to obtain

$$[4.9]\quad \ell p_t \le K_t^{\beta} Z_t \le u P_t \quad \text{or equivalently} \quad \lambda\ell \le \frac{K_t^{\beta} Z_t}{P_t} \le u$$

where the boundaries u and ℓ are constants to be determined. Such a control rule is obviously nonanticipating, since it only depends on current values of the exogenous variables, which in turn have continuous sample paths.

To verify that the control rule has the form [4.9], and to find the optimal control barriers ℓ and u, it is now necessary to compute the expectations of discounted marginal payoff streams appearing in [4.6]-[4.8].

The problem at hand is to find an expression for the conditional expectation appearing in [4.6]-[4.8]; the conditional expectation will, of course, be a function of the current value of the state variables and of the parameters δ, r, u, ℓ, ϑ, ϑ_p, σ_z, σ_p, β. Define a new variable $\eta_t \equiv K_t^{\beta} Z_t$, and define a functional expression for the conditional expectation:

$$[4.10] \quad f\left[\eta_t, P_t;\ \delta,\ r,\ u,\ \ell \ldots\right] \equiv E_t\left\{\int_t^{\infty} \eta_\tau \, e^{-(r+\delta)(\tau-t)} d\tau\right\}$$

Of course, the control rule in [4.9] has to be taken into account when computing the conditional expectation in [4.10]: the process followed by $\{\eta_t\}$ under the control rule needs to be determined.

Now note that when control is not enacted, i.e. dX_t is zero, then $dK_t = -\delta\ dt$, and the derivative of the payoff function with respect to K follows a geometric Brownian motion, being a constant elasticity function of geometric processes; define $\xi_t \equiv \{K_t^{\beta} Z_t;\ dX_t = 0\}$, and use Ito's lemma to find its stochastic differential:

$$[4.11] \quad d\xi_t = \frac{\partial \xi_t}{\partial K_t} dK_t + \frac{\partial \xi_t}{\partial Z_t} dZ_t + \frac{1}{2}\left[\frac{\partial^2 \xi_t}{\partial K_t^2}\left(dK_t\right)^2 + \frac{\partial^2 \xi_t}{\partial Z_t^2}\left(dZ_t\right)^2 + 2\frac{\partial^2 \xi_t}{\partial K_t \partial Z_t}\left(dK_t dZ_t\right)\right] =$$

$$= \beta \quad K_t^{\beta-1} Z_t \left[-\delta \; K_t dt\right] + K_t^{\beta}\left[\vartheta_z \; Z_t dt + \; Z_t \; \sigma_Z \; dW_{zt}\right] +$$

$$+ \frac{1}{2}\left[\beta(\beta-1) K_t^{\beta-2} \; 0 \quad + 0 \; \sigma^2 Z^2 \; dt + 2 \; \beta \; K_t^{\beta-1} \; 0\right] =$$

$$= \xi_t \left[-\delta\beta + \vartheta_z\right] dt + \xi_t \left[\sigma_Z dW_{zt}\right]$$

$$\equiv \xi_t \; \mu \; dt + \sigma_z dW_{zt}$$

where $\mu \equiv -\delta\beta + \vartheta_z$.

If a control policy of the form [4.9] is adopted, the firm will prevent $\{\xi_t\}$ from ever being larger than $u \; P_t$ or smaller than $\ell \; p_t = \ell\lambda P_t$: the derivative of the payoff function with respect to K then follows a regulated geometric Brownian motion, with moving control barriers at $u \; P_t$ and $\ell \; \lambda \; P_t$: i.e. the stochastic process $\{\eta_t\}$ is defined by

[4.12] $$\eta_t = \frac{\xi_t \; L_t}{U_t}$$

where:

(i) $\{\xi_t\}$ is a geometric Brownian motion process, with stochastic differential

$$d\xi_t = \xi_t \; \mu \; dt + \xi_t \; \sigma_z dW_{zt}$$

and initial condition ξ_0;

(ii) $\{U_t\}$ and $\{L_t\}$ are increasing and continuous processes, with $L_0 = U_0 = 1$;

(iii) $\{L_t\}$ only increases when $\eta_t = \ell\lambda P_t$, and $\{U_t\}$ only increases when $\eta_t = uP_t$, where u and ℓ are given positive real numbers and P_t follows a geometric Brownian motion, with stochastic differential

$dP_t = \vartheta_p P_t\ dt + P_t\ \sigma_p\ dW_{pt}$, $dW_{zt} dW_{pt} = \rho\sigma_z\sigma_p$

and initial condition P_0, such that $\ell\lambda P_0 \le \xi_0 \le uP_0$;

(iv) $\ell \le \eta_t \le u$ for all $t \ge 0$

These four properties uniquely identify $\{U_t\}$ and $\{L_t\}$; these two processes maintain η_t within the moving barriers using the minimum amount of control, since they only increase when η_t it at the frontiers of the region $[\ell, u]$. Proposition (6), page 22 in Harrison[1985] proves uniqueness formally for the case of a regulated linear Brownian motion process, and it is easy to adapt the proof to the present case of a regulated geometric Brownian motion process: note that $\{\eta_t/P_t\}$ is a geometric Brownian motion process regulated between $\lambda\ell$ and u, implying that $\{\ln(\eta_t/P_t)\}$ is a linear Brownian motion process regulated between $\ln(\lambda\ell)$ and $\ln(u)$, and apply Harrison's proof of uniqueness.

It is now possible to compute the expectations of discounted marginal profitability streams appearing in equations [4.6]-[4.8], as a function of the yet to be determined control points ℓ and u, and of the problem data.

First note that $\{U_t\}$ and $\{L_t\}$ are processes of finite variation, since they never decrease: this means that $(dU_t)^2 = (dL_t)^2 = (dU_t dL_t) = (dU_t d\xi_t) = (dL_t d\xi_t) = 0$.

These relations imply that if we apply Ito's lemma to η_t, which is a continuously differentiable function of ξ_t, U_t and L_t, all the second order terms vanish to yield:

$$[4.13]\quad d\eta_t = d\left[\frac{\xi_t L_t}{U_t}\right] = \frac{L_t}{U_t} d\xi_t + \frac{\zeta_t}{U_t} dL_t - \frac{\zeta_t L_t}{U_t^2} dU_t =$$

$$\frac{L_t}{U_t} \xi_t \mu \, dt + \frac{L_t}{U_t} \sigma_z dW_{zt} + \frac{\zeta_t}{U_t} L_t \frac{dL_t}{L_t} - \frac{\zeta_t L_t}{U_t} \frac{dU_t}{U_t}$$

$$= \eta_t \mu \, dt + \eta_t \sigma_z dW_{zt} + \eta_t \frac{dL_t}{L_t} - \eta_t \frac{dU_t}{U_t}$$

Now consider the conditional expectation defined in [4.10], $f(\eta_t, P_t)$ (the dependence of the function on the time-invariant parameters r, δ, u, ℓ... is suppressed in what follows for typographical convenience); assume that it is a continuously differentiable function of η_t and P_t, and apply Ito's lemma again to obtain (subscripts denote partial derivatives):

$$[4.14]\quad df(\eta_t, P_t) = f_1(\eta_t, P_t) d\eta_t + f_2(\eta_t, P_t) dP_t +$$

$$\frac{1}{2} f_{11}(\eta_t, P_t)(d\eta_t)^2 + \frac{1}{2} f_{22}(\eta_t, P_t)(dP_t)^2 +$$

$$f_{12}(\eta_t, P_t)(d\eta_t dP_t) =$$

$$= f_1(\eta_t, P_t)\left[\eta_t \mu \, dt + \eta_t \sigma_z dW_{zt}\right] - f_1(uP_t, P_t) \; uP_t \frac{dU_t}{U_t}$$

$$+ f_1(\lambda\ell P_t, P_t) \; \lambda\ell P_t \frac{dL_t}{L_t} + f_2(\eta_t, P_t)\left[\vartheta_p P_t \, dt + P_t \sigma_p \, dW_{pt}\right]$$

$$+ \frac{1}{2} f_{11}(\eta_t, P_t)\sigma_z^2 \eta_t^2 \, dt + \frac{1}{2} f_{22}(\eta_t, P_t)\sigma_p^2 P_t^2 \, dt +$$

$$+ f_{12}(\eta_t, P_t)\eta_t P_t \, \rho\sigma_z\sigma_p \, dt$$

where the fact that $dL \neq 0$ only if $\eta_t = \ell\lambda P_t$, and similarly $dU_t \neq 0$

only if $\eta_t = uP_t$, is used in obtaining the last equality.

Now recall the Integration By Parts formula found in Harrison[1985], page 73: if $\{Y_t\}$ is an Ito process (i.e. the stochastic integral $\int dY_t$ is well defined) and $\{X_t\}$ is a continuous process with finite variation, then

$$Y_\nu X_\nu = Y_t X_t + \int_0^\nu Y_\tau dX_\tau + \int_0^\nu X_\tau dY_\tau$$

Apply the Integration by Parts formula to $Y_\nu = f(\eta_\nu, P_\nu)$ and $X_\nu = e^{-(r+\delta)(\nu-t)}$, which always decreases and therefore has finite variation; using $d[e^{-(r+\delta)(\tau-t)}] = -(r+\delta)e^{(r+\delta)(\tau-t)}d\tau$, and $df(\eta_\tau, P_\tau)$ from [4.14], one obtains, after rearranging terms,

$$[4.15]\quad e^{-(r+\delta)\nu} f(\eta_\nu, P_\nu) = f(\eta_t, P_t)$$

$$+\int_t^\nu e^{-(r+\delta)(\tau-t)} \Big[f_1(\eta_\tau, P_\tau)\eta_\tau \mu + f_2(\eta_\tau, P_\tau)\vartheta_p P_\tau + \frac{1}{2} f_{11}(\eta_\tau, P_\tau)\sigma_z^2 \eta_\tau^2 +$$

$$\frac{1}{2} f_{22}(\eta_\tau, P_\tau)\sigma_p^2 P_\tau^2 + f_{12}(\eta_\tau, P_\tau)\eta_\tau P_\tau \rho \sigma_z \sigma_p - (r+\delta) f(\eta_\tau, P_\tau) \Big] d\tau +$$

$$\int_t^\nu f_1(\eta_\tau, P_\tau)\, \eta_\tau \sigma_z dW_{z\tau} - \int_t^\nu f_1(uP_\tau, P_\tau)\, uP_\tau \frac{dU_\tau}{U_\tau} +$$

$$\int_t^\nu f_1(\lambda \ell P_\tau, P_\tau)\, \lambda \ell P_\tau \frac{dL_\tau}{L_\tau} + \int_\tau^\nu f_2(\eta_\tau, P_\tau) P_\tau \sigma_p dW_{p\tau}$$

For the investment rule to be well defined it must be the case that f(.,.) and its derivatives are always finite; this implies that

$$\lim_{\nu \to \infty} \left[e^{-(r+\delta)\nu} f(\eta_\nu, P_\nu) \right] = 0$$

and, for all $\nu \geq t$,

$$E_t\left\{ \int_t^\nu f_1(\eta_\tau, P_\tau) \ \eta_\tau \sigma_z dW_{z\tau} \right\} = E_t\left\{ \int_t^\nu f_2(\eta_\tau, P_\tau) P_\tau \sigma_p dW_{p\tau} \right\} = 0$$

because the expectation of the integral of a bounded function against a Wiener process is always zero (see, for example, Proposition 4.3.7 in Harrison[1985]: this is due to the erratic behavior of the Wiener process, whose increments after time t are completely unpredictable, by definition, on the basis of the information available at t).

Then, let $\nu \to \infty$ and take the conditional expectation of [4.15] at t to obtain

$$[4.16]\quad 0 = f(\eta_t, P_t) + E_t\left\{ \int_t^\infty e^{-(r+\delta)(\tau-t)} \left[f_1(\eta_\tau, P_\tau)\eta_\tau \mu + f_2(\eta_\tau, P_\tau)\vartheta_p P_\tau + \frac{1}{2} f_{11}(\eta_\tau, P_\tau)\sigma_z^2 \eta_\tau^2 + \frac{1}{2} f_{22}(\eta_\tau, P_\tau)\sigma_p^2 P_\tau^2 + f_{12}(\eta_\tau, P_\tau)\eta_\tau P_\tau \rho \sigma_z \sigma_p - (r+\delta) \ f(\eta_\tau, P_\tau) \right] d\tau \right\} + - E_t\left\{ \int_t^\infty f_1(uP_\tau, P_\tau) \ uP_\tau \frac{dU_\tau}{U_\tau} - \int_t^\infty f_1(\lambda \ell P_\tau, P_\tau) \ \lambda \ell P_\tau \frac{dL_\tau}{L_\tau} \right\}$$

Recall, from [4.10], that

$$f\left[\eta_t, P_t\right] \equiv E_t\left\{\int_t^{\infty} \eta_\tau \; e^{-(r+\delta)(\tau-t)} d\tau\right\};$$

In light of [4.16], this can be true only if on the one hand

$$[4.17] \quad f_1(\eta,P)\eta\mu + f_2(\eta,P)\vartheta_p P + \frac{1}{2} f_{11}(\eta,P)\sigma_z^2\eta^2 + \frac{1}{2} f_{22}(\eta,P)\sigma_P^2 P^2$$

$$+ f_{12}(\eta_t, P_t)\eta_t P_t \; \rho\sigma_z\sigma_P - (r+\delta)\; f(\eta,P) = -\eta$$

for all η and P such that $\lambda\ell \leq (\eta/P) \leq u$, and, on the other hand, the last expectation in [4.16] vanishes: this requires that

$$[4.18] \quad f_1(uP,P)uP = f_1(\lambda\ell P,P)\; \lambda\ell P = 0 \quad \text{for all } P$$

We conclude that, in general, the conditional expectation $f(\eta_t, P_t)$ is defined by the functional equation [4.17] with boundary conditions [4.18].

The general form of the solution to [4.17] is a linear combination of power functions and of a linear term,

$$[4.20] \quad f(\eta,P) = B_1\; \eta + B_2\; \eta^{\alpha_1}P^{\beta_1} + B_3\; \eta^{\alpha_2}P^{\beta_2}$$

Imposing the boundary conditions [4.18] on the functional form in [4.20] we find:

$$[4.21] \quad B_1 uP + B_2\alpha_1\; (uP)^{\alpha_1}P^{\beta_1} + B_3\alpha_2\; (uP)^{\alpha_2}P^{\beta_2} = 0$$

$$B_1\lambda\ell P + B_2\alpha_1\; (\lambda\ell P)^{\alpha_1}P^{\beta_1} + B_3\alpha_2\; (\lambda\ell P)^{\alpha_2}P^{\beta_2} = 0$$

These conditions must be satisfied for all $P \geq 0$, which requires that

[4.22] $\beta_1=1-\alpha_1, \quad \beta_2=1-\alpha_2$

[4.23] $B_1 + B_2\alpha_1 \, u^{\alpha_1-1} + B_3\alpha_2 \, u^{\alpha_2-1} = 0,$

$B_1 + B_2\alpha_1 \, (\lambda\ell)^{\alpha_1-1} + B_3\alpha_2 \, (\lambda\ell)^{\alpha_1-1} = 0$

In light of [4.22], we can rewrite [4.20] as

[4.24] $f(\eta,P) = B_1 \, \eta + B_2 \, \eta^{\alpha_1}P^{1-\alpha_1} + B_3 \, \eta^{\alpha_2}P^{1-\alpha_2}$

and proceed to compute the partial derivatives of this function to find that, for [4.17] to be satisfied,

$$B_1=\frac{1}{r+\delta-\mu}$$

and α_1, α_2 must be solutions of

[4.25] $\frac{\sigma^2}{2} \alpha^2 + \left(\mu-\vartheta_p - \frac{\sigma^2}{2} \right) \alpha - \left(r+\delta-\vartheta_p\right) = 0$

where $\sigma^2 \equiv \sigma_z^2 + \sigma_p^2 - 2\rho\sigma_z\sigma_p > 0$

The quadratic equation [4.25] has two real roots of opposite sign provided that $r+\delta-\vartheta_p>0$; let α_1 be the positive root and α_2 be the negative root:

$$\alpha_1 \equiv \frac{-\left(\mu-\vartheta_p - \frac{\sigma^2}{2} \right) + \sqrt{\left(\mu-\vartheta_p - \frac{\sigma^2}{2} \right)^2 + 2\left(r+\delta-\vartheta_p\right) \sigma^2}}{\sigma^2} > 0$$

$$\alpha_2 \equiv \frac{-\left(\mu-\vartheta_p - \frac{\sigma^2}{2} \right) - \sqrt{\left(\mu-\vartheta_p - \frac{\sigma^2}{2} \right)^2 + 2\left(r+\delta-\vartheta_p\right) \sigma^2}}{\sigma^2} < 0$$

Note, for use below, that $r > -\vartheta_z\frac{1}{\beta} + \vartheta_p(1 + \frac{1}{\beta}) + \frac{\sigma^2}{2} \frac{1+\beta}{\beta^2}$ implies that $-\beta\alpha_1>1$, as can be verified noting that $-\beta\alpha_1$ is the positive solution to the quadratic equation

$$[4.26] \quad \frac{\sigma^2}{2\beta^2} X^2 - \left[\vartheta_z \frac{1}{\beta} + \delta - \vartheta_p \frac{1}{\beta} - \frac{\sigma^2}{2}\frac{1}{\beta}\right] X - \left[r+\delta-\vartheta_p\right] = 0$$

The only parameters as yet unsolved for in [4.24] are B_2 and B_3: but insertion of $B_1 = (r+\delta-\mu)^{-1}$ in [4.23] yields a system of two linear equations in B_2 and B_3, with solution

$$[4.27] \quad B_2 = \frac{1}{r+\delta-\mu} \frac{u^{\alpha_2}\lambda\ell - u\,(\lambda\ell)^{\alpha_2}}{\alpha_1\left[u^{\alpha_1}(\lambda\ell)^{\alpha_2} - u^{\alpha_2}(\lambda\ell)^{\alpha_1}\right]}$$

$$B_3 = \frac{1}{r+\delta-\mu} \frac{u(\lambda\ell)^{\alpha_1} - u^{\alpha_1}\lambda\ell}{\alpha_2\left[u^{\alpha_1}(\lambda\ell)^{\alpha_2} - u^{\alpha_2}(\lambda\ell)^{\alpha_1}\right]}$$

This completes the derivation of

$$f(K_t^\beta Z_t, P_t;\ u,\ \ell\ \ldots) \equiv E_t\left\{\int_t^\infty e^{-(r+\delta)(\tau-t)} K_\tau^\beta\ Z_\tau\ d\tau\right\}$$

under the assumption that the firm adopts a control rule that obtains

$$[4.28] \quad \lambda\ell \le \frac{K_t^\beta\ Z_t}{P_t} \le u$$

using the <u>minimum</u> amount of control: control dX_t is applied to K_t <u>only</u> when one of the weak inequalities in [4.28] holds with equality.

To find which constants u and ℓ characterize the optimal control policy, we insert $f(K_t^\beta Z_t, P_t;\ u,\ \ell\ \ldots)$ into the necessary conditions [4.6]-[4.8]:

$$[4.6']\quad f\left[\lambda\ell P_t, P_t\ ;\ u,\ \ell\ \ldots\right] = \lambda P_t \qquad (dX_t<0 \Rightarrow K_t^\beta Z_t=\ell\lambda P_t)$$

$$[4.8']\quad f\left[\ uP_t, P_t\ ;\ u,\ \ell\ \ldots\right] = P_t \qquad (dX_t>0 \Rightarrow K_t^\beta Z_t=uP_t)$$

Using [4.24], these conditions read

$$[4.6"]\quad B_1\ \lambda\ell P_t + B_2\ (\lambda\ell P_t)^{\alpha_1}P_t^{1-\alpha_1} + B_3\ (\lambda\ell P_t)^{\alpha_2}P_t^{1-\alpha_2} = \lambda P_t$$

$$[4.8"]\quad B_1\ uP_t + B_2\ (uP_t)^{\alpha_1}P_t^{1-\alpha_1} + B_3\ (uP_t)^{\alpha_2}P_t^{1-\alpha_2} = P_t$$

P_t can be eliminated from these equations; then, insertion of the expressions found above for B_1, B_2, B_3, and some simplification, yields:

$$[4.6"]\quad \lambda\left[\ell-(r+\delta-\mu)\right]\left[u^{\alpha_1}(\lambda\ell)^{\alpha_2} - u^{\alpha_2}(\lambda\ell)^{\alpha_1}\right]$$

$$+\left[u^{\alpha_2}\lambda\ell - u\ (\lambda\ell)^{\alpha_2}\right]\frac{(\lambda\ell)^{\alpha_1}}{\alpha_1} + \left[u(\lambda\ell)^{\alpha_1} - u^{\alpha_1}\lambda\ell\right]\frac{(\lambda\ell)^{\alpha_2}}{\alpha_2} = 0$$

$$[4.8"]\quad \left[u-(r+\delta-\mu)\right]\left[u^{\alpha_1}(\lambda\ell)^{\alpha_2} - u^{\alpha_2}(\lambda\ell)^{\alpha_1}\right]$$

$$+\left[u^{\alpha_2}\lambda\ell - u\ (\lambda\ell)^{\alpha_2}\right]\frac{(u\ )^{\alpha_1}}{\alpha_1} + \left[u(\lambda\ell)^{\alpha_1} - u^{\alpha_1}\lambda\ell\right]\frac{(u\ )^{\alpha_2}}{\alpha_2} = 0$$

These two equations are highly nonlinear in u and ℓ, but can be easily solved numerically. Although it does not seem possible to prove analytically the uniqueness of their solution, numerical procedures are not at all sensitive to the starting point of the iteration, suggesting that the solution is indeed unique.

Closed-form solutions can be found for special cases: if $\lambda=1$ then the two equations have the same form, and cannot be solved for distinct u and ℓ. But with $\lambda=1$ [4.6] and [4.8] are infact the same equation, implying that $u=\ell$; then $[\partial\pi(K_t,Z_t)/\partial K_t]/P_t$ is <u>constant</u>, and simple integration on [4.6] or [4.8] shows that the solution to the optimal control problem simply equates the current marginal contribution of K to the payoffs to the current "user cost" of K, $(r-\vartheta_p+\delta)P_t$, confirming that when $P_t=p_t$ for all t the firm will continuously exercise control (see the discussion after Proposition 2 above).

If $\lambda=0$ then it is clear that dX_t is never negative ([4.8] never applies) since $[\partial\pi(K_t,Z_t)/\partial K_t]>0$ always for $K>0$: this is the case of <u>irreversible</u> accumulation of K. The firm only exercises positive control, and K only decreases via depreciation.

Given that $\ell=0$, it is possible to find a closed-form solution for u. The requirement that f(.,.) be bounded imposes that $B_3=0$ in [4.20] above, since in the absence of negative control η_t can become arbitrarily close to zero; the boundary condition then provides a closed-form solution for u. The derivation of the solution proceeds along the lines discussed above, and is not reported to save space; the form of u for the case of irreversible control is reported in the next Chapter (see [R]).

The next section uses an alternative approach to the solution of the control problem, and finds a sufficient condition for existence of a solution (finiteness of the value function).

5 - An alternative approach to the solution: Marginal Option Pricing

The dynamic program defined in [4.1] above can be decomposed in a sequence of optimal stopping problems: rather than deciding how much control should be exercised at any given time, the firm should decide when each infinitesimal particle of K should be installed. This approach to the problem (which was first proposed by Pindyck[1986]) uses more familiar mathematics than the approach used in the previous section, but is based on somewhat less economically intuitive thought experiments.

When deciding about the optimal timing for installation of the K^{th} marginal unit of capital, the optimal stopping problem facing the firm is in the form considered by McDonald and Siegel[1986] for the case of $\lambda=0$ (no resale possibility) and extended by Dixit[1987] to the case in which the decision can be reverted.

The general form of the optimal stopping problem considered by these authors is as follows: an asset with value v can be purchased at price P, if not purchased yet, or sold at price p, if already purchased. McDonald and Siegel assume that both the value of the asset v and the purchase price P are uncertain, and follow geometric Brownian motion; but they impose that no resale be possible, so that the exchange of the option for the asset is an irreversible decision. They note that the value of the asset could simply be equal to the present value, appropriately discounted, of the dividends it will provide to its owner (but not, of course, to the holder of the option); the asset must produce dividends, or it would never be profitable to buy it since it cannot be resold. Dixit, on the other hand, considers the case of <u>reversible</u> decisions but restricts P and p to be both fixed (nonstochastic).

In this section the optimal stopping technique is applied to the decision to install or uninstall the <u>currently marginal</u> unit of K: it will be shown that the firm's problem can be reduced to a <u>sequence</u> of stopping problems of the form outlined above.

Consider that at any time, given that the total installed stock of the endogenous variable is K, installation or uninstallation of an additional (K^{th}) infinitesimal unit of capital is possible; installation will produce a stream of marginal payoffs whose expectation is easy to compute, because it only depends on the <u>exogenous</u> probability law of Z.

Define the "dividend" provided by installed marginal units of K (its marginal contribution to the payoff) as

$$[5.1] \quad \eta_t = \eta(K, Z_t) \equiv K^{\beta} Z_t$$

for each K, $K \in [0, \infty)$.

The next task is computing the "value" of the K^{th} installed unit, i.e. the expectation of the stream of marginal payoffs it will produce. The presence of depreciation introduces some complications: note that installation of the K^{th} unit at time t increases the capital stock at any future time τ by only $e^{-\delta(\tau-t)} < 1$ units, since the unit steadily depreciates at rate δ; then note that the total capital stock also depreciates at rate δ, and therefore the depreciation rate appears twice in the expression for the additional payoff produced by installation of the marginal unit of the endogenous factor when the current stock is K:

$$[5.2] \quad v\left[K, Z_t\right] = E_t\left\{\int_t^{\infty} e^{-r(\tau-t)} e^{-\delta(\tau-t)} \left[Ke^{-\delta(\tau-t)}\right]^{\beta} Z_{\tau}\, d\tau\right\} =$$

$$= \int_t^{\infty} e^{-(r+\delta)(\tau-t)} \left[Ke^{-\delta(\tau-t)}\right]^{\beta} E_t\left\{Z_{\tau}\right\} d\tau =$$

$$= \int_t^{\infty} e^{-(r+\delta)(\tau-t)} \left[Ke^{-\delta(\tau-t)}\right]^{\beta} Z_t e^{\vartheta_z(\tau-t)}\, d\tau =$$

$$= \frac{K^{\beta} Z_t}{r+\delta-(\vartheta_z-\delta\beta)} = \frac{\eta_t}{r+\delta-(\vartheta_z-\delta\beta)}$$

The firm holds the right to pay P_t and acquire, at any time, the currently marginal unit of K, and forever receive the stream of marginal payoffs whose discounted expectation is given in [5.2], regardless of future control of the endogenous stock; on the other hand, the firm can always sell the marginal unit of capital, receiving λP_t but giving up the stream of marginal payoffs those units would have produced if left in place.

If the firm does not purchase the K^{th} unit, a "dividend" is given up: uninstalled units do not produce, and payoffs are lower.

But if the right to install the K^{th} unit is exercised, the firm gives up the option to wait and learn information about the evolution of $\{Z_t\}$ and $\{P_t\}$: it is often the case, of course, that installation turns out to be a bad idea ex-post, because Z_t or P_t fall; conversely, uninstallation can turn out to have been a bad idea if Z_t or λP_t rise. This defines an optimal stopping problem for the decision to install the currently marginal unit: the firm has to trade off the foregone dividends and the risk of acting too soon.

The value of the marginal unit available for installation at time t has, given that the already installed capital stock K_t is depreciating at rate $-\delta$, dynamics given by Ito's lemma as

$$dv\left(K_t, Z_t\right) = \frac{d(K_t^\beta\ Z_t)}{r+\delta-(\vartheta_z-\delta\beta)} =$$

$$= \frac{1}{r+\delta-(\vartheta_z-\delta\beta)}\left[\ \beta K_t^{\beta-1} Z_t\ \left(-\delta\ K_t dt\right) + K_t^\beta\left(\vartheta_z\ Z_t dt + \ Z_t\ \sigma_z\ dW_{zt}\right)\right] =$$

$$= v\left(K_t, Z_t\right)\ \left(-\delta\beta + \vartheta_z\right)dt + v\left(K_t, Z_t\right)\ \sigma_z dW_{zt}$$

$$\equiv v\left(K_t, Z_t\right)\ \mu\ dt + v\left(K_t, Z_t\right)\ \sigma_z dW_{zt}$$

By the assumptions at the beginning of section 5, the purchase price of K follows the dynamics given in [4.4] above:

$$dP_t = P_t \vartheta_p dt + P_t \sigma_p dW_{pt}$$

and the resale price of K follows:

$$d\lambda P_t = \lambda P_t \vartheta_p dt + \lambda P_t \sigma_p dW_{pt}$$

Therefore, the value of the marginal unit in place, its purchase cost and its resale price follow geometric Brownian motion processes; the correlation between the increments of the first process and the increments of the latter two processes is ρ, while the purchase cost and the resale price have perfectly correlated increments.

Now define a marginal option valuation problem: at any given time, the firm can exercise the call option to purchase, at price P_t, a package containing the currently marginal unit of K and a put option to sell it at price λP_t; alternatively, the firm can exercise the put option to sell, at price λP_t, the package containing the currently marginal unit of K and the call option to install the next unit of K.

Clearly, the value of the two options must, for given K, depend on the current values of η_t and P_t, which completely describe the state of the system; denote $F(\eta_t, P_t)$ the value of the call option, and $f(\eta_t, P_t)$ the value of the put option. Ito's lemma gives the dynamics followed by the two options when "alive" (not yet exercised). The expected rate of return on any unexercised option must then, by arbitrage, be such that the their holder earns the required rate of return: here, again, the presence of depreciation introduces some complications. The required rate of return on assets is r in the problem considered in Section 5:

since the value of the firm is the expected flow of cash-flow discounted at rate r, the opportunity cost of funds for the firm is equal to r. But unexercised options to install or uninstall units of K should instead earn a rate of return equal to $r+\delta$: when such options go unexercised capital depreciates at rate δ, and the index K of the <u>currently</u> <u>marginal</u> unit steadily decreases. The optimal stopping problem is defined on a "moving target": the asset that can be purchased (the <u>marginal</u> option) changes continuously, as capital depreciates.

Imposing then that unexercised options earn an expected rate of return equal to $(r+\delta)$, the following functional equations are obtained for the value of the two options <u>when alive</u>:

$$[5.3]\quad f_1(\eta,P)\eta\mu + f_2(\eta,P)\vartheta_p P + \frac{1}{2} f_{11}(\eta,P)\sigma_z^2\eta^2 + \frac{1}{2} f_{22}(\eta,P)\sigma_p^2 P^2 - (r+\delta)\, f(\eta,P) = 0$$

$$[5.4]\quad F_1(\eta,P)\eta\mu + F_2(\eta,P)\vartheta_p P + \frac{1}{2} F_{11}(\eta,P)\sigma_z^2\eta^2 + \frac{1}{2} F_{22}(\eta,P)\sigma_p^2 P^2 - (r+\delta)\, F(\eta,P) = 0$$

where subscripts denote partial derivatives.

The solution of these ordinary differential equations must satisfy the following boundary conditions:

$F(0,P)=0$ for $P>0$, because the option that gives its holder the right to purchase the asset with value $\eta/(r+\delta(1+\beta)-\vartheta_z)$ must be worthless when $\eta=0$ forever and the exercise price is positive (0 is an absorbing state for a geometric Brownian motion process);

$f(x,P)\to\infty$ as $x\to\infty$ if $P<\infty$, for the symmetric reason that an option to sell an asset with very large value at a finite price must become worthless as the value approaches infinity.

In view of the parabolic form of the functional equations, we can guess the following functional form for the two options:

[5.5] $F(\eta,P)= B_1 \ \eta^{\alpha_1}P^{\beta_1}$

[5.6] $f(\eta,P)= B_2 \ \eta^{\alpha_2}P^{\beta_2}$

with $\alpha_1>0$, $\alpha_2<0$ to satisfy the boundary conditions above.

The two options will be exercised on loci in (η,P) space that are implicitly defined by a function $\eta^*=g(P^*)$ for the call option and $\eta_*=h(P_*)$ for the call option, and satisfy the following conditions:

$$F(g(P^*),P^*)) + P^* = g(P^*)/(r+\delta(1+\beta)-\vartheta_z) + f(g(P^*),P^*)$$

$$F(h(P_*),P_*) + P_* = h(P_*)/(r+\delta(1+\beta)-\vartheta_z) + f(h(P_*),P_*)$$

These conditions simply impose that the exchanges of assets performed at the exercise points be "fair", i.e. that the value of options surrendered plus the exercise price be in each case equal to the value of assets and options received.

To solve the firm's problem, we need to determine g(.) and h(.): from the definition of η, knowledge of g(.) and h(.) will suffice to implement the optimal policy from the observation of the current values of K, Z and P.

Since the value of the assets to be received in exchange for a price proportional to P is proportional to η, it is intuitive (and can be formally proved in the present setting adapting the arguments of McDonald and Siegel[1986]) that the boundary loci should be homogeneous of degree 0 in η and P, i.e. that for some constants $\mathcal{u}$ and ℓ

$g(P)=\mathcal{u}P$

$h(P)=\lambda\ell P$

The following conditions are then obtained:

[5.7] $F(\mathcal{u}P^*,P^*) + P^* = \mathcal{u}P^*/(r+\delta(1+\beta)-\vartheta_z) + f(\mathcal{u}P^*,P^*)$

[5.8] $F(\lambda\ell P_*,P_*) + P_* = \lambda\ell P_*/(r+\delta(1+\beta)-\vartheta_z) + f(\lambda\ell P_*,P_*)$

Additional boundary equations are needed for determination of the boundaries and of the option values: these are the so-called "smooth pasting" or "high impact" conditions, which can be derived from the fact that the exercise boundary is freely chosen by the holder of the option to maximize the option's value.

Merton[1973, footnote 60] proves the necessity of the smooth pasting condition for a simpler problem (optimal exercise of an American call option), but his arguments are easily extended to the present setting where the exercise boundary is common to <u>two</u> linked option pricing problems (one an American put, one an American call) and the exercise price is uncertain. Merton notes that exercise points are chosen to maximize the value of the options; in the present setting, if extended option values $\mathcal{F}$ and $\mathcal{f}$ are defined over the space of possible exercise loci, it is true that

[5.8] $F(\eta,P)=\max\limits_{\mathcal{u},\ell} \mathcal{F}(\eta,P,\mathcal{u},\ell)$

[5.9] $f(\eta,P)=\max\limits_{\mathcal{u},\ell} \mathcal{f}(\eta,P,\mathcal{u},\ell)$

Replace $\mathcal{F}$ and $\mathcal{f}$ for F and f in [5.7] and totally differentiate with respect to $\mathcal{u}$; similarly, replace $\mathcal{F}$ and $\mathcal{f}$ for F and f in [5.8] and totally differentiate with respect to ℓ, to obtain (subscripts denote partial derivatives):

$$[5.10]\quad \mathcal{F}_1(uP^*,P^*,u,\ell)P^* + \mathcal{F}_3(uP^*,P^*,u,\ell) + P^* = P^*/(r+\delta(1+\beta)-\vartheta_z) + f_1(uP^*,P^*,u,\ell)P^* + f_3(uP^*,P^*,u,\ell)$$

$$[5.11]\quad \mathcal{F}_1(\ell\lambda P_*,P_*,u,\ell)P_*\lambda + \mathcal{F}_4(\ell\lambda P_*,P_*,u,\ell) + \lambda P_* = \lambda P_*/(r+\delta(1+\beta)-\vartheta_z) + f_1(\lambda\ell P_*,P_*,u,\ell)\lambda P_* + f_4(\lambda\ell P_*,P_*,u,\ell)$$

Now note that $\mathcal{F}_i=F_i$ and $f_i=f_i$ (i=1,2) by the definitions in [5.8] and [5.9]; and, since u and ℓ maximize f and F, necessarily $f_j=\mathcal{F}_j=0$ (j=3,4). Simplifying P^* and λP_* out of [5.10] and [5.11], the following additional boundary equations are obtained:

$$[5.12]\quad F_1(uP^*,P^*,u,\ell) + 1 = 1/(r+\delta(1+\beta)-\vartheta_z) + f_1(uP^*,P^*,u,\ell)$$

$$[5.13]\quad F_1(\ell\lambda P_*,P_*,u,\ell) + 1 = 1/(r+\delta(1+\beta)-\vartheta_z) + f_1(\lambda\ell P_*,P_*,u,\ell)$$

It is now easy (though tedious) to insert the functional forms in [5.5] and [5.6] into the differential equations [5.3] and [5.4] and the boundary conditions [5.7], [5.8], [5.12] and [5.13] to determine the exercise points u and ℓ, as well as the unknown constants B_1 and B_2 and the exponents α_1, α_2, β_1, β_2. As in section 4, it is found that $\beta_1=1-\alpha_1$ and $\beta_2=1-\alpha_2$ necessarily, to satisfy conditions [5.12] and [5.13] for all positive prices; replacing the partial derivatives of f(.,.) and F(.,.) into the functional equations [5.3] and [5.4] determines that $\alpha_1>0$ and $\alpha_2<0$

are the solutions to the same second-order equation found in Section 4; boundary conditions [5.12] and [5.13] can be used to solve for B_1 and B_2; and, finally, [5.7] and [5.8] give, after substitution of these values for B_1 and B_2, a pair of nonlinear equations to be solved for u and ℓ that are exactly equivalent to [4.6"] and [4.7"].

The marginal option pricing approach then gives the same solution to the firm's problem as the technique proposed in Section 4, based on the necessary conditions [2.1] and [2.2]. The economic intuition behind [2.1] and [2.2] is probably more straightforward than the thought experiment which underlies the marginal option pricing technique: the reduction of the dynamic program to a series of optimal stopping problems is awkward, especially in the presence of depreciation. But the techniques used in this section are probably more familiar to economists (at least to financial economists) than the mathematics adopted in Section 5; and extension of the model (for example to a Leontief production function as in Pindyck[1986], or to discontinuous processes for the forcing variables $\{W_t\}$) is probably easier if the marginal option pricing approach is adopted than otherwise.

It is now possible to verify existence of the optimal policy, i.e. finiteness of the value function, because an useful byproduct of the option pricing approach is the total value of the firm: this value must be the integral of the value of all infinitesimal units of K, both those installed (whose value is v as given in

[5.2] plus the value of the options to sell them, $f(K^{\beta}Z_t,P_t)$) and those that the firm may decide to install in the future (whose value of these is $F(K^{\beta}Z_t,P_t)$).

The total value of the firm is then

$$[5.14]\quad V^*(K_t,Z_t,P_t) = \int_0^{K_t}\left[\frac{x^{\beta}\,Z_t}{r+\delta-(\vartheta_z-\delta\beta)} + B_2\left(x^{\beta}\,Z_t\right)^{\alpha_2}P_t^{1-\alpha_2}\right]dx\, +$$

$$+\int_{K_t}^{\infty} B_1\left(x^{\beta}\,Z_t\right)^{\alpha_1}P_t^{1-\alpha_1}\,dx = =$$

$$= \frac{Z_t}{r+\delta-(\vartheta_z-\delta\beta)}\left[\frac{1}{1+\beta}\,x^{1+\beta}\right]_0^{K_t} + B_2\left(Z_t\right)^{\alpha_2}P_t^{1-\alpha_2}\left[\frac{1}{1+\beta\alpha_2}\,x^{1+\beta\alpha_2}\right]_0^{K_t} +$$

$$+ B_1\left(Z_t\right)^{\alpha_1}P_t^{1-\alpha_1}\left[\frac{1}{1+\beta\alpha_1}\,x^{1+\beta\alpha_1}\right]_{K_t}^{\infty}$$

The first two integrals in [5.14] are easily seen to be convergent as long as $-1<\beta<0$ and $\alpha_2<0$. Convergence of the last integral requires $1+\beta\alpha_1<0$, or $\alpha_1>-1/\beta$: this completes the study of the control program, since the value function (and hence the optimal policy) is shown to exist if and only if $\alpha_1>-1/\beta$.

Recalling [4.26], we have the following:

<u>Proposition</u> The control problem proposed at the beginning of Section 5 has a solution if and only if

$$r > -\vartheta_z\,\frac{1}{\beta} + \vartheta_P\left(1+\frac{1}{\beta}\right) + \frac{\sigma^2}{2}\,\frac{1+\beta}{\beta^2}$$

This says that the value of the firm will fail to be finite unless the expected real interest rate in terms of K, $r-\vartheta_p$, is sufficiently small compared to ϑ_z, the expected growth rate of the payoffs for given K. Intuitively, when the payoff function grows too quickly and/or the price of K falls too fast, control policies can be devised that produce cash flows streams whose expected value over the infinite future fails to converge when discounted at rate r.

6 - The characteristics of the optimal control process; Connections to the engineering literature

In the previous sections the optimal policy for the firm's problem under linear costs of control has been found, using techniques based on economically intuitive thought experiments. In this section the characteristics of the control process are discussed, and the technical literature that has studied similar problems in the abstract is referenced.

The reader should, first of all, be reminded of the characteristics of the Brownian motion process assumed above for the exogenous variables in the firm's problem, $\{P_t\}$ and $\{Z_t\}$. The Wiener process $\{W_t\}$, or standard Brownian motion, has the following important property: $W_{\tau 1}-W_{\tau 0}$ is distributed independently of $W_{\tau 2}-W_{\tau 1}$ for all $\tau 2>\tau 1>\tau 0$. This is very convenient, because it implies that the state of the system is completely described by the current value of W_t, independently of past events; but if the process has nonzero variance, this property implies that $\{W_t\}$ has infinite variation, i.e. that Brownian motion "fluctuates very fast" (which is the reason why $(dW_t)^2=dt$, a result repeatedly used above): Brownian motion moves both up and down in any interval of time, no matter how small.

It follows that the firm considered in the previous sections has to exercise control very quickly to maintain η_t/P_t (which would be driven by Brownian motion in the absence of control) between the control barriers u and $\ell\lambda$ at all times. In fact, X_t increases or decreases only when η_t/P_t is equal to one of the control barriers, and this only happens at distinct moments in time: control is never exercised throughout the length of any nonempty interval $[\tau_1,\tau_2]$. This is the reason why Proposition 2 in Section 3 above is completely unapplicable if uncertainty is described by functions of Brownian motion: whenever control is exercised, it is known that control will certainly not be applied in the next instant.

The stochastic process followed by the control process $\{X_t\}$ is called "singular" in the technical literature, and corresponds to the "local time" spent at the boundaries by the underlying Brownian motion process (see for example Harrison[1985] or Karatzas and Schreve[1988]): though continuous (since it is a continuous transformation of Brownian motion, which has continuous sample paths) it only increases or decreases on a time set which has total measure zero, being a collection of distinct points. The "singularity" of $\{X_t\}$'s sample paths lies in the fact that, though they are continuous, differentiable almost everywhere and with a derivative always equal to zero, they do decrease or increase so that $p(X_{\tau_2}-X_{\tau_1}\neq 0)>0$ for all $\tau_1<\tau_2$. At its points of increase or decrease, X_t moves infinitely fast, though it never jumps: the rate of control is infinite, making it impossible to use the classic Hamiltonian analysis.

Optimal "singular" control problems have been extensively studied in the engineering and operations research literature[1]. The abstract "tracking" problem considered by those authors is the following one: an operator controls the position of an object in $\mathbb{R}$ with the objective of minimizing the (possibly discounted) time integral of a convex loss function depending on the distance between the object and an exogenous stochastic process in continuous time, assumed to be a function of Brownian motion; control can be exercised to affect directly the position of the object in $\mathbb{R}$, but is costly in both directions ("fuel" is expended when control is exercised), or is actually prohibitively costly when exercised in one direction (the so-called "monotone follower" problem).

The problem considered at the beginning of Section 4 could be reduced in that form by the following transformation: a first-best policy for use of K could be defined as $\{K^{**}\}$, some function of the exogenous processes $\{Z_t\}$ and $\{P_t\}$; the firm would then try and "track" $\{K^{**}\}$ by varying K, but would not achieve perfect tracking because $\{K^{**}\}$ has infinite variation (moves up and down infinitely

---

[1] See for example Beneš et al.[1980], Chow et al.[1985]; Harrison[1985] reports many related results, and proposes applications to economic problems such as inventory or cash balances control.

often), and a control policy with infinite variation is infinitely costly as long as $P_t > p_t$ (the purchase price of K is strictly larger that its resale price). If $P_t = p_t$ then, of course, control is costless and the policy that continuously tracks $\{K^{**}\}$ is feasible.

Rather than deriving an expression for $\{K^{**}\}$ and invoking the technical literature, the two approaches taken above use economic intuition to derive the optimal control boundaries. The technical literature does note the equivalence between the full-fledged singular control problem and a sequence of optimal stopping problems[2], although, once again, the optimal stopping problem is highly abstract and devoid of economic content.

The technical literature also illuminates the connection between the "singular control" problem and the more usual Hamilton-Jacobi-Bellman equation approach to dynamic programming, which assumes that control is always applied at a finite rate: Chow et al.[1985] show how the "variational inequalities" they use to solve the singular control problem can be derived by taking the limit as $\varepsilon \to 0$ of the Bellman equation for a penalized problem, in which the allowable rate of control is bounded by $1/\varepsilon$: the penalized problem would have a solution of the "bang-bang" type, such that control would (continuously) be enacted at the maximum possible rate when it does take place, on time interval of finite size; allowing the maximum rate of control to become unbounded,

[2]See Karatzas and Shreve[1984,1985].

the time intervals over which control is enacted collapse to individual points, and the resulting control process is singular.

8- Concluding comments

This Chapter discusses at length the solution to a class of dynamic control problems, characterized by control costs that are linear in the displacement of the endogenous state variables. The solution is derived from first principles, highlighting the economic intuition underlying the mathematics; the connection to techniques studied in the technical literature, and to previous work by economists on factor demand policies, is made clear.

The equivalence between the dynamic programming and optimal stopping approaches is probably of independent interest. The latter approach is particularly appropriate in the presence of uncertainty, because when there are first-order adjustment costs undertaking any adjustment is a risky proposition: the path of exogenous variables can always turn out ex-post to be such as to render adjustment inappropriate, and reversion of the control decisions is costly. The optimal stopping technique sets out to minimize the risk implicit in any control decision.

The set of techniques discussed here has wide-ranging applications in economics: linear costs of adjustment produce non-certainty equivalence in the firm's factor demand policies, and models of investment and of labor demand have to be revised in the light of the results found here.

CHAPTER 2

IRREVERSIBLE CAPITAL ACCUMULATION AND THE STUDY OF INVESTMENT

"Capital" is the durable factor of production. A firm's decision about use of capital today cannot be unrelated to decisions about use of capital yesterday or tomorrow. If it were possible to rent capital services on a smoothly functioning spot market, firms could continuously adjust the amount of capital used in production, and the user cost of capital (Jorgenson[1963]) could be used in modelling demand for capital as the wage rate is used in modeling demand for labor.

But in reality the use of capital[1] cannot always be costlessly adjusted. To obtain an interesting and tractable dynamic programming problem, the assumption has often been made that although faster adjustment of the capital stock incurs increasing costs, the cost of infinitesimal adjustments is zero - in other words, capital adjustment costs are assumed to be strictly convex, and to achieve a minimum at some level (at zero net investment, at zero gross investment or maybe at some "normal" level of investment). These adjustment costs may be internal to the firm, or external to it and due to decreasing returns in production of investment goods.

Most models of investment assume a certain environment for the firm, and show that the ratio of the value of the firm to the replacement cost of its capital cost (Tobin's Q) is the only determinant of the investment decision if the firm produces under

[1] or of labor: see Chapter 3 in this thesis. Modelling the joint determination of several factor demands under linear costs of adjustment should have high priority in future research.

constant returns to scale and is perfectly competitive (see for example Hayashi[1986]). It is clear that under these assumptions there would be no reason for individual firms to exist, in partial equilibrium: the size of the firm is only bounded by the capital adjustment costs. Under the same assumptions, but allowing for uncertainty, Abel[1983,1985] provides an explicit solution to the dynamic investment problem, making simplifying assumptions (similar to those made below) about the form of the stochastic process facing the firm, of the demand function and of the production function. Lucas and Prescott[1971] and Prescott[1973] prove existence of competitive and oligopolistic equilibria under uncertainty and convex costs of adjustment, giving some characterization results. If costs of adjustment are modelled as quadratic, it is possible to obtain characterizations of the investment process in the presence of uncertainty (see for example Sargent[1979a]), because the Euler equation that the optimal invesmtment rule must satisfy is linear, yielding certainty equivalence.

The models of investment based on convex adjustment costs have not been very successful empirically (see Abel and Blanchard[1986], Hall[1987]), and one explanation of their poor performance may be the assumption of a well-behaved adjustment cost function.

Realistically, investment in productive capital is often irreversible. From a macroeconomic point of view, industrial plants are next to worthless unless used in production, because their direct consumption value is clearly very low. From a microeconomic point of view, many productive facilities are firm-specific, their reconversion to alternative uses being costly if at all possible. Irreversibility is then potentially very important for empirical studies of investment behavior.

Moreover, investment irreversibility can provide insights for the theoretical and empirical treatment of aggregate prices, production and employment: the dynamic behavior of these variables will be non-standard if investment decisions are irreversible, because firms will sometimes find themselves stuck with a larger capital stock than the desired one. Ex-post, the cost of installed capital is sunk; positive shocks can have very long lasting consequences if they induce firms to invest and the capital stock can then only be reduced by depreciation. In an open economy this is probably relevant to some degree for the medium and long run effects of exchange rate shocks.[2] On the other hand investment will occur in spurts, when costs and prices are favorable enough for firms to exercise their option to invest; such a nonlinear investment function is very different from the one assumed in the standard multiplier-accelerator models, and could give firm theoretical foundations to Hicksian trade cycle theory.

Arrow [1968], Nickell [1974] and others have studied irreversible investment decisions in a partial equilibrium,

[2] see Baldwin and Krugman [1986] for a seminal treatment of these issues, and Dixit[1987] for a more sophisticated treatment.

dynamic optimization framework, assuming however that firms hold certain expectations about the cyclical path of exogenous variables. The same certainty equivalence assumption underlies the (mostly empirical) literature on "putty-clay" models of investment, which assume that not only machine tools have no value unless used in production but the labor requirement of existing machines is fixed (see for example Ando et al.[1974]).

Sargent[1979b] developed a simple general equilibrium model of irreversible capital accumulation, adopting a framework similar to the one in Lucas and Prescott[1971] and obtaining mostly qualitative results.

Irreversible investment <u>under uncertainty</u> has been extensively studied by financial economists (see McDonald and Siegel [1986] and their references, as well as Ingersoll and Ross[1987] for the case of interest rate uncertainty). Option pricing techniques provide elegant solutions in the case of an individual irreversible investment project with uncertain payoffs: such a project will be adopted only when the expected discounted payoff from investment <u>exceeds</u> the cost by an amount that depends on the level of uncertainty, and can be impressively large for plausible parameter values. Even risk neutral firms are, in a sense, reluctant to invest when projects are irreversible and the future is uncertain: if the irreversible project is adopted the option to wait for some of the uncertainty to be resolved is given up, and options are valuable even to risk-neutral agents. These models primarily study the decision to adopt an individual investment project whose payoff is independent of past and future investment decisions. The results are clearly relevant to the study of investment from a macroeconomic point of view: Bernanke[1983] notes that the level of uncertainty perceived by firms is likely to vary cyclically, and emphasizes that irreversibility effects are important for the understanding of the cyclical behavior of aggregate investment.

But most of the option valuation models so far available only consider the optimal timing for the adoption of an individual project with given characteristic: neglecting the availability of many investment projects of different sizes and with different characteristics at different points in time, they do not provide a proper dynamic investment model. The important fact that adoption of an investment project today changes the menu of projects available tomorrow should be explicitly taken into account to clarify the relationship of the option pricing models to the more traditional macroeconomic investment models. Pindyck[1986] applies option pricing techniques to the _marginal_ investment decision: his firm sequentially decides the optimal amount of capacity to be installed, knowing that future demand (and production) are uncertain and follow a geometric Brownian motion stochastic process with known parameters.

This Chapter studies _irreversible_ putty-putty investment under uncertainty: capital has no value unless used in production, but can ex-post be optimally combined with other factors to adapt to news about the variables exogenous to the firm. Using the results of Chapter 1, it is possible to reconcile the option-based financial literature with Arrow's and Nickell's results under certainty. The solution to the irreversible capital accumulation problem has a closed form under a simplifying (but not totally unrealistic) set of assumptions. Using this closed form solution permits a straightforward comparison between the behavior of investment under irreversibility and its behavior in the more usual convex-costs-of-adjustment models; moreover, the dynamic behavior and the ergodic distribution of the value of the firm, of marginal Q and of average Q are easily derived, and the implications of investment irreversibility for empirical work on investment can be discussed.

Section 1 describes a simplified model of a firm faced by exogenous uncertainty. Section 2 obtains the solution to the investment problem, and discusses the dependence of its form on the degree of uncertainty facing the firm. Section 3 derives the long-run characteristics of irreversible capital accumulation; Section 4 discusses the implications of investment irreversibility for empirically relevant observable variables: the value of the firm, the shadow price of capital relative to the purchase price of capital, and Tobin's Q. Section 5 considers the implications of the results for the interpretation of empirical evidence on investment behavior, and comments on the realism and relevance of the irreversibility constraint. Section 6 concludes.

1 - A model of production and sales.

In partial equilibrium, a firm is characterized by its production and demand functions, and by the form of the stochastic processes it takes as given. Consider then a firm endowed with a Cobb-Douglas production function[3] and a constant elasticity demand function:

$$[1.1] \quad Q_t=\left[K_t^{\alpha}(A_tL_t)^{1-\alpha}\right]^{\phi} \qquad 0<\alpha<1 \ , \ \phi>0$$

$$[1.2] \quad B_t= D_tQ_t^{\mu-1} \qquad 0<\mu\phi<1$$

where Q_t denotes production and sales at instant t (inventories are assumed away for simplicity); ϕ indexes the return to scale in production: constant returns to scale are given by $\phi=1$. B_t is the product price at time t, and μ indexes the firm's monopoly power: the inverse of the markup factor is equal to μ, and the firm's monopoly power increases as μ approaches zero. For a competitive firm μ equals 1. The factor of production L_t, "labor", is perfectly flexible and can be rented at the instantaneous price w_t; A_t is an index of technological progress. The capital stock,

---

[3] the Cobb-Douglas production function is of course the workhorse of investment theory; recent applications related to this Chapter, in that the uncertainty facing the firm is modeled in continuous time, include: Dietrich and Heckerman [1980], who solve for the one-time choice of capital stock by a competitive firm producing with decreasing returns to scale; Abel [1983, 1985], who studies the investment problem under constant returns and perfect competition when there are constant-elasticity costs of capital stock adjustment; McDonald and Siegel [1985] extend Dietrich and Heckerman's model to the case where there is a fixed cost for production, so that the factory may be shut down; and McDonald and Siegel [1986] derive the value of an investment opportunity in a Cobb-Douglas plant of fixed size, and the optimal timing of its adoption.

K_t, can be increased at any time by paying the unit price P_t, but installed capital has no value unless used in production (i.e. its resale price is zero). The parameter D_t influences the position of the demand curve (it may be a function of the consumers' income, or of a price index for substitutes, that the firm takes as given in its optimization): if $\mu=1$, $D_t=B_t$ is simply the market price, which a competitive firm takes as given.

Define the operating-profits function

$$[1.3]\quad \Pi\left[K_t,\ w_t,\ D_t\right] \equiv \max_{L_t}\ B_tQ_t-w_tL_t \qquad \text{subject to [1.1], [1.2]}$$

Since there are no constraints on the adjustment of labor input, this maximum is always attained by the firm; some algebra shows that operating profits can be written as

$$[1.4]\quad \Pi\left[K_t,\ w_t,\ D_t\right] = \frac{1}{1+\beta}\, K_t^{\,1+\beta}\, Z_t$$

where

$$\beta \equiv \frac{\phi\mu-1}{1-(1-\alpha)\phi\mu}\ , \quad -1<\beta<0$$

$$Z_t \equiv \frac{\alpha\phi\mu}{1-(1-\alpha)\phi\mu}\left[\left[\phi\mu(1-\alpha)\right]^{\frac{(1-\alpha)\phi\mu}{1-(1-\alpha)\phi\mu}} - \left[\phi\mu(1-\alpha)\right]^{\frac{1}{1-(1-\alpha)\phi\mu}}\right]\left[1+\beta\right] D_t^{\frac{1}{1-(1-\alpha)\phi\mu}}\left[w_t/A_t\right]^{\frac{-(1-\alpha)\phi\mu}{1-(1-\alpha)\phi\mu}}$$

The variable Z_t summarizes at every instant the business conditions for the firm: it is higher the higher is the demand indicator D_t, and the lower is the ratio of the flexible-factor rental cost w_t to its productivity A_t.

Uncertainty is introduced in the model by the assumption that $\{w_t\}$, $\{A_t\}$, $\{D_t\}$ and the purchase price of capital, $\{P_t\}$, are stochastic processes described by geometric Brownian motion in continuous time. In words, it is assumed that wages, price of capital and demand are always expected to grow at some constant rate, but the growth rate fluctuates randomly, so that the outlook farther and farther ahead is increasingly uncertain. Future values of the exogenous variables are jointly lognormally distributed, conditional on their current values, with variance proportional to the length of the forecast interval.[4]

It is then easy to show that $\{Z_t\}$, being a constant-elasticity function of geometric Brownian motion process, follows itself a geometric Brownian motion process (multiplicative functions of lognormally distributed random variables are lognormally distributed); denote ϑ_z the drift parameter of $\{Z_t\}$ and σ_z its standard deviation parameter, and similarly define parameters for the stochastic process of the purchase price of capital:

[1.5] $dZ_t=Z_t\vartheta_z dt+Z_t\sigma_z dW_{zt}$

[1.6] $dP_t=P_t\vartheta_p dt+P_t\sigma_p dW_{pt}$

[4]All variables shculd of course be expressed in real terms, i.e. wages and demand should be deflated by the price of the consumption basket of the firm's owners.

where dW_{zt} and dW_{pt} are the increments of (possibly correlated) standard Wiener processes.

The parameters ϑ_z and σ_z are linear combinations of the drifts, variances and covariances of the wage, productivity and demand processes, with weights depending on technology and demand parameters. In empirical work it would be necessary to estimate the importance of each source of uncertainty for the firm; but here $\{Z_t\}$, the shifter of the reduced-form profit function, is taken as the primitive exogenous variable in the firm's problem.

The parameter β indexes the concavity of the reduced-form profit function with respect to the installed capital stock. The shape of the reduced-form profit function depends on the degree of monopoly power and returns to scale, as well as on the Cobb-Douglas share of the flexible factor in production. If $\mu\phi=1$ then $\beta=0$, and the reduced form profit function is linear in K and the firm is indifferent to the level of the capital stock; if $\mu=0$ (a unit-elastic demand function), then $\beta=-1$ and arbitrarily large profits can be obtained by producing arbitrarily small amounts. In both these cases the value-maximization problem of the firm is ill-defined.

2 - Optimal irreversible investment decisions

Assume that the objective of the firm's managers is to maximize the discounted expected value of profits, and that

- r, the required rate of return, is constant;
- the installed capital stock depreciates at the constant exponential

rate δ.

The firm's problem at any time t is then to choose a contingent investment rule $\{X_t\}$, or - which is the same - a stochastic process $\{K_t\}$, to maximize its value: the state of the system is completely described by three state variables: the installed capital stock K_t; the state of affairs Z_t; and P_t, the price at which additional capital can be purchased. Denote $V^*(K_t,Z_t,P_t)$ the firm's value if the managers follow the optimal investment rule.

As noted in Chapter 1, "irreversibility" is implied by the resale price of capital being equal to (or lower than) zero, given that the marginal contribution of capital to profits as defined above is nonnegative: the firm will never discard installed capital if the resale price is zero, since capital in place cannot decrease the operating profits. But the optimality conditions can be better understood if the irreversibility constraint is explicitly imposed and constrained maximization theory is used on the value function.

Define then the value function, and the related optimization problem, as follows:

$$[2.1] \quad V^*(K_t,Z_t,P_t) = \underset{\{X_\tau\}}{\mathrm{Max}}\, E_t\left\{\int_t^\infty e^{-r(\tau-t)}\left[\frac{1}{1+\beta}K_\tau^{1+\beta}\, Z_\tau d\tau - P_\tau dX_\tau\right]\right\}$$

subject to $dK_\tau = -\delta K_\tau d\tau + dX_\tau$ (capital stock dynamics)

and to $dX_t \geq 0$ (irreversibility constraint)

Note that jumps in the capital stock are not excluded a priori: the gross investment rate dX_t/dt could well be infinite, and for this reason the amount of control is not multiplied by dt in the capital dynamics equation. The second integral in [2.1] is defined in the Stiltjes sense, with $\{X_t\}$ as the integrating function.

The expectation E_t in [2.1] is taken over the joint distribution of the $\{K_t\}$, $\{P_t\}$ and $\{Z_t\}$ processes, conditional on the information available at time t, taking into account that investment decisions will be taken optimally (subject to the irreversibility constraint) in the future.

At any time t, the irreversibility constraint imposes that $dX_t \geq 0$, or, which is the same, $K_{t+} \geq K_{t-}$; heuristically, the following Kuhn-Tucker first order conditions with respect to K_t (or to dX_t) must necessarily hold:

$$[2.2] \quad E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} K_\tau^\beta \, Z_\tau \, d\tau\right\} = P_t \quad \text{if } dX_t>0$$

$$[2.3] \quad E_t\left\{\int_t^{\infty} e^{-(r+\delta)(\tau-t)} K_\tau^\beta \, Z_\tau \, d\tau\right\} < P_t \quad \text{if } dX_t=0$$

The left-hand side of [2.2] and [2.3] is the derivative of V^* with respect to K_{t+}, the shadow price of capital. These first order conditions are based, as usual, on a feasible perturbation argument in which all controls except one are assumed given (in probability distribution): the currently marginal unit of installed capital is therefore viewed, allowing for depreciation, as the marginal unit throughout the infinite future. The firm knows that conditions like [2.2] and [2.3] will be satisfied at all future times: this defines the (yet to be found) probability distribution of future capital stocks, which is used in taking the expectation in [2.2] and [2.3].

We now provide sufficient conditions for existence of a solution to [2.1], and sketch a proof of uniqueness of the optimal investment rule:

<u>Proposition</u>

a) Provided that

$$r > \left[\vartheta_p \left(1 + \frac{1}{\beta}\right) - \vartheta_z \frac{1}{\beta} + \frac{\sigma^2}{2} \frac{1+\beta}{\beta} \frac{1}{\beta}\right]$$

where

$$[2.4] \quad \sigma^2 \equiv \sigma_z^2 + \sigma_p^2 - 2\,\rho\,\sigma_z\sigma_p \;, \quad \rho \equiv \left[dW_{zt} dW_{zt}\right]/dt$$

is the variance of the rate of increase of the process $\{Z_t/P_t\}$, the irreversible investment problem has a solution, i.e. the value function V^* defined in [2.1] exists (is bounded);

b) if an investment rule [R] can be found that satisfies [2.2] or [2.3] at all times, then [R] is the unique solution of the optimization problem [2.1].

<u>Proof</u>:

a) If investment were reversible, so that the firm could at any time buy or sell capital at price P_t, the risk-neutral manager's problem would be the same as if there were a rental market for capital, with instantaneous rental rate $(r-\vartheta_p+\delta)P_t$; then the capital stock would always satisfy the first order condition

$$[2.5] \quad K_\tau^\beta \, Z_\tau = (r-\vartheta_p+\delta)P_\tau$$

implying that $$K_\tau = \left[(r-\vartheta_p+\delta)\frac{P_\tau}{\quad}\right]^{1/\beta}$$

and that $$\Pi\left(K_\tau, Z_\tau\right) = \frac{1}{1+\beta}\left[(r-\vartheta_p+\delta)\frac{P_\tau}{Z}\right]^{\frac{1+\beta}{\beta}} Z_\tau$$

Consider the present discounted value of operating profits under conditions of <u>reversible</u> investment: noting that the order of expectation and integration can be reversed, by Fubini's theorem, as long as both operations are well-defined, we have

$$[2.6] \quad E_t\left\{\int_t^\infty e^{-r(\tau-t)} \frac{1}{1+\beta} K_\tau^{1+\beta} Z_\tau \, d\tau\right\} = \int_t^\infty e^{-r(\tau-t)} \frac{1}{1+\beta} E_t\left\{K_\tau^{1+\beta} Z_\tau\right\} d\tau =$$

$$= \int_t^\infty e^{-r(\tau-t)} \frac{1}{1+\beta}\left(r-\vartheta_p+\delta\right)^{\frac{1+\beta}{\beta}} E_t\left\{\left[\frac{P_\tau}{Z_\tau}\right]^{\frac{1+\beta}{\beta}} Z_\tau\right\} d\tau$$

Now note that $\left[\frac{P_\tau}{Z_\tau}\right]^{\frac{1+\beta}{\beta}} Z_\tau$, being a constant-elasticity combination of lognormal variables, is lognormally distributed given the information available at time $t<\tau$: ; the easiest way to find its expectation is to use Ito's lemma, that gives

$$d\left[P_t^{1+1/\beta} \, Z_t^{-1/\beta}\right] =$$

$$= P_t^{1+1/\beta} \, Z_t^{-1/\beta}\left[\left(\vartheta_p(1+\frac{1}{\beta}) - \vartheta_z\frac{1}{\beta} + \frac{\sigma^2}{2}\frac{1+\beta}{\beta}\frac{1}{\beta}\right) dt + \right.$$

$$+ \frac{1+\beta}{\beta}\sigma_p \, dW_{pt} - \frac{1}{\beta}\sigma_z \, dW_{zt} \Bigg]$$

It follows that

$$E_t\left\{\left[\frac{P_\tau}{Z_\tau}\right]^{\frac{1+\beta}{\beta}} Z_\tau\right\} = \left[\frac{P_t}{Z_t}\right]^{\frac{1+\beta}{\beta}} Z_t \quad e^{\left[\vartheta_p(1+\frac{1}{\beta})-\vartheta_z\frac{1}{\beta} + \frac{\sigma^2}{2}\frac{1+\beta}{\beta}\frac{1}{\beta}\right](\tau-t)}$$

and it is easy to see that the integral in [2.6] converges if and only if

$$[2.7] \quad r > \left[\vartheta_p(1 + \frac{1}{\beta}) - \vartheta_z \frac{1}{\beta} + \frac{\sigma^2}{2}\frac{1+\beta}{\beta}\frac{1}{\beta}\right] \text{, as was to be shown.}$$

Imposing the irreversibility constraint and subtracting investment expenditures can only decrease the value of the firm, which therefore is bounded.[5]

b)[6] By standard Kuhn-Tucker theory, uniqueness of the optimal investment policy follows from concavity of the maximand, V^*, at times when investment is positive: the first order condition [2.2] is then sufficient as well as necessary. (As to the alternative first order condition [2.3], it is clearly not optimal to invest when it applies, since positive investment would immediately decrease the value of the firm).

Concavity of $V^*(K_t,Z_t)$ follows from concavity in K_t of the

---

[5] [2.7] turns out to be <u>necessary</u>, as well as sufficient, for existence of the value function. See the expression for the value of the firm in [4.1] below.

[6] Ricardo Caballero suggested this line of proof.

instantaneous operating profit function $\Pi(K_t, Z_t)$; since K_{t+} increases linearly with dX_t, the instantaneous cash flow function $\Pi(K_t, Z_t) - P_t dX_t$ is concave in dX_t. The value function $V^*(K_t, Z_t)$ is the integral of instantaneous cash flows sample paths, over states of nature and time, with positive measure (the joint probability measure of density of Z_τ, P_τ, K_τ, and the discount factor); to prove concavity of $V^*(.,.)$ in its first argument it is sufficient to show that K_τ and $-dX_\tau$ are nondecreasing in K_t, $t<\tau$.[7] But investment irreversibility implies that more investment at time t, all else being equal, will never result in a lower K_τ or in a higher dX_τ ($\tau>t$), and therefore concavity of $V^*(K_t, Z_t)$ is guaranteed.
[end of proof]

The firm's dynamic optimization problem is, on the basis of this proposition, completely solved if an investment rule that satisfies [2.2] and [2.3] can be found; this is typically a difficult task, but in our framework the following can be shown to be true using the results of Chapter 1:

1) Under the assumptions given above, conditions [2.2] and [2.3] are satisfied if the firm invests following the rule:

---

[7] a nondecreasing function of a concave function is weakly concave; the <u>current</u> cash flow is strictly concave; and sums (or integrals) of concave functions are strictly concave if one of the elements is strictly concave.

Whenever possible, install more capital so as to satisfy

$$[R] \quad \frac{\partial \pi(K_t, Z_t)}{\partial K_t} \frac{1}{P_t} = \frac{K^\beta Z_t}{P_t} = \frac{A}{A-1}\left[r + \delta + \delta\beta - \vartheta_z\right] \equiv c^*$$

where

$$A = \frac{-\left[\vartheta_z - \delta\beta - \vartheta_p - \frac{\sigma^2}{2}\right] + \sqrt{\left[\vartheta_z - \delta\beta - \vartheta_p - \frac{\sigma^2}{2}\right]^2 + 2\left[r + \delta - \vartheta_p\right]\sigma^2}}{\sigma^2} > -\frac{1}{\beta} \quad 8$$

and $\sigma^2 \equiv \sigma_z^2 + \sigma_p^2 - 2\,\rho\,\sigma_z\sigma_p$

NOTE: c^* , the ratio of marginal productivity of capital to the purchase price of capital at times when investment is positive, is a constant, i.e. it does not depend on time nor on P_t nor on Z_t.

2) The shadow value of capital is, if the firm adopts rule [R],

$$[2.8] \quad E_t\left\{\int_t^\infty e^{-(r+\delta)(\tau-t)} K_\tau^\beta Z_\tau \, d\tau\right\} = \frac{K^\beta Z_t - \frac{1}{A}\left[\frac{K_t^\beta Z_t}{c^* P_t}\right]^{A-1} K_t^\beta Z_t}{r + \delta + \delta\beta - \vartheta_z}$$

---

8 $(-\beta A)$ is the positive solution to the quadratic equation

$$\frac{\sigma^2}{2\beta^2} X^2 - \left[\vartheta_z \frac{1}{\beta} + \delta - \vartheta_p \frac{1}{\beta} - \frac{\sigma^2}{2}\frac{1}{\beta}\right] X - \left[r + \delta - \vartheta_p\right] = 0$$

and it is possible to verify that $-\beta A > 1$, implying that $A > -1/\beta$, as long as the condition in [2.7] holds.

3) The value of each installed unit of capital, regardless of the possibility of installing more capital in the future, is

$$[2.9] \quad v\left[K, Z_t\right] = \frac{K^{\beta} Z_t}{r+\delta-(\vartheta_z-\delta\beta)} \quad , \; K \le K_t$$

while the value of the opportunity to install capital in the future is given by[9]

$$[2.10] \quad F\left[K, Z_t, P_t\right] = \frac{1}{A-1}\left[\frac{K^{\beta} Z_t}{c^{*} P_t}\right]^{A-1} \frac{K^{\beta} Z_t}{c^{*}} \quad , \; K \ge K_t$$

It is possible to verify that, under the investment rule [R], whenever investment is positive the shadow value of capital is equal to P_t (so that condition [2.2] is satisfied) <u>and</u> the following holds true:

$$v\left[K_t, Z_t\right] = P_t + F\left[K_t, Z_t, P_t\right]$$

which can be interpreted to say that whenever a marginal unit of capital is installed it is the case that v(.), the expected discounted value of marginal profits from the currently marginal units, exactly compensates the out-of-pocket cost of installation,

---

[9]This formula is the same as that found by McDonald and Siegel[1986] for a single investment opportunity of given size. As noted in Chapter 1, each infinitesimal increment of the installed capital stock defines an investment opportunity (of infinitesimal size) of the kind considered by McDonald and Siegel's optimal stopping problem.

P_t, plus the opportunity cost of immediate installation, F(.): the firm could delay installation and learn something about the future evolution of the business conditions, as summarized by $\{Z_t\}$, and of the price of capital $\{P_t\}$; this opportunity to wait has value, because in the immediate future the price of capital might decrease (making delayed installation of the same unit less costly) or business conditions might deteriorate, decreasing v(.) and making installation of the currently marginal unit of capital unprofitable.

Figure 1 plots c^* as a function of σ for several values of β; values for the other parameters are given in the figure. The firm is more reluctant to invest the higher is the variability of its environment, summarized by σ (note that σ is a combination of variances and covariances of the processes for demand, wage, productivity, and capital price).

This is not surprising, because higher demand variability worsens the "worst case" scenario, in which the firm regrets the irreversible investment decision. Higher variance does not symmetrically improve the "best case" scenario: whenever demand increases, or wage decreases, or price of capital falls, the firm can easily increase the capital stock. The irreversibility constraint only binds in the case of adverse realizations of uncertainty, and from the point of view of the firm effectively truncates the (lognormal) probability distribution of future states of nature. Bernanke[1983] refers to this insight as the " 'bad news principle of irreversible investment' ... of all

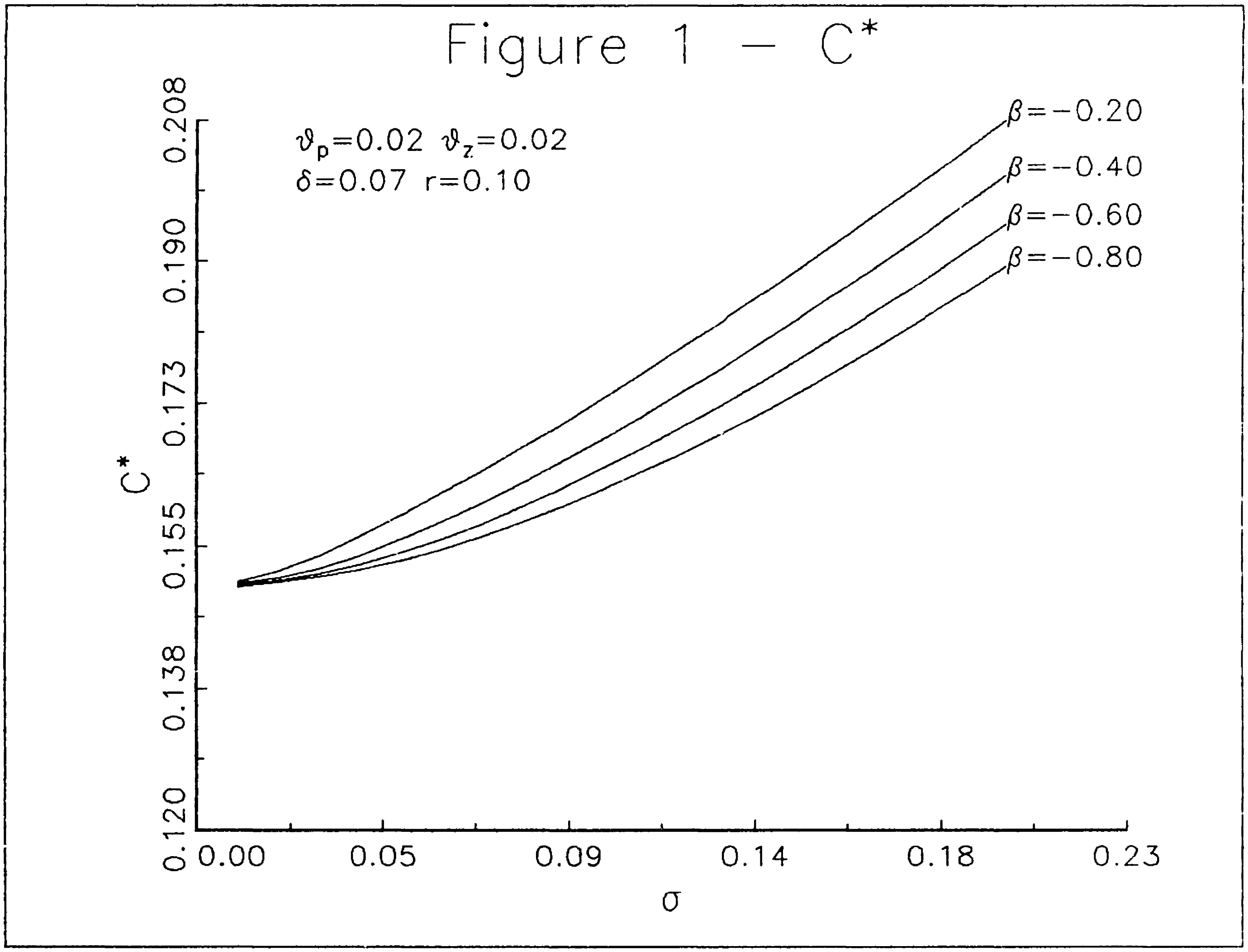


Figure 1 – C*
ϑp=0.02 ϑz=0.02
δ=0.07 r=0.10
β=−0.20
β=−0.40
β=−0.60
β=−0.80
C*
0.120
0.138
0.155
0.173
0.190
0.208
0.00
0.05
0.09
0.14
0.18
0.23
σ

possible future outcomes, only the unfavorable ones have a bearing on the current propensity to undertake a given project ...".

When $\sigma \to 0$ it appears from the figure that $c^* \to (r-\vartheta_p+\delta)$, the Jorgenson[1963] rental cost of capital: it can be verified algebraically (using l'Hopital's rule) that this is indeed the case as long as

[2.11] $\vartheta_z - \delta\beta - \vartheta_p > 0$

[2.11] implies that, under certainty, the irreversibility constraint is never binding (the firm's desired dotation of capital steadily increases).

If [2.11] is not satisfied, in the absence of uncertainty the firm never wants to increase the capital stock, except when it is set up and a capital stock is chosen taking the irreversibility constraint into account: it is again possible to verify that, as $\sigma \to 0$, [R] converges to the appropriate limit for this case if [2.11] is not true.

The value of the marginal productivity of capital that triggers investment is higher, for a given σ, if β is lower in absolute value, i.e. the instantaneous operating profit function is less concave in K. The reason for this is that, for a given $\delta > 0$, a large $|\beta|$ implies that the marginal profitability of capital increases more rapidly when gross investment is zero.

It is of course possible to invert the marginal condition in [R] and find an expression for the firm's desired capital stock as a function of the current value of Z and P:

$$K^{*}\left[Z_t, P_t\right] = \left[c^{*}\ \frac{P_t}{Z_t}\right]^{1/\beta} = \left[\frac{A}{A-1}\ \left[r + \delta + \delta\beta - \vartheta_z\right]\ \frac{P_t}{Z_t}\right]^{1/\beta}$$

Recall that $\beta<0$, so the desired capital stock is higher the higher is Z (i.e. the lower is the wage and the stronger is demand), and the lower is P_t. Figure 2 plots the desired capital stock for <u>given</u> Z and P, as a function of σ, for several values of β: higher uncertainty implies a lower desired capital stock, because the firm knows that the "worst case" is very likely to occur and hedges against possible decreases in Z or P.of capital is decreasing in $|\beta|$ for given P, Z and σ. Desired capital decreases in $|\beta|$ for given P, Z and σ: a firm with high $|\beta|$ has more monopoly power and/or more strongly decreasing returns to scale, hence tends to supply less, and use less capital, to maximize its profits.

This completes the normative analysis of the irreversible capital accumulation problem: if the current capital stock K_t is smaller than $K^{*}(Z_t, P_t)$, the firm immediately invests so that $K_t=K^{*}$; if K_t is larger than K^{*}, the firm does not invest (and the capital stock is reduced by depreciation). Of course, since downward fluctuations of Z are possible, the firm will sometimes regret the investment decision; similarly, the firm will regret having invested when P decreases at a rate higher than the required rate of return: the same investment opportunity exercised yesterday could be more profitably exercised today.

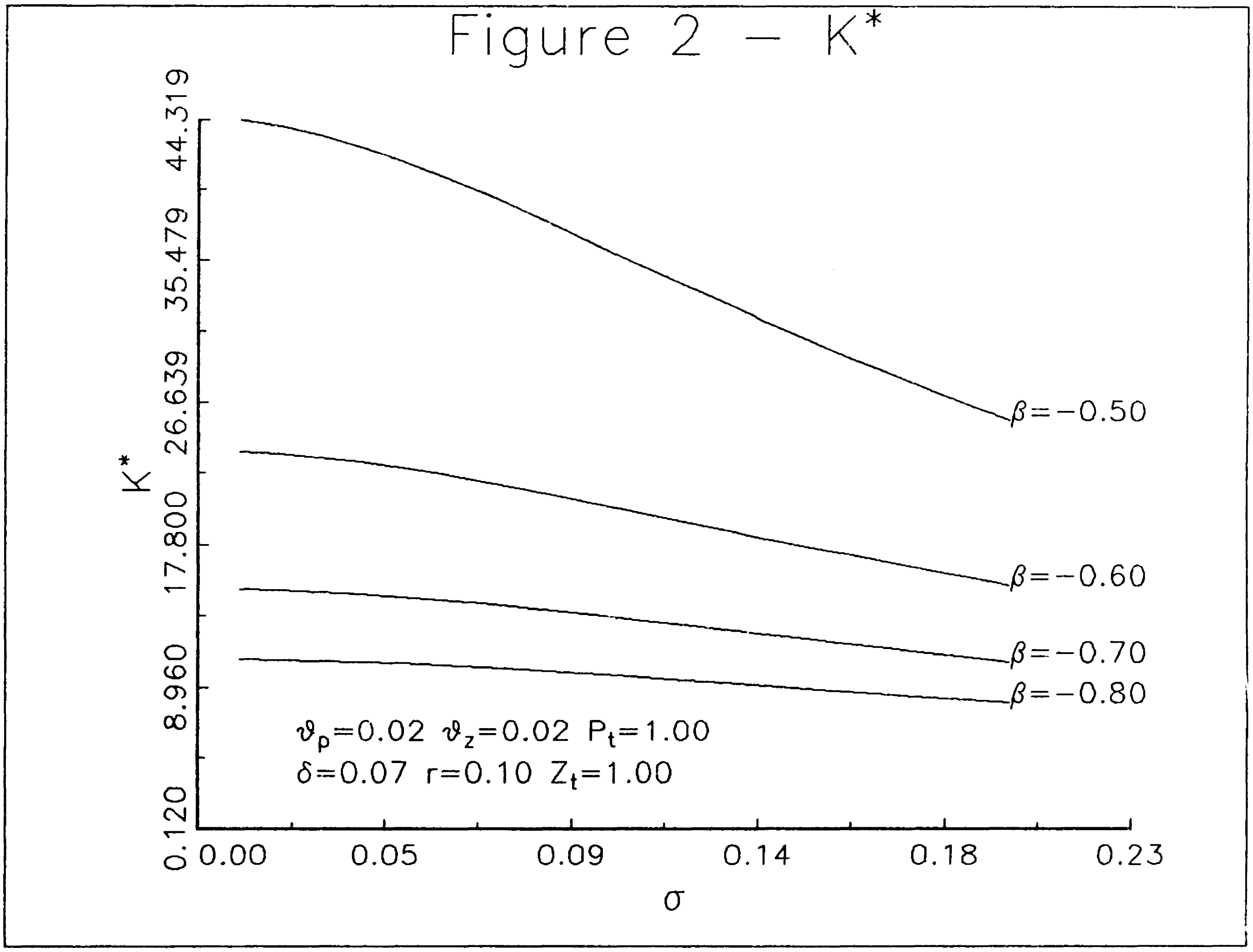


Figure 2 — K*
K*
44.319
35.479
26.639
17.800
8.960
0.120
0.00
0.05
0.09
0.14
0.18
0.23
σ
β=−0.50
β=−0.60
β=−0.70
β=−0.80
ϑp=0.02 ϑz=0.02 Pt=1.00
δ=0.07 r=0.10 Zt=1.00

3 - The steady-state characteristics of irreversible capital accumulation

The previous section has derived and characterized the optimal irreversible investment rule, which could be of interests to some firm's managers. Economists might also be interested in a somewhat different perspective on the problem: if in fact managers knew all along the optimal irreversible investment rule (and therefore did not need to read the previous section), what would be the empirically observable characteristics of optimal irreversible capital accumulation?

Usually, normative theory straightforwardly characterizes the behavior of endogenous variables: agents are provided by a set of rules that at all times determines their behavior, and agents' behavior uniquely determines the endogenous variables' paths. But this is not true of the irreversible investment problem: the managers of the firm do <u>not</u> continuously control the capital stock. The model's endogenous variables have autonomous dynamics most of the time, and determination of the characteristics of the model from the point of view of an outside observer is of independent interest.

The dynamic behavior of the model is interesting, though difficult to describe formally. Investment occurs in spurts, whenever the price of capital and business conditions are sufficiently favorable.

This section is more simply concerned with the <u>long-run</u> characteristics of irreversible capital accumulation: given that the firm is following the rule described in the previous section, what should an observer (who is ignorant of the firm's history) expect about the relationship of the firm's capital stock to other observable quantities at any given moment in time?

To answer this, we have to compute the "ergodic" or "steady state" distribution of some variable. The exogenous processes are assumed above to be nonstationary, and therefore fail to possess a steady-state distribution: but there are functions of the exogenous processes and of the installed capital stock which do possess a steady state distribution.

The ratio of the marginal profitability of currently installed capital to the current purchase price of capital plays an important role in the investment rule; it will be convenient to define

[3.1] $\xi_t = K_t^{\beta} Z_t / P_t$

This quantity follows a <u>regulated</u> geometric Brownian motion if the firm follows rule [R]; Ito's lemma can be used to derive the stochastic differential of ξ_t when the firm is not investing:

$$d\xi_t = d[\, K_t^{\beta} Z_t P_t^{-1} \,] =$$

$$= \beta \; K_t^{\beta-1} Z_t P_t^{-1}(-\delta K_t dt) + K_t^{\beta} P_t^{-1}(dZ_t) - K_t^{\beta} Z_t P_t^{-2}$$

$$(dP_t) + 2 \frac{1}{2} K_t^{\beta} Z_t P_t^{-3} (dP_t)^2 - K_t^{\beta} P_t^{-2}(dZ_t dP_t) =$$

$$= \xi_t(-\delta\beta+\vartheta_z-\vartheta_p+\sigma_p^2-\rho\sigma_z\sigma_p)dt + \xi_t(\sigma_z dW_{zt} - \sigma_p dW_{pt})$$

Denote m the drift parameter of the geometric Brownian motion process followed by $\{\xi_t\}$ when no investment takes place, and note that σ (as defined in [2.4] above) is its standard deviation parameter:

$$m = -\delta\beta+\vartheta_z-\vartheta_p+\sigma_p^2-\rho\sigma_z\sigma_p$$

$$\sigma = \sqrt{\sigma_z^2 + \sigma_p^2 - 2\rho\sigma_z\sigma_p}$$

The investment policy [R] imposes on $\{\xi_t\}$ an upper control barrier at c^*: capital will be installed as necessary to prevent ξ_t from being larger than c^*. ξ_t is then a geometric Brownian motion with a reflecting barrier at c^*, and applying Ito's lemma it can be shown that $-\ln(\xi_t) + \ln(c^*)$ is a linear Brownian motion process with a control barrier at zero, drift $-(m-\sigma^2/2)$ and standard deviation σ; the ergodic ditribution for such a process is known[10] to exist as long as the drift is negative, i.e. (from the definitions above for m and σ) as long as

$$[3.2] \quad -\delta\beta+\vartheta_z-(\sigma_z^2/2) > \vartheta_p-(\sigma_p^2/2)$$

This requires both that $\{\xi_t\}$ have a tendency to drift (upwards) towards the investment point, and that there not be too much "noise" in the model: σ_z^2 and σ_p^2 should be reasonably small compared to the drift parameters δ, ϑ_z and ϑ_p. If [3.2] is not satisfied, the density of $\{-\ln(\xi_t) + \ln(c^*)\}$ goes to zero everywhere in $[0,\infty)$, implying that the density of ξ degenerates to a spike arbitrarily close to zero (the log function has infinite slope at zero, but a geometric Brownian motion process can never

[10] see for example Cox and Miller[1965] page 225.

reach zero from positive values): ξ converges to zero in probability for all initial conditions. Some intuition about this degeneracy can be obtained considering the certainty case: if $\sigma_z^2 = \sigma_p^2 = \sigma^2 = 0$ and [3.2] is not satisfied, the firm would never undertake a dynamic investment strategy, but would limit itself to a once-and-for-all acquisition of capital when it is set up (compare [3.2] with [2.11]): then the ratio of capital's marginal profitability to its purchase price would certainly converge to zero as $t \to \infty$. In the presence of uncertainty, the firm would invest not only at the beginning of time but also at other points in time, when business conditions and/or the price of capital are favorable enough to obtain $\xi_t = c^*$, even though [3.2] is not true; but in the limit the probability of observing $\xi > 0$ goes to zero all the same, because good business conditions and/or low price of capital are so very unlikely if [3.2] is not true.

If [3.2] is satisfied, the ergodic distribution of the $\{\xi_t\}$ process is well defined and is exponential (see Cox and Miller[1965], p.225):

$$\text{Prob}\left[-\ln(\xi) + \ln(c^*) \le x\right] = 1 - 1(x \ge 0)\, e^{\frac{2(m - s^2/2)}{s^2} x}$$

It is then a simple matter to invert the monotonic function $f(\xi) = -\ln(\xi_t) + \ln(c^*)$ and find the ergodic cumulative distribution function of ξ:

$$[3.3] \quad \text{Prob}\left[\xi \le x\right] = \left[\frac{x}{c^*}\right]^{2\frac{m}{\sigma^2} - 1} 1\left[0 \le x \le c^*\right]$$

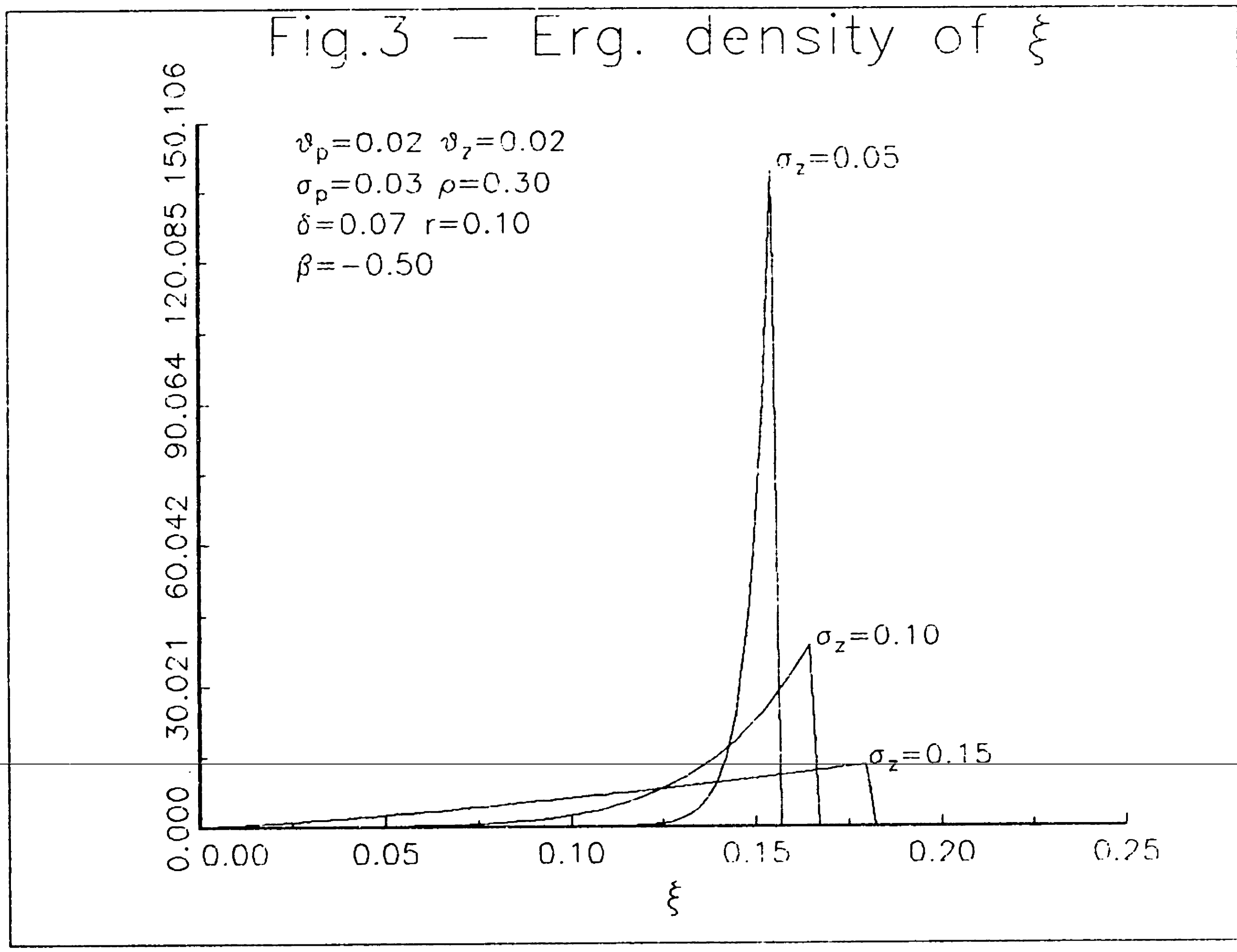


Fig.3 – Erg. density of ξ
ϑp=0.02 ϑz=0.02
σp=0.03 ρ=0.30
δ=0.07 r=0.10
β=−0.50
σz=0.05
σz=0.10
σz=0.15
0.000
30.021
60.042
90.064
120.085
150.106
0.00
0.05
0.10
0.15
0.20
0.25
ξ

The steady-state density of ξ is plotted in Figure 3 for different levels of undertainty. Naturally, if there were no uncertainty this density would degenerate to a spike located at $(r-\vartheta_p+\delta)$; as uncertainty becomes more important, and the upper limit of the distribution shifts to the right as c^* increases, more and more probability density is located at low values of ξ.

Simple integration shows that, in the ergodic steady state, the mean of ξ is

$$[3.4] \quad \bar{\xi} = \left[\frac{m - \sigma^2/2}{m}\right] c^* = \frac{-\delta\beta+\vartheta_z-(\sigma_z^2/2) - \vartheta_p+(\sigma_p^2/2)}{-\delta\beta+\vartheta_z-\vartheta_p+\sigma_p^2-\rho\sigma_z\sigma_p} c^*$$

Of course, $0<\bar{\xi}<c^*$ (as long as [3.2] is satisfied and the expectation is well defined), since $0<\xi_t\leq c^*$ for all t.

If [3.2] is satisfied, this expression can be shown to be equal to $r+\delta-\vartheta_p$ when there is no uncertainty ($\sigma_z^2 = \sigma_p^2 = 0$), and to be strictly <u>less</u> than that when there is uncertainty and the irreversibility constraint is sometimes binding.

Figures 4a and 4b plot the ergodic mean of ξ as a function of σ_z and σ_p, for several values of the technology and demand parameter β. It is apparent from the figures that when the parameters are such as to make the firm more reluctant to invest ex-ante, they are also such that the firm will ex-post be using <u>more</u> capital compared to the reversible investment case.

The presence of uncertainty, while making the firm more reluctant ex-ante to undertake irreversible investment, also makes adverse realizations of business conditions or decreases in the

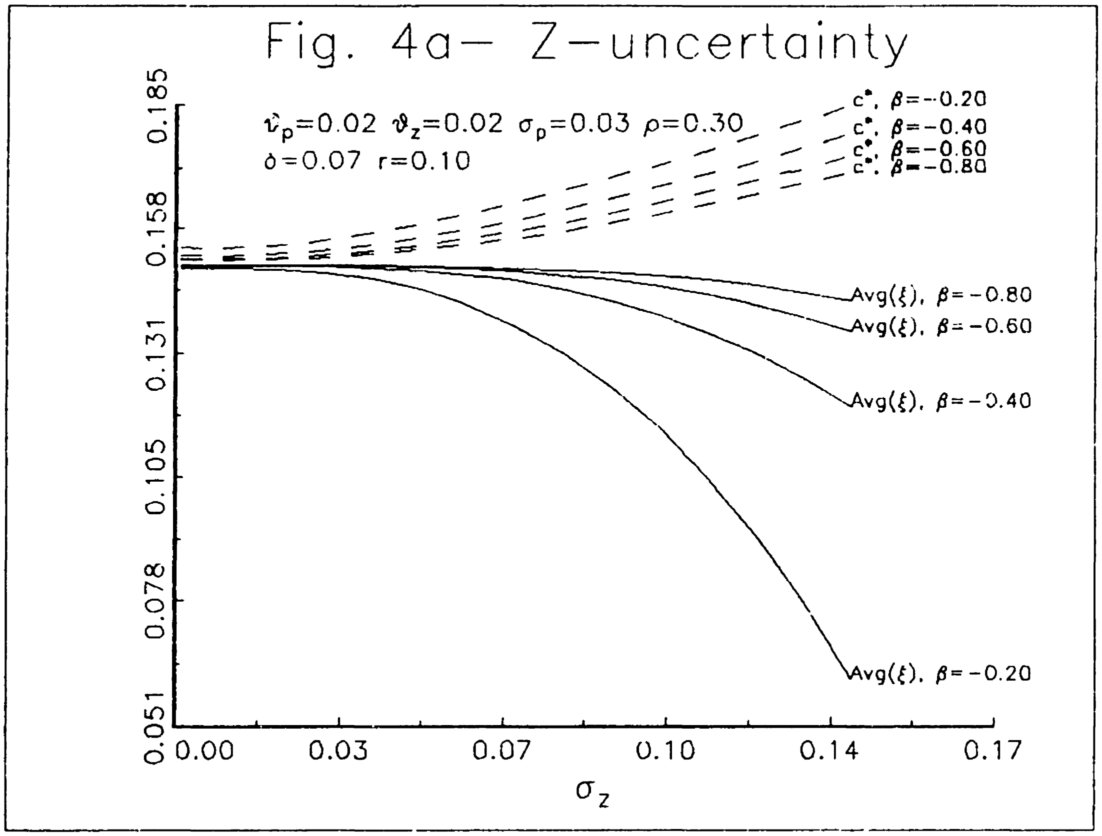


Fig. 4a– Z-uncertainty
ϑ_p=0.02 ϑ_z=0.02 σ_p=0.03 ρ=0.30
δ=0.07 r=0.10
c*, β=−0.20
c*, β=−0.40
c*, β=−0.60
c*, β=−0.80
Avg(ξ), β=−0.80
Avg(ξ), β=−0.60
Avg(ξ), β=−0.40
Avg(ξ), β=−0.20
0.051
0.078
0.105
0.131
0.158
0.185
0.00
0.03
0.07
0.10
0.14
0.17
σ_z

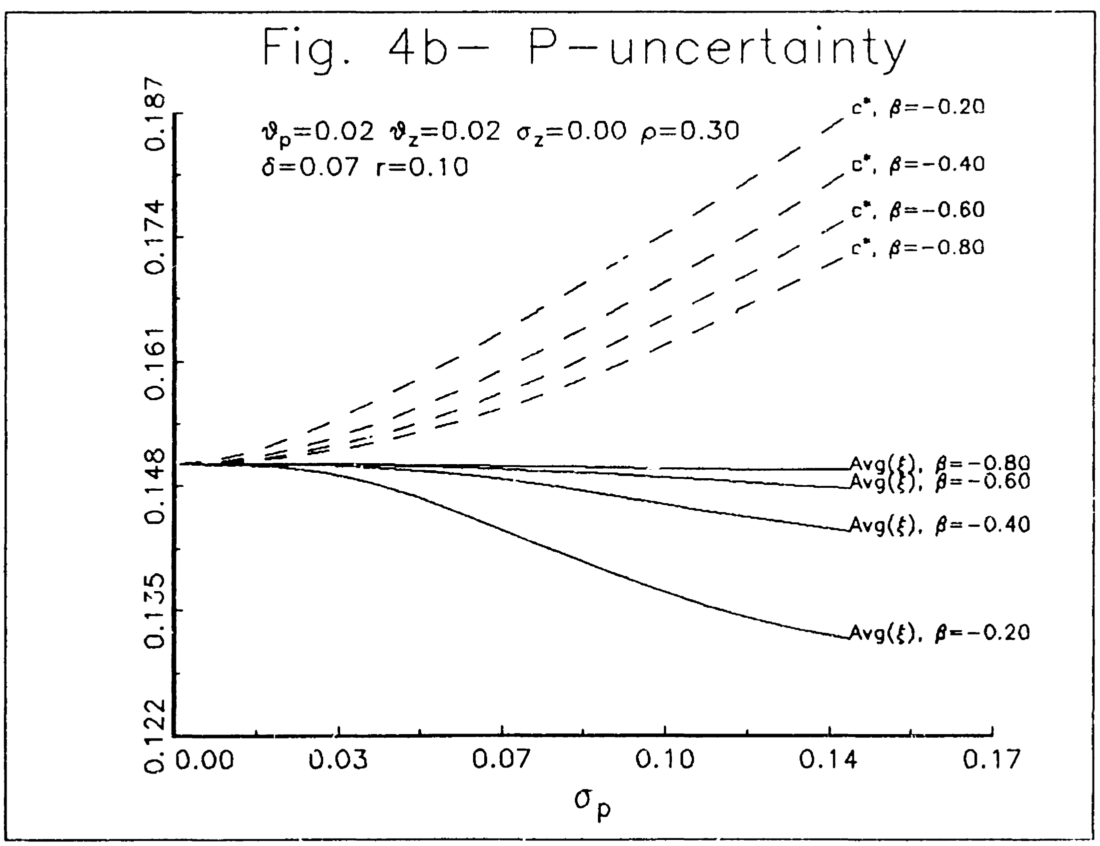


Fig. 4b– P-uncertainty
ϑ_p=0.02 ϑ_z=0.02 σ_z=0.00 ρ=0.30
δ=0.07 r=0.10
c*, β=−0.20
c*, β=−0.40
c*, β=−0.60
c*, β=−0.80
Avg(ξ), β=−0.80
Avg(ξ), β=−0.60
Avg(ξ), β=−0.40
Avg(ξ), β=−0.20
0.122
0.135
0.148
0.161
0.174
0.187
0.00
0.03
0.07
0.10
0.14
0.17
σ_p

price of capital so likely that, ex-post, irreversible capital accumulation results on average in higher capital intensity of production. The empirical implications of this are discussed in section 5 below.

4 - The value of the firm, average Q and marginal Q

It is possible to compute the value of the firm by integrating the value of all marginal units of capital, both those installed and those yet to be installed, given in [2.9] and [2.10]. If the firm never installed any more capital, each of the currently installed units would still, while progressively depreciating, produce a cash-flow with present expected value as given in [2.9]. Moreover, the firm does hold the option to install further units: an option is always valuable, since it provides its owner with the right, but not the obligation, to acquire an asset. In the framework considered here, the option's value can be shown to equal the expression F(.) in [2.10].

Simple integration then obtains the following expression for the firm's value:

$$[4.1] \quad V^*\left(K_t, Z_t, P_t\right) = \int_0^{K_t} v\left(x, Z_t\right)\, dx + \int_{K_t}^{\infty} F\left(x, Z_t, P_t\right)\, dx =$$

$$= \int_0^{K_t} \frac{x^{\beta}\, Z_t}{r+\delta-(\vartheta_z-\delta\beta)}\, dx + \int_{K_t}^{\infty} \frac{1}{A-1}\left(\frac{x^{\beta}\, Z_t}{c^* P_t}\right)^{A-1} \frac{x^{\beta}\, Z_t}{c^*}\, dx =$$

$$= \frac{1}{1+\beta} \frac{K_t^{1+\beta} Z_t}{r+\delta-(\vartheta_z-\delta\beta)} + \frac{1}{A-1} \frac{1}{-\beta A-1}\left(\frac{K_t^{\beta} Z_t}{c^* P_t}\right)^A P_t K_t$$

Note that $-\beta A>1$ is necessary and sufficient for convergence of the second integral above: this is guaranteed to be true as long as the condition in [2.7] above is satisfied. If $\beta A\leq 1$, the value function fails to exist because the value of the options to install capital in the future does not converge.

Also note that $K_t<K^*$ is never observed, because if that were the case the capital stock would instantaneously be increased to K^*. The value function is defined over the $K<K^*$ region as $V^*\left[K^*(Z_t,P_t),Z_t,P_t\right]-P_t\left[K^*(Z_t,P_t)-K\right]$: if for any reason the firm finds itself with $K<K^*$, (K^*-K) units of capital are immediately purchased and $V^*(K^*,Z,P)$ is obtained. V^* is continuously differentiable, is concave in the relevant region, and its slope equals P_t at or below $K^*(P_t,Z_t)$.

Figure 5a plots the value of the firm against K for several values of σ_z (the values for the other parameters are given in the figure); and Figure 5b performs the same experiment for several values of β.

For given K, Z and P, the firm is more valuable the more volatile the business conditions process is: as Pindyck[1986] notes, when demand is very volatile the options to invest are worth more. Most of the value of firms faced by high uncertainty consists of the opportunity to invest in the future (the "growth

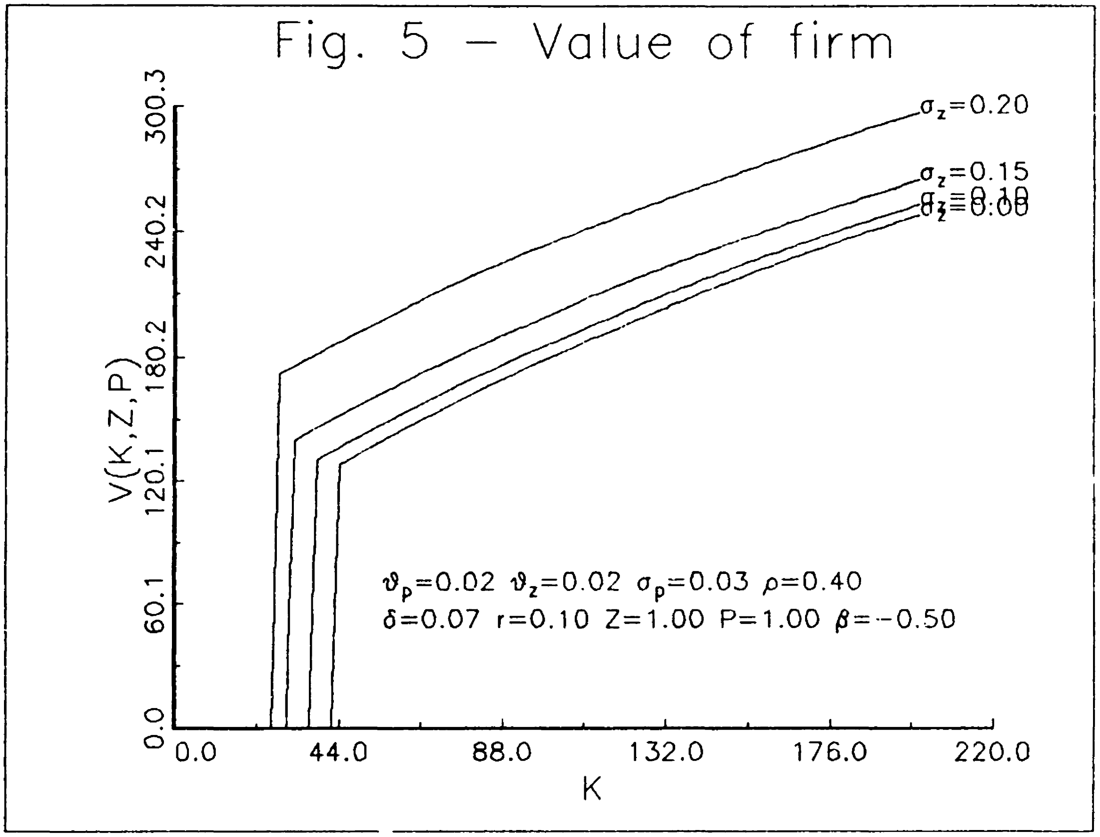


Fig. 5 — Value of firm
V(K,Z,P)
K
0.0
60.1
120.1
180.2
240.2
300.3
0.0
44.0
88.0
132.0
176.0
220.0
σz=0.20
σz=0.15
σz=0.10
σz=0.00
ϑp=0.02 ϑz=0.02 σp=0.03 ρ=0.40
δ=0.07 r=0.10 Z=1.00 P=1.00 β=-0.50

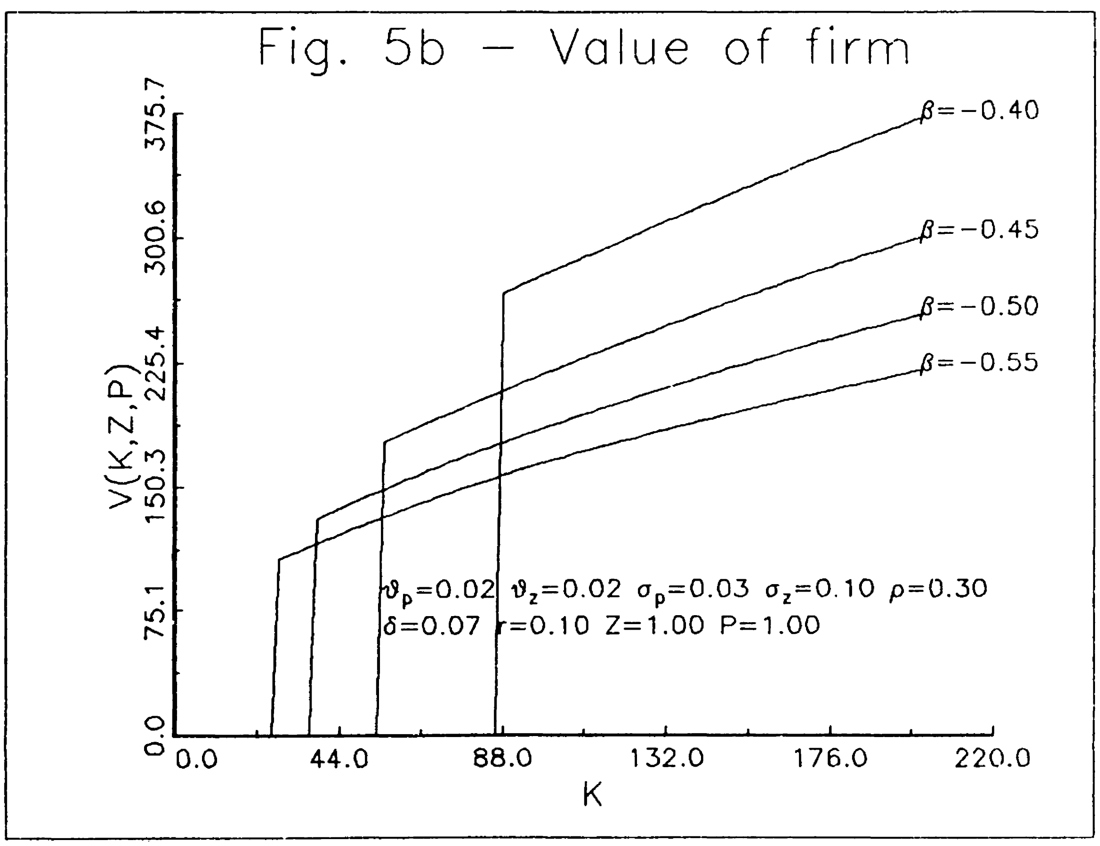


Fig. 5b — Value of firm
V(K,Z,P)
K
0.0
75.1
150.3
225.4
300.6
375.7
0.0
44.0
88.0
132.0
176.0
220.0
β=-0.40
β=-0.45
β=-0.50
β=-0.55
ϑp=0.02 ϑz=0.02 σp=0.03 σz=0.10 ρ=0.30
δ=0.07 r=0.10 Z=1.00 P=1.00

options").[11]

It is also worth noticing that firms whose β has a low absolute value (indicating high monopoly power and/or strongly decreasing returns to scale) have higher value, given K, for the same value of Z and P: such firms earn large monopoly profits or inframarginal rents.

It is easy to derive the dynamic behavior of the firm's value, by Ito's lemma:

$$[4.2]\quad dV^*\left(K_t, Z_t, P_t\right) = \frac{1}{1+\beta}\,\frac{K_t^{1+\beta}\;Z_t}{r+\delta-(\vartheta_z-\delta\beta)}\left[\left(-\delta(1+\beta)+\vartheta_z\right)dt+\sigma_z dW_{zt}\right]$$

$$+\frac{1}{A-1}\,\frac{1}{-\beta A-1}\,\frac{K_t^{\beta A+1}Z_t^A}{c^{*A}\;P_t^{A-1}}\left[\left(-\delta(\beta A+1)+A\vartheta_z+(1-A)\vartheta_p+\frac{1}{2}A(A-1)\sigma^2\right)dt\right.$$

$$\left.+A\sigma_z dW_{zt}+(1-A)\,\sigma_p dW_{pt}\right]$$

Noting that $-\delta(\beta A+1)+A\vartheta_z+(1-A)\vartheta_p+\frac{1}{2}A(A-1)\sigma^2 = r$ from the definition of A, it is possible to verify that

$$E_t dV^*\left(K_t, Z_t, P_t\right) = rV^*\left(K_t, Z_t, P_t\right)-\left[\frac{1}{1+\beta}K_t^{1+\beta}\;Z_t dt - P_t\left(dK_t-\delta K_t dt\right)\right]$$

This is true by construction, since V^* is defined in [2.1] above as the present discounted value of cash flows: the expected

[11] Note that we assume here that the demand function completely describes the market situation of the firm: in particular, if good business conditions would induce other firms to enter the market, then the process for the demand index D_t should take this into account. In practice, it would be difficult to accurately specify such a demand process for any individual firm, because barriers to entry are not observable.

proportional return (cash flow plus capital gains) from holding the firm's stock for one instant is equal to the required rate of return.

However, the fluctuations of the return around its expected values are <u>not</u> normally distributed. In other words, the value of the firm does not follow a geometric Brownian motion, even though the processes that characterize the firm's environment do:

$$dV^*\left[K_t, Z_t, P_t\right] - E_t V^*\left[K_t, Z_t, P_t\right] =$$

$$= \frac{1}{1+\beta}\,\frac{K_t^{1+\beta}\;Z_t}{r+\delta-(\vartheta_z-\delta\beta)}\,\sigma_z dW_{zt} + \frac{1}{A-1}\,\frac{1}{-\beta A-1}\,\frac{K_t^{\beta A+1} Z_t^A}{c^{*A}\;P_t^{A-1}}\left[A\sigma_z dW_{zt} + (1-A)\;\sigma_p dW_{pt}\right]$$

and this expression is not proportional to V^*.

The firm's value fails to have a conditionally lognormal distribution because the cash-flow process follows geometric Brownian motion almost always, but has a singular component at times of positive gross investment. From another point of view, the total firm value is given by the sum of discounted operating profits from currently installed capital and of the "growth options": each of the components has normally distributed returns under the assumptions made above, but the relative importance of the two components varies as the firm finds itself closer or farther from the investment point.

When far from the investment point, the options to invest in the future are less valuable and their weight decreases: since the options are more volatile than the profits from installed capital, the conditional variance of returns <u>decreases</u> after an increase in the price of capital or a deterioration of the business conditions. Since such occurrences also decrease the total value of the firm, it appears that investment irreversibility should imply lower variability of returns after abnormally low realizations of returns: this is not true in the data (see for example Nelson[1987]), and more research will be needed to clarify the relationship between the firms' irreversible investment decisions and the volatility of stock market returns.[12]

The non-normality of returns from holding the firm's stock implies that r, the required rate of return, cannot be computed by a simple application of a Capital Asset Pricing Model. Further research should relate this finding to the Capital Budgeting literature's "project beta" concept, and obtain a relationship between the required rate of return on the sum total of the firm's assets and the rate of return to be used in evaluating the opportunity of undertaking incremental investment in the firm's capital stock.

An important variable in empirical research on investment is the so-called "Marginal Q", defined as the ratio of the shadow price of capital to the market price of uninstalled capital: Abel

---

[12] It will also be necessary to study the implications of the model proposed here in a richer financial environment, allowing for a choice of financing instruments (stocks or bonds). There probably exist dynamic leverage policies that can reconcile the model with evidence on stock values and produce smooth dividend payments. In the model above, firms engaged in rapid investment pay negative dividends (issue shares) at times of positive investment, when $K^{\beta} Z\, dt < dX\, P$.

and Blanchard[1986] argue that not only should investment be related to this quantity, but that marginal Q should be the <u>only</u> determinant of investment decisions.

Under the assumption of investment irreversibility, it is easy to compute marginal Q from the results of the previous sections of this chapter: recalling the expression for the shadow value of capital in [2.8] above, and using the definition of ξ_t in [3.1],

$$[4.3] \qquad Q_m = \frac{K_t^\beta Z_t - \frac{1}{A}\left(\frac{K_t^\beta Z_t}{c^* P_t}\right)^{A-1} K_t^\beta Z_t}{r + \delta + \delta\beta - \vartheta_z} \frac{1}{P_t} =$$

$$= \frac{\xi_t - \frac{1}{A}\left(\frac{\xi_t}{c^*}\right)^{A-1} \xi_t}{r + \delta + \delta\beta - \vartheta_z}$$

It is easy to check that, under the investment rule [R], $Q_m \le 1$ always, and $Q_m=1$ when the firm is investing[13]: moreover, Q_m is monotonically increasing in ξ_t in the relevant range $0<\xi_t\le c^*$, and it is therefore possible to compute the ergodic distribution of Q_m using ξ's distribution derived above.

Unfortunately, the function $Q_m=Q_m(\xi)$ does not have an inverse in closed form, and it is necessary to invert it numerically to find $\xi=Q_m^{-1}(q)$. Differentiating [3.3] above we find the ergodic

---

[13]the same nonlinear relationship between irreversible investment and Q is noted by Sargent[1979b], in his general equilibrium model.

density of ξ:

$$[4.4] \quad f_\xi\left(\xi\right) = \left(2\frac{m}{\sigma^2} - 1\right)\left(\frac{\xi}{c^*}\right)^{2\frac{}{\sigma^2} - 2}\frac{1}{c^*}$$

and it is then possible to compute the ergodic density of Q_m using the relationship:

$$f_{Q_m}\left(q\right) = \left[\frac{\partial Q_m(\xi)}{\partial \xi}\Bigg|_{\xi = Q_m^{-1}(q)}\right]^{-1} f_\xi\left(Q_m^{-1}(q)\right)$$

As long as condition [3.2] holds true, implying that investment occurs "often" and ergodic distributions exist, the ergodic density of marginal q has a sharp spike at 1: but lower values of Q_m can be observed with positive probability, when firms face very unfavorable business conditions or the price of capital is very high.

Once numerical values are found for Q_m's density, an approximate ergodic mean can be computed by summation rather than integration; Table 1 reports the results.

The mean of Q_m is strictly less than one, and is lower for smaller $|\beta|$ [14] and larger σ. When $\sigma=0$ and [3.2] is true, marginal Q is identically equal to one, since the firm is always investing; if uncertainty is large, lower values will be more often observed.

---

[14] a small $|\beta|$ implies that the shadow value of capital is only very slowly decreased by depreciation, and investment is less likely to occur.

TABLE 1 - ERGODIC MEAN OF MARGINAL Q

	$\sigma_z=0.05$	$\sigma_z=0.10$	$\sigma_z=0.13$
$\beta=-0.25$	0.898	0.734	0.568
$\beta=-0.50$	0.929	0.861	0.784
$\beta=-0.75$	0.938	0.900	0.856

Note to the table: mean of the ergodic distribution, obtained by numerical approximation of the density; $Q_m(\xi)$ was numerically inverted at a grid of ξ points, 0.005 apart. Parameter values: P=1, Z=1, $\vartheta_P=0.02$, $\sigma_P=0.03$, r=.15, $\delta=.07$, $\vartheta_z=.02$, $\rho=.3$.

The formula for V^* could in principle be used to correct the specification of empirical investment equations that use stock market data to compute "average q". Hayashi[1986] shows in a convex-costs-of-adjustment model that average q is the correct independent variable in an investment equation only under perfect competition and constant returns to scale. Here it is necessary to violate at least one of these conditions to obtain the investment rule, but the results still provide useful insights. Average Q is the ratio of the market value of the firm, V^*, to the replacement cost of the currently installed capital stock: this can easily computed, for the model proposed here, as

$$Q_A = \frac{1}{1+\beta}\,\frac{\dfrac{K_t^{\beta}\,Z_t}{P_t}}{r+\delta-(\vartheta_z-\delta\beta)} + \frac{1}{A-1}\,\frac{1}{-\beta A-1}\left(\frac{K_t^{\beta}\,Z_t}{c^*P_t}\right)^A$$

$$= \xi_t \left[\frac{1}{(1+\beta)(r+\delta-(\vartheta_z-\delta\beta))} + \frac{1}{A-1}\,\frac{1}{-\beta A-1}\left(\frac{\xi_t}{c^*}\right)^{A-1}\frac{1}{c^*} \right]$$

Since $\xi_t \leq c^*$ under [R], and $A-1>0$, there is an upper bound to average Q:

$$\mathrm{Max}(Q_A) = c^* \left[\frac{1}{(1+\beta)(r+\delta-(\vartheta_z-\delta\beta))} + \frac{1}{A-1}\,\frac{1}{-\beta A-1}\,\frac{1}{c^*} \right] =$$

$$= \frac{A}{A-1} + \frac{1}{A-1}\,\frac{1}{-\beta A-1} = \frac{A(-\beta A-1)+1}{(A-1)(-\beta A-1)}$$

Positive investment is only observed when average Q attains its maximum value. There is however no lower bound on Q_A, as there is no lower bound on Q_m: a firm can be <u>very</u> unlucky, and experience such a bad drop in Z or such a large increase in P that as to bring average and marginal Q arbitrarily close to zero.

Since $A>1$ and $-\beta A>1$, average Q is a monotonic function of ξ_t and therefore, like marginal Q, possesses an ergodic distribution as long as ξ_t does. The same numerical procedure that was used for marginal Q can be used to compute average Q's ergodic distribution and ergodic mean: the results are reported in Figures 6a and 6b and in Table 2.

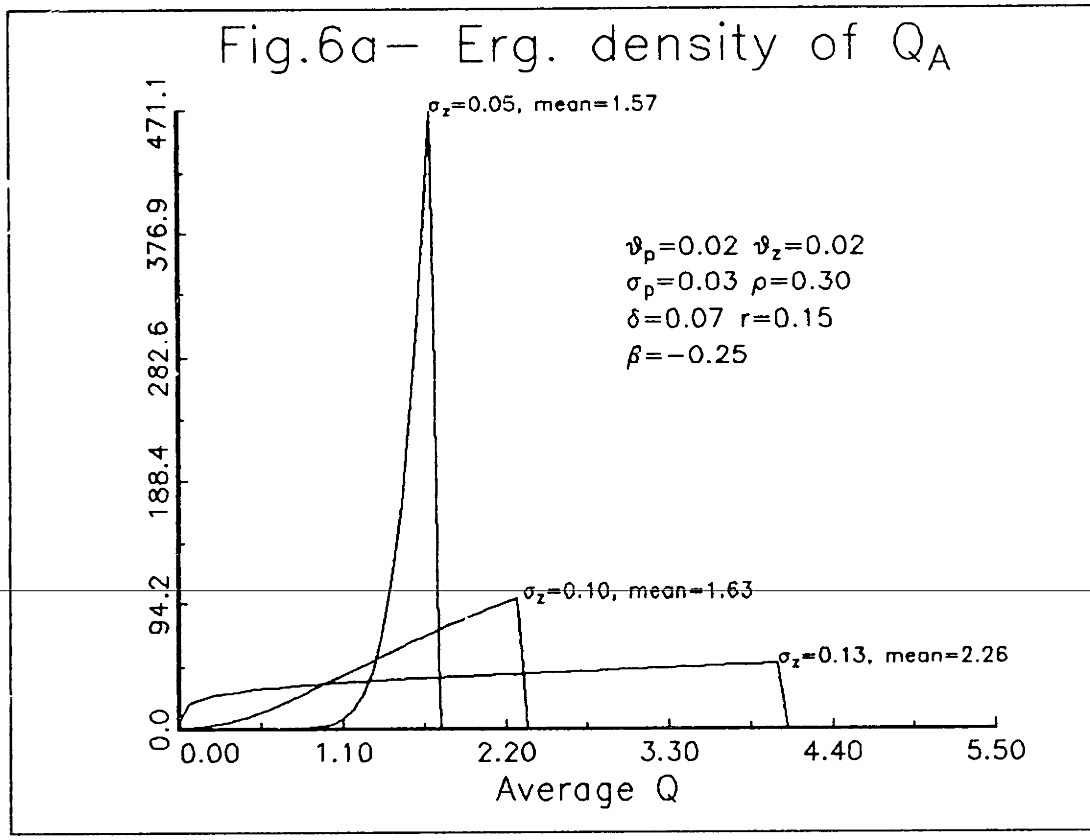


Fig.6a– Erg. density of Q_A
σ_z=0.05, mean=1.57
σ_z=0.10, mean=1.63
σ_z=0.13, mean=2.26
ϑ_p=0.02 ϑ_z=0.02
σ_p=0.03 ρ=0.30
δ=0.07 r=0.15
β=−0.25
0.0
94.2
188.4
282.6
376.9
471.1
0.00
1.10
2.20
3.30
4.40
5.50
Average Q

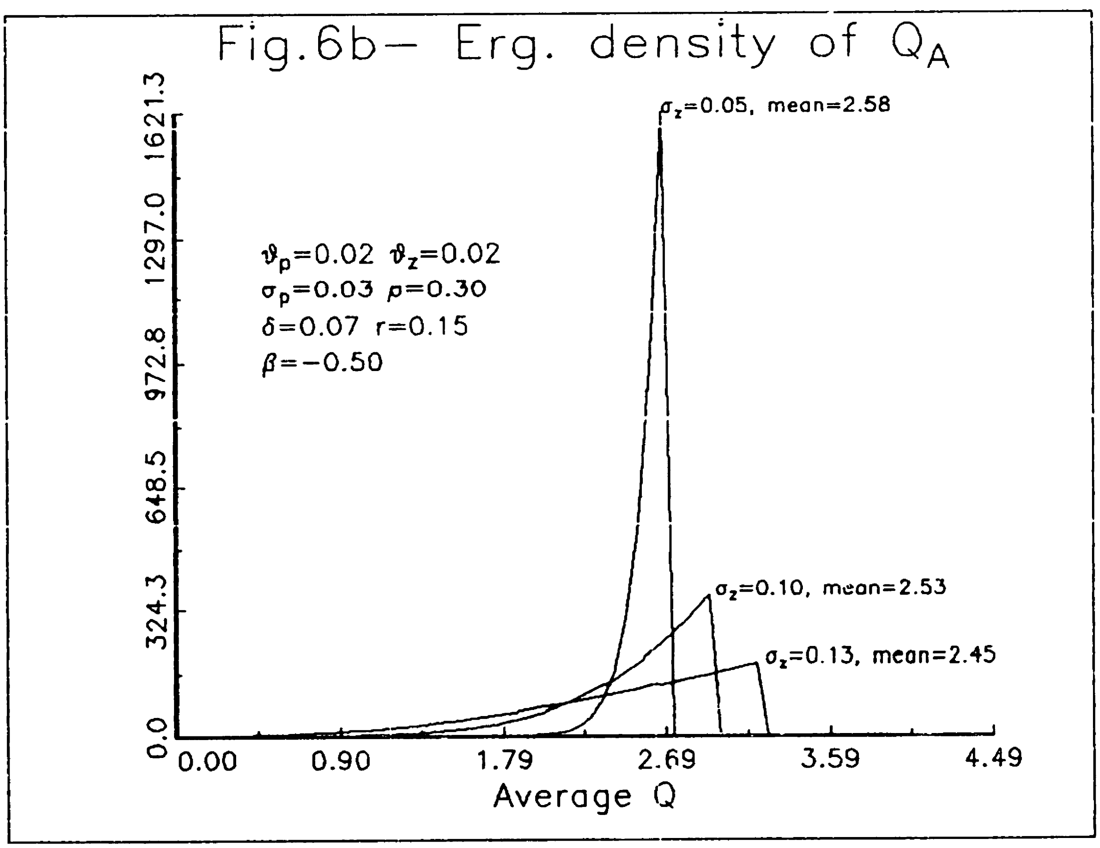


Fig.6b– Erg. density of Q_A
σ_z=0.05, mean=2.58
σ_z=0.10, mean=2.53
σ_z=0.13, mean=2.45
ϑ_p=0.02 ϑ_z=0.02
σ_p=0.03 ρ=0.30
δ=0.07 r=0.15
β=−0.50
0.0
324.3
648.5
972.8
1297.0
1621.3
0.00
0.90
1.79
2.69
3.59
4.49
Average Q

TABLE 2 - ERGODIC MEAN OF AVERAGE Q

	σ_z=0.05	σ_z=0.10	σ_z=0.13
β=-0.25	1.506	1.625	2.244
β=-0.50	2.538	2.520	2.436
β=-0.75	5.600	5.535	5.447

Note to the table: mean of the ergodic distribution, obtained by numerical approximation; $Q_A(\xi)$ was numerically inverted at a grid of ξ points, 0.005 apart. The averages reported in Figure 6 do not correspond exactly to those in Table 2 because a more widely spaced grid was used in constructing the figure. Parameter values: P=1, Z=1, ϑ_p=0.02, σ_p=0.03, r=.15, δ=.07, ϑ_z=.02, ρ=.3.

Average Q is very large if β is large in absolute value, as implied by high monopoly power or strongly decreasing returns to scale: Hayashi[1986] shows that average Q should fluctuate around one for perfectly competitive firms operating under constant returns to scale and convex costs of capital adjustment, while monopoly power and decreasing returns would produce higher values of average Q. Empirically, average Q is not very high (see for example Hoshi and Kashyap[1987]), which would seem to rule out values of $|\beta|$ much larger than .5 for the model proposed here if the other parameters have the values reported in the note to the figure.

The dispersion of Q_A's ergodic distribution is larger for large σ_z (see Figure 6), but the value of its ergodic mean is not very sensitive to the degree of uncertainty facing the firm[15]. This is not surprising, however, since it has been shown above that more uncertainty causes more capital to be installed in the ergodic steady state, increasing the denominator of Q_A, but also implies that the value of the firm is higher for any installed capital stock, which increases the numerator of Q_A. The net effect of higher uncertainty appears very small.

5 - Implications of investment irreversibility for empirical work

The basic irreversibility insight should prove useful in empirical work on investment. Naturally, the simple model proposed in Section 1 should be made more realistic for empirical work, including for example inventory fluctuations, time-to-build[16] (see Majd and Pindyck[1987]), a downward-sloping supply function for capital goods, and so on.

This section discusses the relevance of the results reported

---

[15] at least, no systematic relationship is found by the computations reported in Table 2; it might be the case that the numerical procedure used is not precise enough to uncover a (shallow) relationship.

[16] It is straightforward to include _fixed_ delivery lags in the model proposed above: if the firm knows with certainty that orders will be filled, say, in a year, then the optimal stopping problem for the installation of the marginal unit of capital is easy to solve since the value of _orders_ can be computed by modifying $v(K_t, Z_t)$ to be the expectation of discounted marginal profits from the time of delivery on. Delivery lags make the firm _less_ reluctant to invest, if capital can be paid at time of delivery: both good and bad developments can occur before the time of delivery.

above for the interpretation of empirical work on investment[17] based on more traditional theoretical models.

It is not easy to evaluate the quantitative relevance of capital accumulation irreversibility: the constraint is more important the more uncertain is the firm's environment, the lower is the expected rate of growth of the economy, the lower is the depreciation rate, the steeper is the marginal profitability of capital schedule. The more binding is the irreversibility constraint, the less attractive is capital accumulation ex-ante (firms try to use less capital under severe uncertainty) but, conversely, the more capital intensive is production ex-post.

In aggregate data, gross investment is positive in all periods, all countries, all sectors: on the one hand, this is evidence in favor of investment irreversibility, because if capital accumulation were in fact reversible negative gross investment would sometimes be observed. But on the other hand, if gross investment were in fact continuously positive then the irreversibility constraint, though present, would never be binding and would be completely irrelevant to the empirical study of investment. Aggregate uncertainty is, in fact, sufficiently low that desired negative aggregate investment is very unlikely for realistic depreciation rates.

[17] The potential importance of investment irreversibility for the interpretation of all economic variables should not be neglected: adverse shocks will cause the capital stock to be (ex-post) too large; if flexible factor demands and prices are correlated to the exogenous processes (wages, demand...), a high degree of permanence in the effects of positive and negative shocks will be found. Moreover, the response to positive and negative shocks will be asymmetric if the former induce the firm to invest.

Of course, even though gross investment is never observed to be zero, it is conceivable that decreasing the aggregate capital stock could sometimes be desirable; the mere possibility of a very bad drop in (say) demand or prices should reduce the desired capital stock (increase the marginal productivity of capital required for additions to the existing capital stock to be desirable) even in periods of positive gross investment, while the installed capital stock would ex-post turn out to be excessive after the realization of the bad shock. These effects could in principle be large, as shown in Figure 4.

More importantly, it is likely that the strict positivity of gross investment at the aggregate level masks binding irreversibility constraints at the level of individual firms, or maybe individual capital goods. After all, it is self-evident that not all firms invest in all types of capital goods, at all times, in all locations[18]: steel mills in Pittsburgh were left unutilized when the U.S. steel industry underwent the 70s crisis; and any firm exposed to international competition must, during the first half of the 80s, have regretted investment decisions made in the late 70s. The variability of individual stock prices is 3 or 4 times larger than the variability of aggregate stock price indexes; and, as documented by Romer[1987], production and sales are several times more variable at the industry level than at the level of the whole manufacturing sector - suggesting that

[18] The same point was made by the discussants of Hall[1977].

uncertainty is even larger at the firm level. Given large disaggregate uncertainty, the irreversibility constraint will be more binding the more specific to a firm's needs capital goods are: trucks can obviously be sold to luckier firms when a negative idiosyncratic shock hits, but machine tools and plants generally have very little value unless employed for their original intended use. This obviously introduces difficult cross-section and time aggregation problems in the study of aggregate investment, which are best left to future research.

It should be easier to apply irreversible investment theory at a more disaggregate level: but to accurately gauge the parameters in the individual firm's problem it would necessary to obtain estimates of the elasticity of the reduced-form profit function to capital (which depends on the elasticity of demand as well as on the substitutability of more flexible factors to capital), and of the expected rates of growth and variances of demand, wages and capital prices. This is clearly a formidable task: identification of demand disturbances requires a specification of the market structure, while all the endogenous variables are jointly determined as functions of all exogenous variables; even in the simplest model, the relationship between the variables is highly nonlinear, and sophisticated identification assumptions would be needed to obtain estimates.

In particular, it should be noted that the characteristics of the firm's production and demand functions are important in determining the weights given to the exogenous stochastic processes' parameters when deriving the parameters of the stochastic process followed by the business conditions index Z_t: for example, the rate of wage inflation and its variability are less important if $(1-\alpha)$, the share of labor in the Cobb-Douglas production function, is small; and if $\phi\mu$ is close to zero (the firm has a lot of monopoly power, or produces under strongly decreasing returns to scale) the weight given to demand uncertainty is large. The "degree of uncertainty" facing the firm depends on the specification of technology and demand.

Leaving the aggregation problems to future research, it is still possible to comment on the results obtained by empirical research on aggregate and firm data: researchers often relate investment in a period to average Q (see Hayashi[1982] and his references) or to marginal Q (see Abel and Blanchard[1986]); other researchers relate the change in the capital stock to the marginal profitability of installed capital and to the user cost of capital (see Jorgenson[1963] and Hall[1986]).

If in fact investment is irreversible at a more disaggregate level, and idiosyncratic shocks are large compared to aggregate ones (or, equivalently, there is low correlation among the evolution of business conditions for different firms), empirical research should be observing the average of distributions similar to the ones derived in Sections 3 and 4 above, and depicted in Figures 6 and 7. The distributions derived above are the steady-state ones: in every period and sector the <u>spread</u> of average and marginal Q for individual firms and individual capital goods will be determined by the history of idiosyncratic shocks, while the location of the distribution will be determined by the history of aggregate shocks.

To fix ideas, think of a researcher trying to explain investment in a sector by the aggregate value (for the sector) of observables such as ξ (the ratio of marginal profitability of capital to its purchase price), Q_m (marginal Q, the ratio of the shadow price of capital to its purchase cost) or Q_A (average Q, the ratio of firm's total value to the replacement cost of their installed capital stock). News affecting the marginal productivity of capital and/or its price for the whole aggregate will shift the distribution of ξ_t: all the individuals that are then brought against the investment barrier c^* will be prevented from crossing it by a spurt of investment, and the mean of the distribution of ξ_t will increase. Idiosyncratic shocks will, on the other hand, simply "stir" or "mix" the distribution of ξ_t, without affecting its location: as individual units get pushed against the investment barrier c^*, noisy movements will be generated in the investment series, movements completely unrelated to aggregate observables.

It has been shown above that both Q_A and Q_m are monotonically related to ξ_t if investment is irreversible: then aggregate (over time, firms and capital goods) Q_A and Q_m should be positively related to the amount of aggregate investment in any given period.

In practice, researchers (see for example Hoshi and Kashyap[1987] for average Q; and Abel and Blanchard[1986] for marginal Q[19]) do find that Q has a fair amount of explanatory power when investment is regressed on it, but observe that much of investment's variability is left unexplained, that variables other than Q enter significantly in the regression, and that the implied speed of adjustment to shocks does not make much sense when interpreted in the framework of the convex-cost-of-adjustment theoretical models that underly this empirical literature.

Even admitting that costs of adjustment are in fact convex, there are of course many explanations for the shortcomings of Q's explanatory power. But it should be noted that investment irreversibility is consistent with these findings: while in the aggregate average and marginal Q should be positively related to investment, there is no presumption that the functional form of the relationship should be linear or loglinear, nor that no other variables should be significant in a linear regression. Operating cash flow, production and sales are also related to investment (for a given price of capital), since they are positively related to the marginal profitability of installed capital. Moreover, any variable related to the idiosyncratic shocks affecting individual firms and regions could well turn out significant in a regression.

[19] Abel and Blanchard estimate marginal Q via a vector autoregression, under restrictive assumptions as to the functional form of the profit function: without a model of aggregation, it is not possible to tell whether this is appropriate under investment irreversibility. The strong nonlinearity of the irreversible investment model would suggest caution.

Other researchers (see Hall[1977] for an excellent exposition) have bypassed the convex-cost-of-adjustment assumption and, noting that gross investment is never zero in U.S. aggregate industry data, have obtained the implication that (since condition [2.2] above applies at all times) the optimal investment policy should simply sets the marginal productivity of capital equal to the Jorgenson[1963] user cost of capital. The implied control policy is not dynamic; such an investment rule has little hope of fitting the data, unless supplemented by a (largely unexplained) lag structure in the relationship between the user cost of capital and investment: such a lag structure could, of course, inadvertently fit the complex dynamics implied by investment irreversibility.

Hall[1986] tests empirically the equality of marginal productivity and rental cost of capital, looking at long-run averages of the two to eliminate the dynamic issues. He finds that, in the industries he considers, the former is significantly lower than the latter, and interprets this finding as evidence of purposeful overinvestment, possibly as an entry deterrent on the part of incumbent firms, or as evidence of increasing returns. Of course, both of these explanations may be true in reality; but the optimality condition tested by Hall does not hold if capital accumulation is irreversible at the firm level: as shown above, investment irreversibility implies that in the long run capital's marginal profitability <u>should</u> be lower than the conventionally measured user cost of capital, even under constant returns to scale, and this may well explain Hall's empirical finding.

6 - Concluding comments

Sections 1 and 2 of this Chapter propose a solution to the problem of irreversible sequential investment under uncertainty; choosing Cobb-Douglas technology and constant elasticity demand, it is possible to solve for the firm's investment rule in closed form as special case of the problem solved in Chapter 1.

The rule implies that, under uncertainty, the marginal productivity of capital that triggeres investment is higher than the conventionally measured cost of capital, because of investment irreversibility, even though the firm's owners are assumed to be risk neutral. As noted by Pindyck[1986], there is informal evidence that managers often discount the expected revenues from an investment project at a rate far higher than the one implied by any reasonable risk premium: the model considered here shows that, under certain conditions, this may indeed be very close to the optimal investment rule.

But the positive implications of investment irreversibility are very different from its normative implications: although ex-ante a higher marginal profitability of capital is needed to trigger investment, suggesting that investment irreversibility would make production less capital intensive, Section 3 finds that ex-post the average marginal productivity of capital is lower if investment is irreversible.

Section 4 derives expressions for observables variables that are empirically relevant for the study of the firm; and Section 5 offers preliminary considerations about the empirical relevance of the results.

It seems likely that most investment projects are irreversible to a large extent, and that idiosyncratic uncertainty is large enough to make the irreversibility constraint important; but to apply these observations and the model proposed in this chapter to the study of aggregate investment behavior it will be necessary to solve complex aggregation and estimation problems, and to devise realistic and tractable assumptions about the degree of flexibility in the use of installed capital and about used capital markets.

Apart from the empirical study of investment, irreversibility should interest macroeconomists by its interesting implications for the dynamic behavior of prices, production and employment across "business cycles". Of course, the stochastic process assumed above corresponds to a random walk with drift, not to a stationary process around a deterministic trend, and the "cycles" would have to be redefined accordingly: in the model above investment occurs repeatedly, possibly generating a fairly regular cycle.

CHAPTER 3

FIRING COSTS AND LABOR DEMAND

(Joint with Samuel Bentolila)

1- Introduction

High unemployment rates have become the major problem for most European countries in the 1980s. Several factors are behind the large increases in unemployment after 1973.

The leading cause of the increase in unemployment after the first oil price shock was probably the rise in real wages to levels well above those compatible with full employment. Contractionary demand policies followed by inflation-fighting, budget-balancing governments surely bear the main responsibility for the further rise of unemployment in the 1980s.

Nevertheless, certain aspects of the European unemployment experience suggest that a lack of labor market flexibility may also have contributed to the worsening of the problem. In particular, it is frequently argued that the adjustment to market forces is inhibited by a number of features of the welfare state: generous unemployment benefits, restrictions on hiring and firing, restrictions on wage competition, etc. Pervasive state intervention is thought to have led to rigidified, "Eurosclerotic" -as dubbed by Giersch[1985]- economies, which could not cope with the big shocks of the 1970s.

Flexibility (or the lack of it) is an ambiguous concept. We can distinguish between price and quantity rigidity. The first refers to the unresponsiveness of the wage level and the sectoral wage structure to labor market disequilibria. The second refers to the lack of labor mobility, the existence of restrictions on hiring and firing by firms and of regulations on the arrangements for the utilization of labor (such as rules on the number of hours per worker or the length of labor contracts). This Chapter focusses on the second type of rigidity.

Blanchard *et al.*[1986] argue that in the high and stable growth scenario of the 1960s firms could costlessly consent to tenure and severance pay demands by workers, since employment growth assured that excess hiring was a mistake of at most a few months. The change to an environment of low and volatile demand growth after 1973 then made the severance payments and tenure arrangements set up in the 1960s very costly for firms. As Dornbusch[1986] stresses, now taking on a worker is making a near-irreversible investment: dismissing an employee is very costly and workers seldom quit because of the slim chances of finding another job. Consequently, firms have become much more reluctant to hire, for fear of high firing costs in downturns. From this argument these authors and others have derived the policy recommendation that European labor markets should be made more flexible.

This Chapter studies the effects of hiring and, especially, firing costs on labor demand, with an application to four European countries. It makes two basic contributions. First, a simple continuous-time model for a firm's labor demand decision in the presence of adjustment costs and demand uncertainty is provided. It includes many of the variables that are often mentioned as being relevant to the European case. Then, based on the implications of the model and on results from solving the model for realistic parameter values, we argue that changes in firing costs should not be expected to have large effects on hiring decisions,nor on the average level of employment; rather, the dynamic response of employment to exogenous shocks should be strongly affected by the presence and size of firing costs.

The applicability of the model presented below is limited: it does not deal with unemployment, but only with employment; it is just partial equilibrium and it does not consider labor supply, only labor demand at the level of the firm. In consequence, statements about aggregate magnitudes cannot be made without heroic aggregation assumptions. Nevertheless, we argue below that the application of this model to the European unemployment problem may still be very relevant.

A final caveat is that we do not consider the rationale for the presence of firing costs or the larger set of labor market institutions of which they are a part. In this sense, the restrictions on dismissals may appear here to be "artificial" (in Piore's[1986] words). The existence of firing costs presumably reflects the value attached to employment security, and therefore any comparison with a world with no firing costs may not be very compelling.

The Chapter is structured as follows. The next section reviews the nature and evolution of firing costs in Europe in the recent past, discusses why firing costs may be relevant in explaining the European employment performance, and briefly surveys previous literature on the subject. Section 3 introduces the model and discusses its implications, both in terms of the marginal propensities to hire and fire, and of average labor demand. In section 4 we attempt to interpret the European experience by solving the model for realistic parameter values and by looking at the effect of firing costs on the dynamics of labor demand and on its long run average. Section 5 concludes.

2- Facts and Motivation

a) Firing costs in Europe:

The main component of firing costs in many European countries is the legal regulation of dismissals. Individual dismissal legislation protects workers from being "unfairly" fired. The underlying notion is that the employment relationship is permanent and so a dismissal is fair only if caused by the employee's gross misconduct or lack of qualifications or by economic reasons (redundancy). While the first category allows for summary dismissal without compensation, the latter involve the following procedures: prenotification to workers, their representatives and government agencies, consultation with workers' representatives, rights of appeal of the employer's decision to labor courts, and severance payments. All these are avoided if the worker quits, except in Italy, where severance payments are paid regardless of whether the worker is fired or quits.[1]

The period of notice usually increases, among other things, with the length of service, which also determines how many months' wages is the severance payment. If the dismissal is appealed to the courts and declared unfair (events which are not very likely, see Appendix 3) the payment is significantly higher.

Whenever a certain minimum number of workers is dismissed within a certain period of time, collective dismissal legislation applies, which basically lengthens notice and consultation periods.

The burden of these regulations varies among countries.[2] The ordering from less to more restrictive in the countries we focus on is: United Kingdom, Germany, France and Italy.[3]

Normally very small firms (1 to 19 employees) are exempted from the employment security laws.[4] Employment in those firms in France in 1985 was 25.8% of total employment (up from 23.3% in

---

[1]We neglect this issue in what follows, and therefore probably overestimate the pecuniary cost of firing in Italy in section 4. We feel justified in doing so since non-monetary firing costs have been very large in Italy, due to union militancy and social custom.

[2]For a summary see Piore[1986] or Emerson[1987].

[3]This ranking is confirmed by a 1985 employer survey by the E.E.C. (European Economy[1986]) and by data in Lazear[1987]. The Italian economy, however, may not be so inflexible, given that State financed temporary layoffs ("Cassa Integrazione Guadagni") are possible and given the relative importance of the informal sector.

[4]In Italy, the threshold is 35 workers.

1979) and 17.2% in Germany[5] in 1970 (Sengenberger and Loveman[1987]); thus the regulations cover firms representing around 75% of total employment. The actual coverage is, however, somewhat smaller because newer, part-time and fixed-term contract employees are also excluded from this legislation.

As to the evolution over time, employment security provisions in the law and in collective agreements were introduced in the late 1960s, mainly induced by the social unrest in those years. They were then strengthened around 1975 to protect workers against the income loss caused by unemployment. Unions also showed a strong opposition to mass firings. Finally, in the 1980s, new laws -prompted by the concern with labor market inflexibility- have tried lowered firing costs and allowed for more unstable forms of employment (see OECD[1986b]), while the attitude of unions towards dismissals has also eased.

b) The relevance of firing costs:

In order to understand why firing costs can be relevant for the explanation of the European unemployment experience, it is necessary to quickly review a few stylized facts. We provide some data in Table 1.

[5] 1 to 10 employees.

Table 1: Labor Market Indicators

	FRANCE	GERMANY	ITALY	U.K.
1. Unemployment: 1973	2.6	0.8	6.2	3.0
1986	10.3	6.9	10.5[a]	11.5
2. Change in part-time employment 1973-83	3.3	2.1	-2.0	1.7
3. Labor turnover:				
Accessions: 1973	22.0[b]	34.0	33.0	32.0
1984	13.0	25.0[c]	8.0[c]	19.0
Separations: 1973	19.0[b]	33.0	26.0	31.0
1984	14.0	25.0[c]	14.0[c]	21.0
4. Long term unempl.				
1979	30.3	19.9	35.8	24.8
1985	46.8	31.0	47.9[d]	41.0
5. Youth unemployment				
1980	15.0	3.9	25.2	14.1
1985	25.6	9.5	33.7	21.7

Notes: (a) 1985; (b) 1971; (c) 1982; (d) 1984.

Sources: (1) Standardized unemployment rates from OECD[1987a], Table R12; (2) OECD[1985], Table 11; (3) Labor turnover (including mobility between establishments) in the whole economy -in manufacturing for France and U.K.- from OECD[1986b], Table II-3, (4) Persons unemployedfor a year or more from OECD[1986b], Table K; (5) OECD[1986a], Table 10.

The steady increase in unemployment since 1973 and its acceleration in the 1980s resulted from a growing labor force and a flat employment level, the latter being the net outcome of declining agricultural and industrial employment and a slowly rising employment in services. We first describe the employment composition and flows and then those of unemployment.

In the last fifteen years, new jobs in Europe have had two characteristics: they have been created mainly by small firms and (except in Italy) they are part-time. Also, labor turnover rates -for both accessions and separations- have fallen considerably. There are no data on quits (except for Italy), but the available information points to a fall in both the quit and the layoff rate.

The composition of unemployment has shifted towards the unskilled, the young and the long term unemployed. With regard to the flows in and out of unemployment, the former have -surprisingly- grown modestly while the latter have dramatically decreased. As shown by Flanagan[1987], the likelihood of entering unemployment has not changed very much but the probability of finding a job once unemployed has sharply declined.

The review of these facts reveals that European unemployment is not a problem of excessive job destruction but of lack of job creation, so that any theory needs to explain why firms have become more reluctant to hire and/or why the unemployed have become choosier about jobs.

The interest in firing costs comes from their consistency with those stylized facts. The change to a bleaker outlook after 1973 made firms want to stop hiring and to dismiss more. But the increase in firing costs around 1975 would make them fire much less and also hire less, hence the dramatic fall in turnover and the stagnant employment level.

Second, as pointed out by Krugman[1987], with low flows into and out of jobs, both the duration of unemployment and youth unemployment have to increase. Finally, jobs not covered by employment protection legislation should increase at the expense of those covered by it. Hence the rise in small-firm, part-time and fixed-term contract jobs.

c) Survey of the previous literature:

The quasi-fixity of labor has been recognized by economists for a long time, starting with the seminal work of Oi[1962].

Adjustment costs have usually been modeled as being strictly convex and, more specifically, quadratic;[6] then, since faster adjustment is increasingly costly, the reaction to any shock takes place over a prolongued period of time.

On the one hand, the convexity assumption is hardly ever justified in the literature. Holt *et al.*[1960] say it is a "tolerable approximation over a range". It is normally used for ease of computation: the linear-quadratic framework yields certainty equivalence. As the model we present below makes clear, this simplification leaves out an important aspect of a firm's optimization problem: the degree of uncertainty about the future is in reality one of the main determinants of a firm's employment policy.

On the other hand, with employment protection laws the main source of firing costs in Europe today, it seems that a fixed cost per employee is a better approximation to reality (as noted in Nickell[1987]). Therefore we use linear asymmetric adjustment

[6] For example Holt *et al.*[1960], Solow[1968] or Sargent[1978].

costs in our model.[7]

Some theoretical work has been devoted to model labor demand with linear adjustment costs. Some models, in continuous time, ignore uncertainty: Kemp and Wan[1974], Nickell[1978, 1987] and Leban and Lesourne[1980]; the latter two study the implications of such costs on labor demand over the cycle. There are papers that do consider uncertainty in discrete time: Caplin and Krishna[1986] study the optimality of the labor demand rule with discrete Markov shocks and Gavin[1986] has a three-period model with only firing costs and serially correlated shocks to the marginal product. Kelsey[1986] and Bentolila[1987] present infinite-horizon models with uncertainty arising from serially independent shocks. In this Chapter we model in continuous time the labor demand decision of a firm subject to nonstationary demand uncertainty.

On the empirical side, apart from the work on quadratic costs in the U.S.,[8] Hamermesh[1987] shows that while the quadratic and the fixed adjustment costs models cannot be distinguished at the aggregate level, the latter performs much better at the individual plant level. Burda[1986] estimates a model with linear and quadratic adjustment costs, finding the latter non-significant in several European countries. Finally, Burgess and Dolado[1987]

---

[7]It would be possible to solve the model under the assumption that hiring and/or firing entails a fixed cost, independent of the number of workers involved (the techniques in Harrison *et al.*[1983] could be adapted for this purpose). This is certainly realistic at the firm level, and generates (S,s)-type employment policies which are the very opposite of the optimal policies under convex costs of adjustment.

[8]For example Sargent[1978], Kennan[1979], Pindyck and Rotemberg[1983] or Shapiro[1986].

estimate, for U.K. manufacturing, a model of variable quadratic costs of changing output and find them significant. We present no econometrics here, but perform some simulations on our model, with realistic parameter values, in section 4 below.

3- A simple model of firing costs

In this section we propose a model of dynamic labor demand under uncertainty, and we briefly illustrate the solution technique.

a) Problem setup:

Consider a firm with a linear constant returns to scale production technology, that uses only homogeneous labor, L, as a factor of production, and faces a constant elasticity demand function:

[1] $$Q_t \leq A_t L_t$$

[2] $$P_t = (z_t / Q_t)^{1-\mu} \qquad 0 < \mu < 1$$

where Q_t denotes production and sales at instant t (inventories are ignored), P_t is the product price and μ is the inverse of the markup factor, so that the firm's monopoly power decreases when μ rises. A_t is labor productivity, which is assumed to grow at a deterministic exponential rate ϑ_a.

The position of the (direct) demand function depends on an index z_t, which evolves in continuous time as a geometric Brownian motion with constant mean growth rate ϑ and standard deviation σ:

[3] $$dz_t = z_t \, \vartheta \, dt + z_t \, \sigma \, dW_t$$

where W_t is a standard Wiener process. From [3], demand is expected to grow at exponential rate ϑ, but fluctuates randomly so that the outlook further and further in the future is increasingly uncertain.

The firm pays a wage, w[9], to its workers and it also bears labor adjustment costs: a hiring cost, H, per new employee and a firing cost, F, per dismissed worker. However, if the worker leaves voluntarily, the firm bears no firing cost.[10] The instantaneous exponential attrition rate is δ.

Since the marginal revenue from a constant elasticity demand function is always positive, equation [1] holds with equality and, from [1] and [2], revenues are equal to $z_t^{1-\mu}(A_t L_t)^{\mu}$.

The firm chooses an employment and pricing policy to maximize its value, defined as the expected present value of its cash flow over the infinite future:

$$\max E_t\left\{\int_t^{\infty} e^{-r(\tau-t)}\left[\left(z_\tau^{1-\mu}(A_\tau L_\tau)^{\mu} - wL_\tau\right)d\tau - \left(1(dX_\tau>0)H - 1(dX_\tau<0)F\right)dX_\tau\right]\right\}$$

$$\text{s.t. } dL_t = dX_t - \delta L_t dt$$

where X is a cumulative labor turnover process (dX>0 means hiring,

---

[9]The wage is assumed constant for simplicity. It would be possible to allow for a stochastic wage (a geometric Brownian motion process) and use the technique proposed in Chapter 1, as long as firing and hiring costs were proportional to the current level of the wage.

[10]As stated above, this is not appropriate for Italy. If capital markets function well, severance payments like the Italian ones can be simply modeled in this framework as part of the wage.

dX<0 means firing), r is given rate of return and 1(.) is the indicator function. By the usual feasible perturbation argument, in which the currently marginal worker -allowing for attrition- is viewed as the marginal worker through the infinite future, the following first order conditions can be derived (see Chapter 1):

$$[4a]\quad E_t\left\{\int_t^{\infty}\left[\mu A_\tau\left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-t)}\, d\tau\right\} = \frac{w}{r+\delta} - F$$

if $dX_t<0$

$$[4b]\quad \frac{w}{r+\delta} - F < E_t\left\{\int_t^{\infty}\left[\mu A_\tau\left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-t)}\, d\tau\right\} < \frac{w}{r+\delta} + H$$

if $dX_t=0$

$$[4c]\quad E_t\left\{\int_t^{\infty}\left[\mu A_\tau\left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-t)}\, d\tau\right\} = \frac{w}{r+\delta} + H$$

if $dX_t>0$

These conditions are easily interpreted. When firing, in [4a], the firm equates the discounted expected marginal revenue product (MRP) given up by dismissing a worker to the discounted wage cost saved from doing so, minus the dismissal cost paid today. When hiring, in [4c], the firm equates the discounted expected MRP that the newly hired worker will provide to the discounted wage cost plus the hiring cost today.

If there were no adjustment costs, the firm would hire whenever the expected MRP at the existing labor force was higher than the future discounted wage cost, and it would fire otherwise. The equations in [4] would then collapse to the simple rule that the MRP be at all times equal to the wage. However, with adjustment costs such a policy is not optimal since it implies a high turnover, which is now costly. Therefore, the firm does not necessarily hire immediately if demand picks up, since there are current hiring costs and future expected firing costs. Conversely, it is also more cautious before firing after a demand slowdown due to current firing and future expected hiring costs.

This reasoning explains condition [4b], i.e. a range or "corridor"[11] around the no-cost-of-adjustment labor demand, where inaction ($dX_t=0$) is optimal. This is shown in Figure 1.

To decide its labor demand, the firm has to calculate the expectation of the MRP in the future, which will in turn depend on its employment policy. Concavity of the revenue function makes the conditions in [4] sufficient as well as necessary to identify the unique optimal employment policy.

With the assumptions laid out above,[12] the optimal policy is simple: allow the MRP to fluctuate between a lower (ℓ) and an upper (u) control barrier, which are constant (see below). If the MRP goes below ℓ, dismiss workers so as to raise the MRP to ℓ; if

---

[11] Leijonhufvud[1973] coined this word in a different context. Dornbusch[1987] uses the term in the sense we do here.

[12] The crucial simplifying assumption is the independence over time of demand increments. Otherwise, the barriers would be state-dependent.

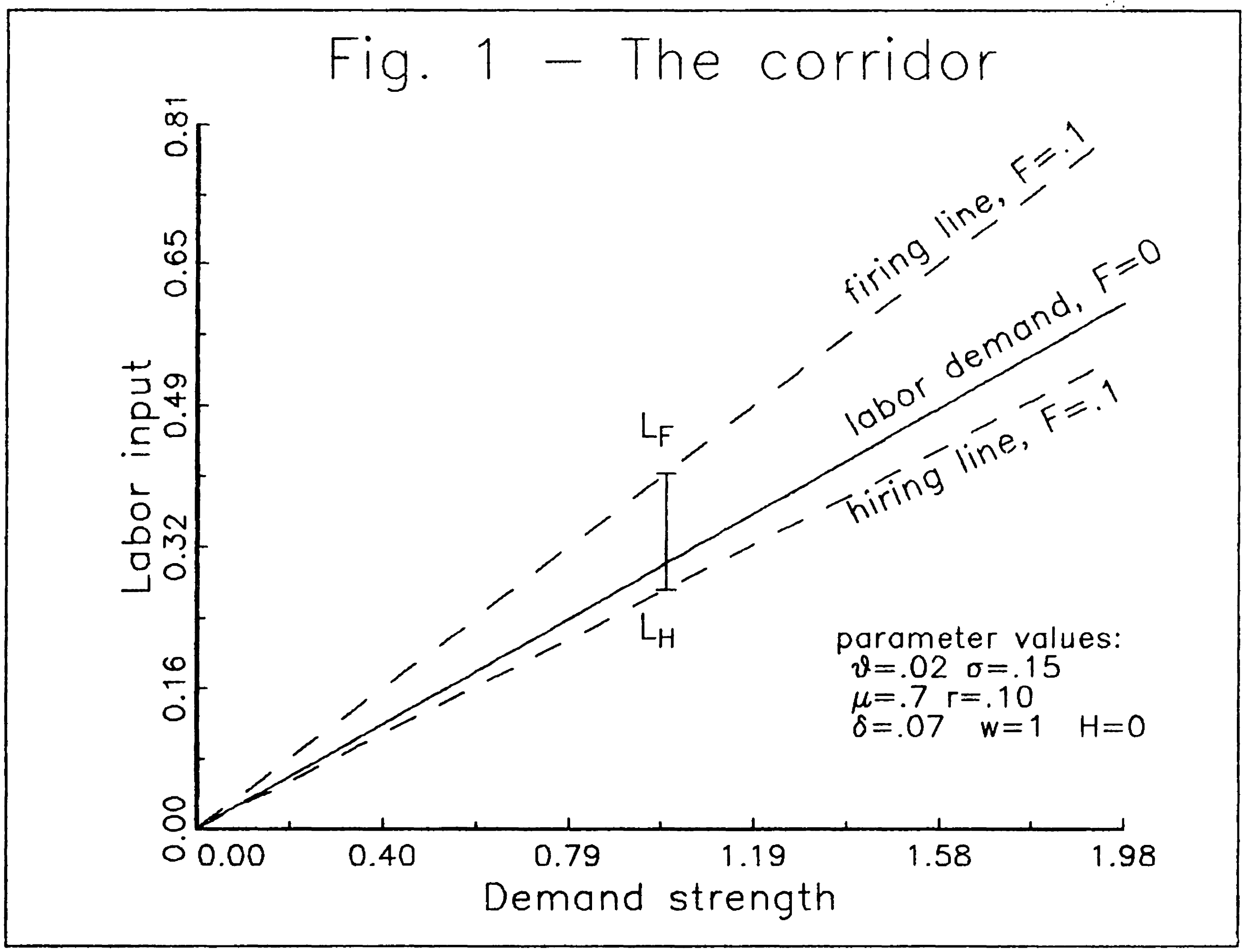


Fig. 1 — The corridor
Labor input
0.00
0.16
0.32
0.49
0.65
0.81
firing line, F=.1
labor demand, F=0
hiring line, F=.1
L_F
L_H
parameter values:
ϑ=.02 σ=.15
μ=.7 r=.10
δ=.07 w=1 H=0
0.00
0.40
0.79
1.19
1.58
1.98
Demand strength

it goes above u, hire new employees to bring the MRP back down to u.

There are two equivalent ways to characterize the rule just described. First, for a given demand level z_t, there exist two boundary labor force levels, L_F and L_H, which satisfy [4a] and [4c], respectively (see Figure 1). Suppose the firm has a given labor force inherited from the previous period, L_I. If L_I is higher than L_F the firm will fire down to L_F; if L_I is lower than L_H it will hire up to L_H; for L_I between L_H and L_F it will keep L_I.

A second interpretation is to take the current labor force as given and define two boundary demand index levels, z_F and z_H, which make [4a] and [4c] hold true, respectively. For all values of z_t below z_F the firm will fire, for values above z_H it will hire and for z_t between z_F and z_H it will stay put.

b) Solution of the model:

The task at hand is to find the optimal resetting points u and ℓ, as a function of the parameters. It is possible to show (by Ito's lemma, see e.g. Harrison[1985]) that the MRP follows a geometric Brownian motion when neither hiring nor firing is taking place:

$$d\left[\mu A_t\left(\frac{z_t}{A_t L_t}\right)^{1-\mu}\right] = m\left[\mu A_t\left(\frac{z_t}{A_t L_t}\right)^{1-\mu}\right] dt + s\left[\mu A_t\left(\frac{z_t}{A_t L_t}\right)^{1-\mu}\right] dW_t$$

with drift $m \equiv \mu\vartheta_a+(1-\mu)(\vartheta+\delta-\mu\sigma^2/2)$ and instantaneous standard deviation $s \equiv (1-\mu)\sigma$. The optimal labor demand policy just described implies that the firm regulates (in the sense of Harrison[1985]) the MRP process, not allowing it to go below ℓ, or above u.

It is then possible to compute the discounted expectation of the MRP appearing in the first order conditions, as a special case (nonstochastic adjustment costs) of the general problem in Chapter 1; the following is proved in Appendix 1 to this Chapter:

<u>Result</u>: Let $d\eta_t = m\,\eta_t dt + s\,\eta_t\,dW_t$ be a regulated geometric Brownian motion, with starting value η_0 at time 0, upper control barrier at u and lower control barrier at ℓ; then,

$$E_0\left\{\int_0^\infty \eta_t e^{-\lambda t}\,dt\ ;\ \eta_0, u, \ell\right\} =$$

$$= \frac{\eta_0}{\lambda-m} + \frac{\eta_0^{\alpha_1}\left[u^{\alpha_2}\ell - u\,\ell^{\alpha_2}\right]}{(\lambda-m)\alpha_1\left[u^{\alpha_1}\ell^{\alpha_2} - u^{\alpha_2}\ell^{\alpha_1}\right]} + \frac{\eta_0^{\alpha_2}\left[u\,\ell^{\alpha_1} - u^{\alpha_1}\ell\right]}{(\lambda-m)\alpha_2\left[u^{\alpha_1}\ell^{\alpha_2} - u^{\alpha_2}\ell^{\alpha_1}\right]}$$

$$\equiv f(\eta_0, u, \ell; m, s, \lambda)$$

where α_1 and α_2 are (respectively) the positive and negative roots of the equation $(s^2/2)\,\alpha^2 + (m - s^2/2)\,\alpha - \lambda = 0$.

We can now insert the expectation of the MRP (a highly nonlinear expression in u and ℓ) into the conditions in [4], which then read:

$$[4a] \qquad f(\ell, u, \ell\ ;\ m, s, r+\delta) = \frac{w}{r+\delta} - F \qquad \text{if } dX_t < 0$$

$$[4c] \qquad f(u,u,\ell\ ;\ m,s,r+\delta) = \frac{w}{r+\delta} + H \qquad \text{if } dX_t>0$$

If F = H = 0, then it is possible to show that $u = \ell = w$ solves [4a,c]. If $F \geq w/(r+\delta)$ then hiring decisions are effectively irreversible (firing is never optimal), and a closed form solution can be found as in Chapter 2. But in general it is not possible to solve in closed form, and so the two equations have to be solved numerically. Once u and ℓ are known, the corresponding hiring and firing points for labor (L_H and L_F) as a function of the strength of demand, z_t, and productivity, A_t, can be found by inverting the MRP: this completely describes the firm's dynamic labor demand policy.

We are also interested in finding out where, between the barriers, is a firm likely to be at any point in time. It is possible to derive the steady-state distribution of the MRP between u and ℓ[13], along the lines described in Chapter 2. Appendix 1 shows that the ergodic density function of the MRP is a power function (the logarithm of the MRP is exponentially distributed), with mean given by the following nonlinear function of u and ℓ:

$$\bar{\eta} = \left[\frac{m - s^2/2}{m}\right] \frac{u^{\left[\frac{2m}{s^2}\right]} - \ell^{\left[\frac{2m}{s^2}\right]}}{u^{\left[\frac{2m}{s^2} -1\right]} - \ell^{\left[\frac{2m}{s^2} -1\right]}}$$

---

[13]Labor demand does not possess a steady-state distribution, since the demand index is nonstationary.

Of course, since u and ℓ are not in closed form, we have to resort to the numerical calculation of the ergodic mean.

It is also possible to solve the dynamic optimization problem of the firm by marginal option valuation. When hiring a worker, the firm is exercising a call option: the option to purchase (at a price equal to the worker's discounted wage bill, plus a hiring cost) a package containing an asset which pays dividends equal to the currently marginal worker's MRP, and a put option to sell the same asset (at a price equal to the worker's discounted wage bill, minus the firing cost). When firing a worker the firm engages in the symmetric operation: it sells the asset-cum-put-option package and receives the call option. The optimal timing for these operations can be derived by arbitrage arguments, and it can be shown that the resulting employment program is the same as that derived by the dynamic programming arguments above.[14]

The analysis below focusses on how the boundary labor levels L_H and L_F and the average labor demand depend on the firing cost, and also on how the other parameters in the model affect such dependence.

In what follows it will be assumed that firing costs are significantly larger than hiring costs, as seems to be the case in Europe (see section 4 and Appendix 3).

[14] See Chapter 1. The model in this Chapter is isomorphic to the general problem considered there, with P_t=H+w/(r+d) for all t and p_t=-F+w/(r+δ) for all t.

c) Effects of a reduction in firing costs:

We now study the effects of a decrease in dismissal costs.[15] This is an exercise in comparative dynamics, since we are not allowing for parameter changes when we solve the firm's problem: we are comparing unrelated economies, each endowed with a set of immutable parameter values.

A fall in F makes both current and future firing less expensive. This makes future firing more likely and, through a reduced likelihood of inaction, future marginal profits are more heavily discounted. The narrowing of the corridor makes the expected time elapsed before hitting a barrier decrease. In the first order conditions in [4] these effects appear in the marginal revenue product process and workers' expected tenure length.

For firing decisions (so that [4a] applies), the fall in the cost of dismissing implies that the expected MRP of the marginal worker has to be higher, i.e., the firm fires more. The magnitude of this effect is large, since the current firing cost is neither uncertain nor discounted.

The impact of a fall in F on hiring decisions is not immediate from equation [4c], and so we rewrite it with F appearing explicitly. Define T (a random variable as of time t) as the first firing time after hiring time t. Then, by an application of the strong Markov property of Brownian motions (see Appendix 2) equation [4c] becomes:

---

[15]For a detailed exposition of the discrete time case, see Bentolila[1987].

[4c']

$$E_t\left\{\int_t^T\left[\mu A_\tau\left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right]e^{-(r+\delta)(\tau-t)}d\tau\right\} =$$

$$= H + \frac{w}{r+\delta}\left[1-E_t\left\{e^{-(r+\delta)(T-t)}\right\}\right] + F\ E_t\left\{e^{-(r+\delta)(T-t)}\right\}$$

where $dX_t>0$, $dX_T<0$, $T>t$.

This equation has a fairly intuitive interpretation: the firm will hire a worker when the marginal revenue he is expected to provide before he is fired (or quits) equals the hiring cost plus the present value of his wages while in the firm, plus the firing cost at the (random) firing time T, discounted to the present. The latter is the "shadow hiring cost" component of firing costs, which shows that the existence of such costs inhibits hiring.

The reduction in firing costs is the fall in F in the right hand side; but it also shortens workers' expected tenure, (E_tT-t), which affects all the other terms in [4c']. More specifically, since $r+\delta>0$, it increases the discount factor multiplying the firing cost F, reduces that multiplying the wage cost and it also decreases the number of periods over which the MRP is taken into account.

The total effect is positive, i.e. a decrease in firing costs will make the firm less reluctant to hire; but the larger F the smaller is this effect. The reason is that the larger F, the longer is expected tenure (E_tT is higher), and the smaller is the discount factor multiplying F (i.e. closer to zero), and so a given reduction in F will not reduce the right hand side of [4c'] as much. Thus, the increase in labor demand will be smaller too.

The implication is that a reduction in firing costs need not have a significant effect on the propensity to hire, while -as shown above- it will definitely increase the propensity to fire. On the other hand, the propensity to hire is very much affected by <u>hiring</u> costs, which are not explicitly dealt with in this section.

Figure 2 shows the asymmetry in the impact of F on the labor demand boundaries (in terms of the MRP, the graph would be the mirror image, i.e. u -which corresponds to L_H- would slightly increase while ℓ would strongly decrease). The dashed line is the steady-state mean of labor demand for a given value of the demand index z. For the realistic parameters values used in the Figure, the size of firing cost has practically no influence on average steady-state employment, which is however slightly <u>higher</u> if firing costs are large. Similar computations reveal that average labor demand is a strongly <u>increasing</u> function of firing costs if the attrition rate δ and the growth rate ϑ are small, and the uncertainty parameter σ is large.

Figure 3 compares the steady-state distribution of the MRP (derived in Appendix 1) for economies with different firing costs.

In the absence of both hiring and firing costs, the MRP has a degenerate distribution: it is a spike at w=1. Small adjustment costs (in this case F) already cause the distribution to spread out considerably. Higher F means that in bad times the firm will be less willing to fire and so the likelihood of observing lower values of the MRP (higher employment) increases, and that is why

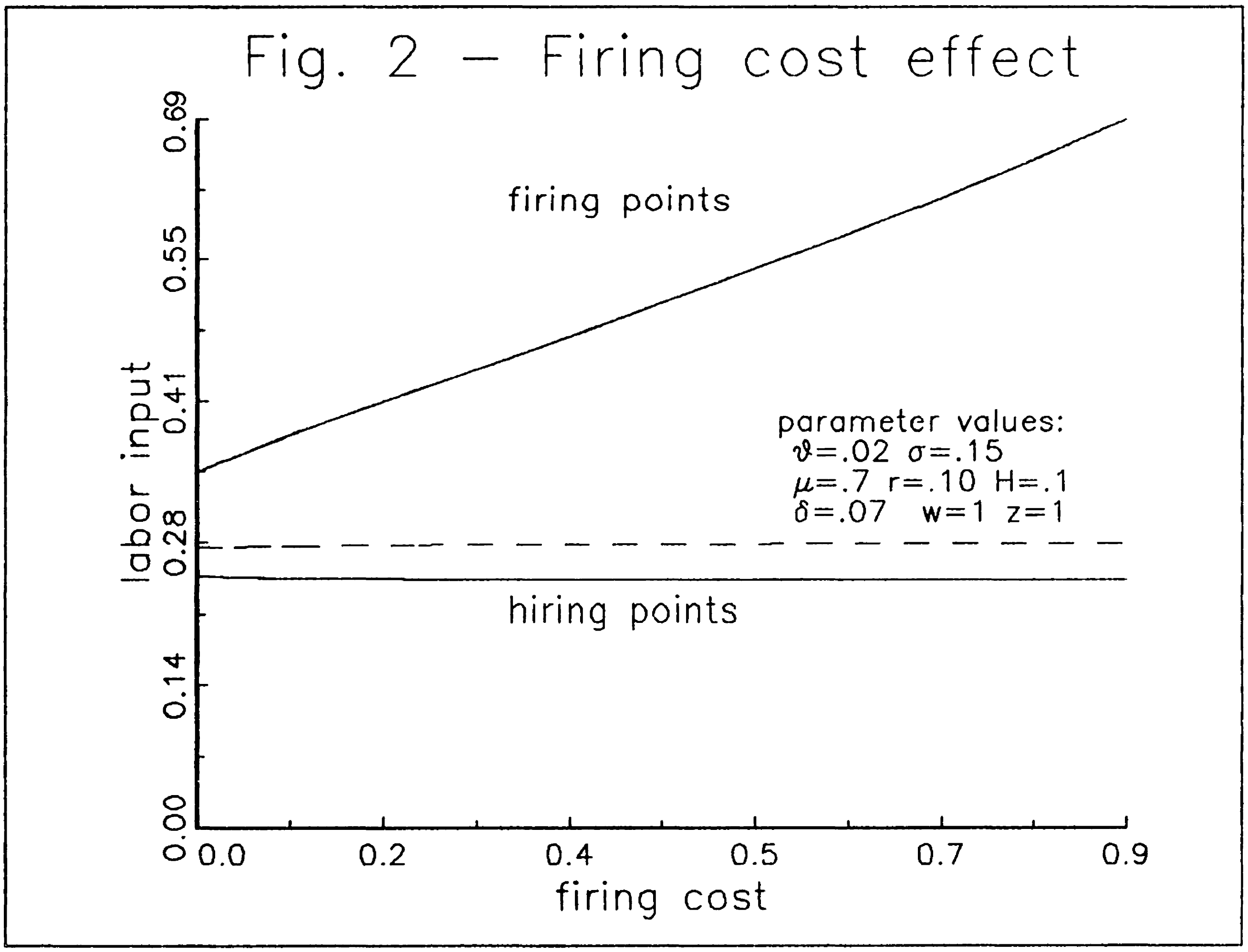


Fig. 2 – Firing cost effect
firing points
parameter values:
ϑ=.02 σ=.15
μ=.7 r=.10 H=.1
δ=.07 w=1 z=1
hiring points
labor input
0.00
0.14
0.28
0.41
0.55
0.69
0.0
0.2
0.4
0.5
0.7
0.9
firing cost

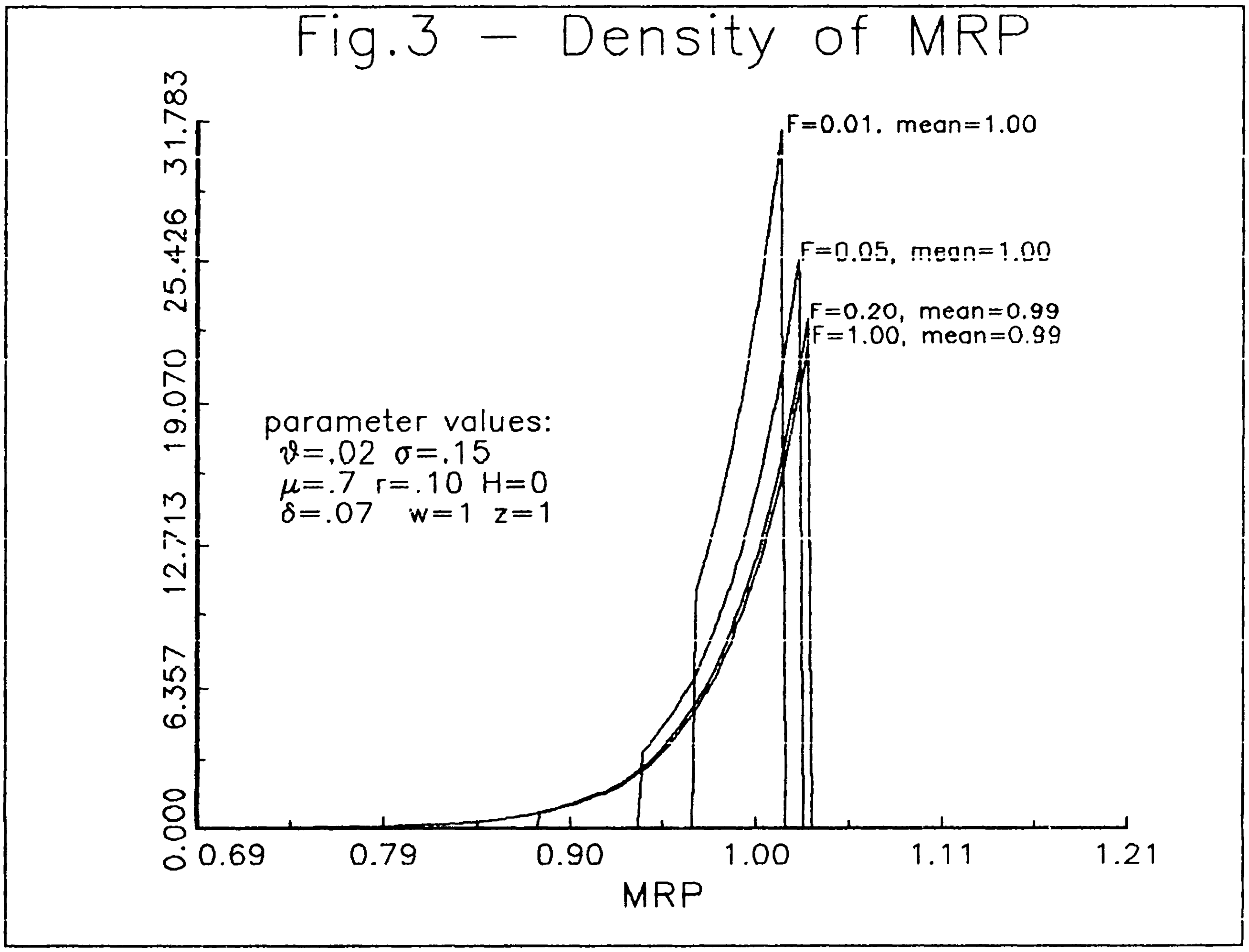


Fig.3 – Density of MRP
F=0.01, mean=1.00
F=0.05, mean=1.00
F=0.20, mean=0.99
F=1.00, mean=0.99
parameter values:
ϑ=.02 σ=.15
μ=.7 r=.10 H=0
δ=.07 w=1 z=1
0.000
6.357
12.713
19.070
25.426
31.783
0.69
0.79
0.90
1.00
1.11
1.21
MRP

the mean of L increases in Figure 2.

The distribution is strongly skewed towards the hiring barrier because there is a natural tendency to hire if, realistically, the demand growth and attrition rates are positive. Thus the probability of a dismissal is quite low already at moderate firing cost levels. This explains the lack of a discernible effect on the hiring boundary in Figure 2, and makes the large rise in the firing boundary relatively unimportant: once firing costs are large, the firm will very seldom be near the firing barrier.

c) Comparative dynamics:

Once the effects of adjustment costs on labor demand are known, it is interesting to ask how other parameters enhance or dampen their impact. Again this is an exercise in comparative dynamics, i.e. we study the behavior of firms facing different environments but we do not model the firms' reaction to changes in the parameters.

An increase in the mean growth rate of demand, ϑ, reduces the probability of future desired firing and increases that of future hiring. Workers' expected tenure increases, making firing costs less important (but hiring costs more important). This raises expected marginal profits and thus raises both boundary labor levels, L_F and L_H (see Figure 4). As ϑ rises, the firm is more likely to be near the hiring barrier, therefore the distribution

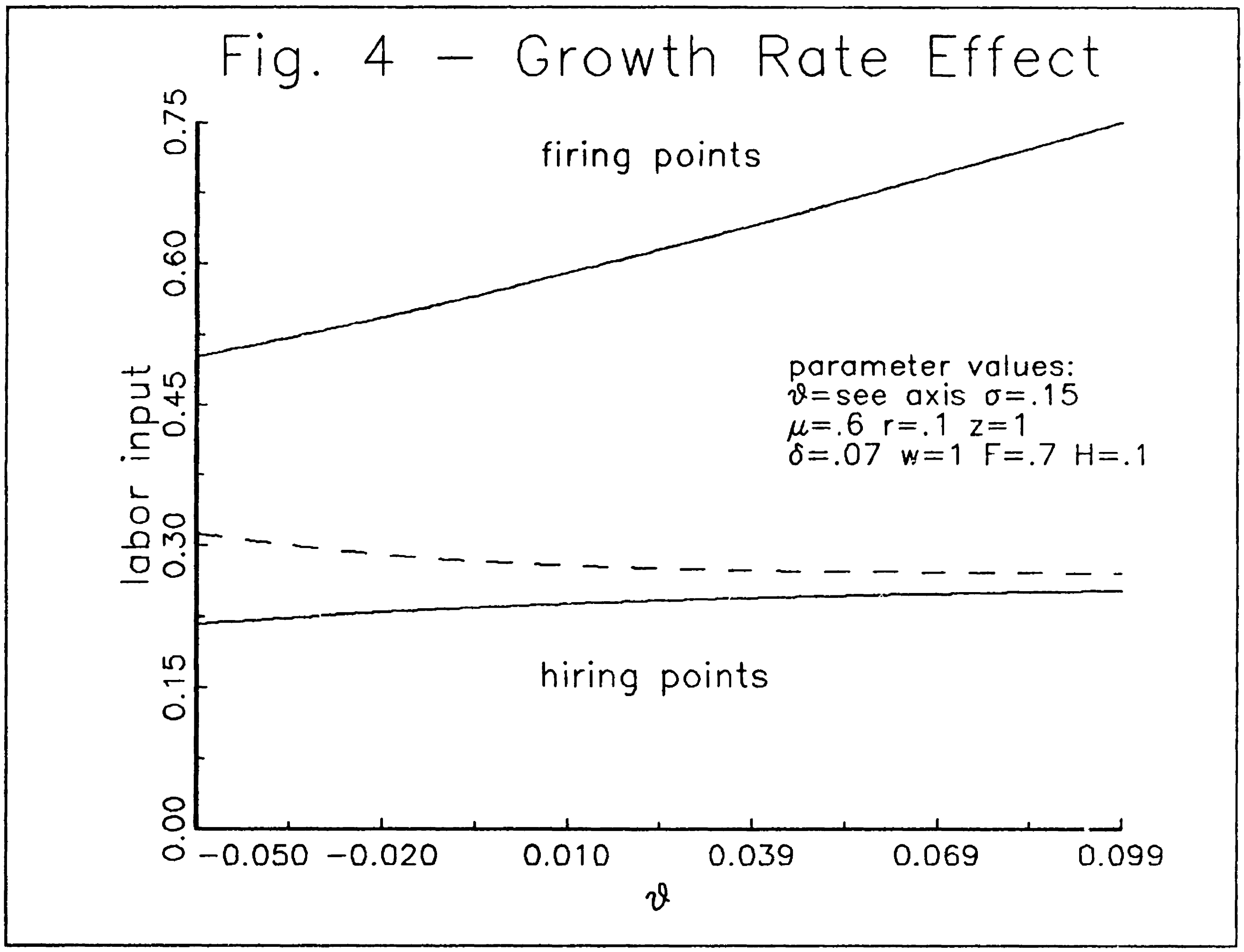


Fig. 4 – Growth Rate Effect
firing points
hiring points
parameter values:
ϑ=see axis σ=.15
μ=.6 r=.1 z=1
δ=.07 w=1 F=.7 H=.1
labor input
0.00
0.15
0.30
0.45
0.60
0.75
-0.050
-0.020
0.010
0.039
0.069
0.099
ϑ

skews towards L_H and so does its mean (dashed line in the graph).[16]

The effect of an increase in the rate of productivity growth, ϑ_a, is qualitatively similar to that of demand growth: both productivity and demand growth increase revenues (although price and quantity are affected differently).

When the variance of demand, σ^2, increases, the likelihood of large changes in the demand index rises, so that the firm would be likely to dismiss and hire more often, and both types of adjustment costs become more important. This discourages both hiring and firing, and the corridor opens up on both sides in Figure 5: the firm will start firing at a higher L_F and hiring at a lower L_H. With low σ, F higher than H and positive growth and attrition, the distribution is skewed towards hiring. As σ increases, the distribution spreads out and so the mean shifts towards the firing barrier (dashed line in the graph).

An increase in the attrition rate, δ, works very much like demand growth (Figure 4), making firing costs less relevant (workers leave voluntarily more often) but hiring costs more important (there are more quits to replace and train); i.e. both L_F and L_H increase.[17] Attrition makes hiring more likely for the firm; the mean shifts towards L_H and is lower the higher is the

---

[16]This is a statement about the steady-state distribution of the MRP: the picture depicts labor demand for a given Z value. Of course, an economy with higher ϑ will have higher labor demand over time as Z will grow faster.

[17]If H is high enough, L_H may fall, as a function of δ, over a range of high attrition rates.

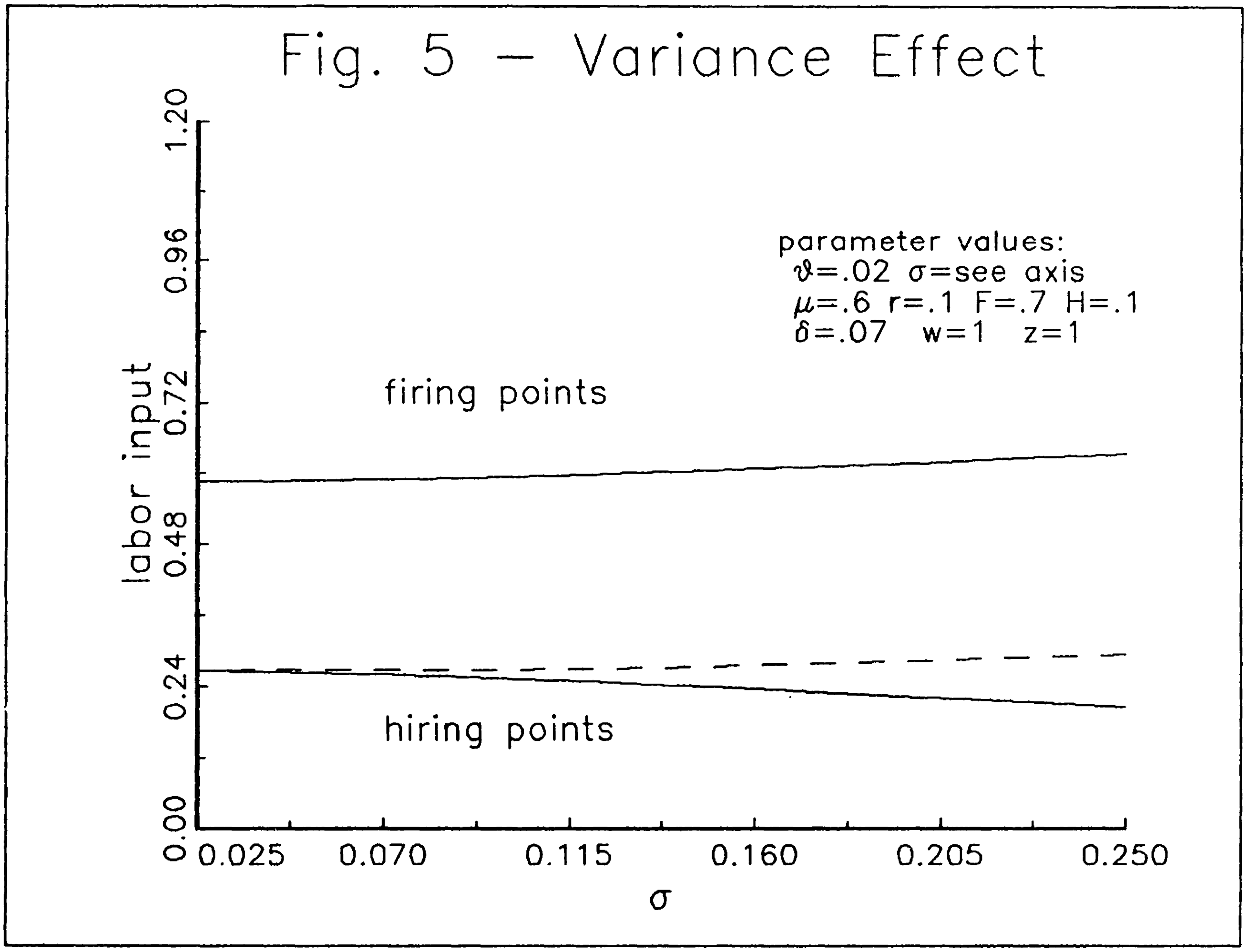


Fig. 5 - Variance Effect
parameter values:
ϑ=.02 σ=see axis
μ=.6 r=.1 F=.7 H=.1
δ=.07 w=1 z=1
firing points
hiring points
labor input
0.00
0.24
0.48
0.72
0.96
1.20
0.025
0.070
0.115
0.160
0.205
0.250
σ

attrition rate.

A rise in the required rate of return, r, also lowers the weight of future profits, so the impact is like that of an increase in δ.

Finally, consider an increase in μ, the inverse of the markup factor, i.e. a reduction of the firm's monopoly power. This reduces the elasticity of the profit function to employment: the marginal profit curve is flatter. In addition to supplying more output for any given value of z_t (i.e. the labor demand mean moves towards the hiring barrier), the firm is not very much hurt by deviations of employment from the no-frictions optimum, and the corridor widens.

The combined effects of all parameters are hard to gauge. But the most interesting interaction is the one between the magnitude of firing costs and the degree of uncertainty about the future evolution of demand, which is illustrated in Figure 6: firing costs are more relevant the more uncertain is the firm's outlook. The main contribution of the model proposed in this section is the explicit treatment of uncertainty and dynamics.

4- Numerical solution of the model

a) The two regimes:

We now solve the simple model laid out in the preceding section, for realistic parameter values, in an attempt to quantify the effects of firing costs in the European experience.

We take the first oil price shock as a watershed between two

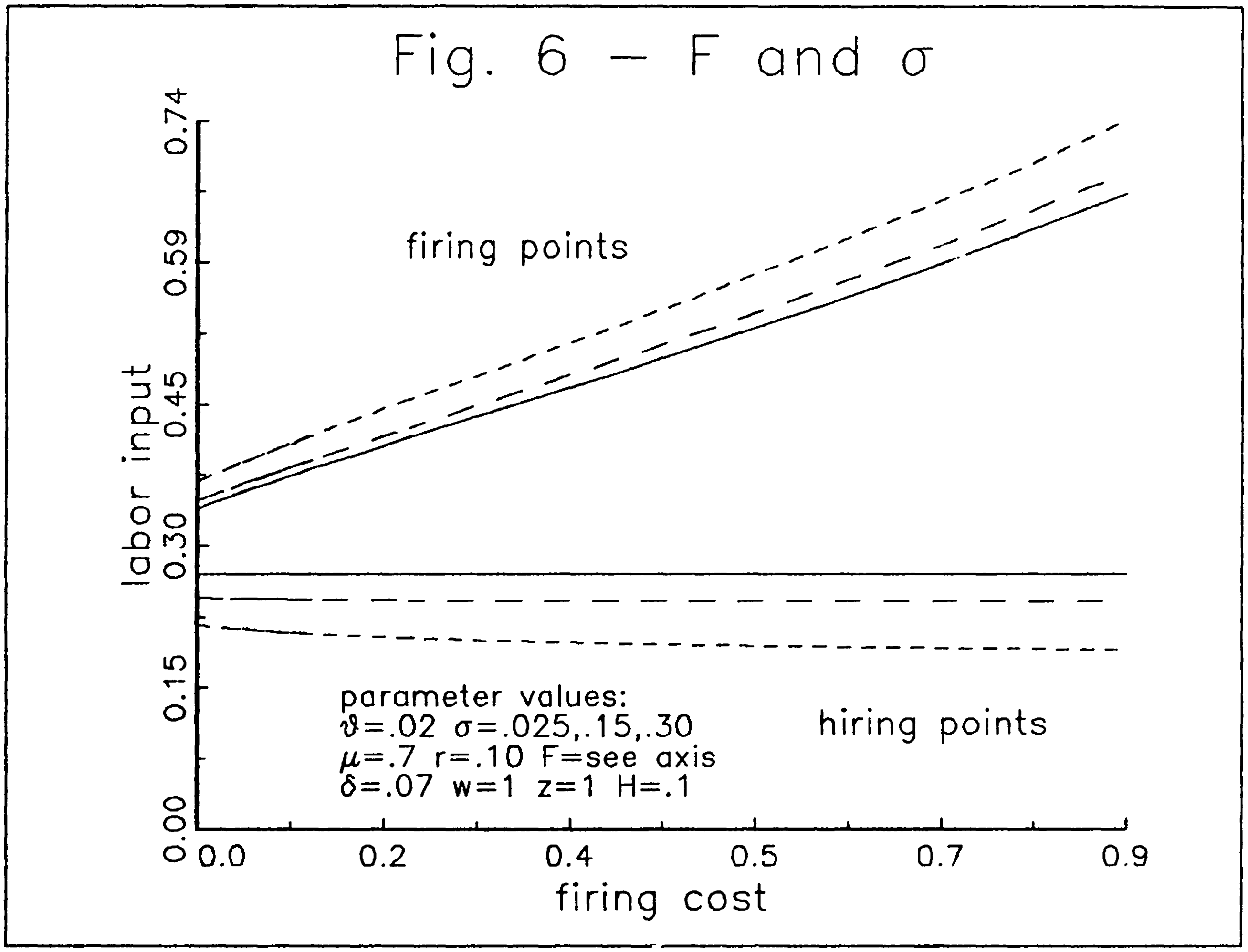


Fig. 6 — F and σ
labor input
0.00
0.15
0.30
0.45
0.59
0.74
firing points
parameter values:
ϑ=.02 σ=.025,.15,.30
μ=.7 r=.10 F=see axis
δ=.07 w=1 z=1 H=.1
hiring points
0.0
0.2
0.4
0.5
0.7
0.9
firing cost

distinct "regimes". Before 1973 demand growth was quick and steady, productivity growth was high and workers were not afraid of quitting, because jobs could easily be found. In contrast, after 1973 demand growth became low and more volatile, productivity slowed down and workers became reluctant to quit because of the difficulty in finding jobs. Finally, around 1975, labor security provisions were tightened by governments and unions trying to avoid massive dismissals.

We define the years from 1961 to 1973 as the first regime and 1975 to 1986 as the second regime.[18] Table 2 shows the sample period averages for the two regimes. ϑ and σ are obtained from the industrial production index. σ is equal to three times the standard deviation for the industrial sector as a whole, as a way to capture the fact that firms suffer from idiosyncratic shocks that are averaged out in the aggregate data (for the U.S., this has been documented by Romer[1987] at the industry level). The estimate of δ comes from several sources[19] and we derive rough proxies for F from the legislation and actual data (see Appendix 3).[20] With two exceptions (σ in Germany and ϑ_a- productivity growth - in U.K.) the stylized facts are confirmed by the data.

---

[18] We exclude the year 1974 to have a clean separation of regimes.

[19] In Italy, the drop in δ might be exaggerated, since δ corresponds to firms with 10 workers or more in Regime 1 and with 50 workers or more in Regime 2, and there is evidence showing that labor turnover is highest in small firms in Italy, see Contini and Revelli[1987].

[20] In the absence of enough data on the rise of F, we set it to one third.

Table 2: Parameter values

Regime*	FRANCE 1	FRANCE 2	GERMANY 1	GERMANY 2	ITALY 1	ITALY 2	U.KINGDOM 1	U.KINGDOM 2
ϑ	5.5	0.7	4.8	1.4	6.2	1.2	2.9	1.0
σ	6.5	12.4	11.4	10.5	10.0	16.2	8.6	11.6
δ	12.7	7.9	22.0	14.1	14.7	6.5	22.7	12.7
ϑ_a	4.9	2.0	4.6	2.9	5.1	2.4	3.4	3.7
F	.687	.916	.562	.750	.812	1.083	.187	.250

Other parameter values: r=15, μ=.6, H= 1 month's wages (i.e. .083).
Notes: Parameters are in percentage terms, at annual rate; hiring and firing costs are in terms of years of wages (see Appendix 3).

* 1=1961 to 1973, 2=1975 to 1986.

The results we are about to present should be taken with caution. First, ϑ, σ and ϑ_a are calculated from production and -for ϑ_a- employment, which are the net result of endogenous decisions of all firms in the industrial sector.

Second, for complete realism we should consider many sources of uncertainty (such as energy prices, exchange rates, and monetary and fiscal policy), and allow the firm to have several sources of flexibility, such as the substitution of capital or materials for labor, the use of overtime, temporary layoffs or subcontracting, and the management of inventories. We think, however, that industrial production summarizes a firm's stochastic environment sufficiently well for the exercise performed here, and we subsume all the firm's flexibility in the single parameter μ, which determines the instantaneous revenue function's elasticity to employment.

Third, legally mandated firing costs started to go down early in the U.K. (1979) and Italy (1980), so that our regime 2 was in these two countries much shorter than we assume. Finally, we have not calibrated the results with the (very scarce) available data on tenure lengths: with firing costs as large as those reported in the table, firms should very seldom fire in our model, and this may or may not be confirmed by the data.

Let us focus on the effects of the regime change on the boundary levels L_H and L_F, shown in Table 3. The changes in ϑ, ϑ_a and, especially, σ lower the hiring boundary L_H, i.e. firms have a lower labor force before they start hiring. The reason for the diverging impact of δ is that lower attrition makes hiring costs matter less.

On the other hand, the reductions in ϑ, ϑ_a and δ make L_F fall. With lower expected demand and productivity growth, and less quitting, firms start firing sooner. The increase in volatility (σ) makes firms be more cautious, so this (small) effect goes in the opposite way, as does the increase in F, which is the single most important determinant of L_F.

The effects on the hiring barrier are smaller than those on the firing barrier. This is reasonable since the fall in ϑ, ϑ_a and δ all make firing costs more important and hiring costs less relevant. Still, the effect of higher uncertainty and lower growth on the hiring boundary is far from negligible: in France and Italy, where firing costs are more important, the change in the parameters induces a 4-7% lower <u>marginal</u> desired employment in our model, with the wage and the demand level given. Although we do not model the supply side of the labor market, and real wage stickiness is of course necessary for any unemployment pattern to be observed, this increased reluctance to hire can probably explain the prolonged high levels of unemployment even in the comparatively strong economic environment of the 1980s: this is precisely the point stressed by Blanchard *et al.*[1986].

The steady-state average labor demand is hardly affected by the parameter changes, except marginally for the variance and more importantly for the attrition rate. The combined effect is an increase in the mean: firms are on average closer to the firing barrier in regime 2 than in regime 1 (the percentage changes are: France, 1.8%, Germany, 1.7%; Italy, 4%; and U.K., 2.2%).

Table 4 shows the effects of the change in regime on the marginal revenue product boundaries u and ℓ. This figures are interesting because, by reversing their sign, they show by how much would the wage have to fall to induce hiring or firing at the same point as before the regime change, other things (such as demand strength) equal. Finally, Table 5 shows the effects of different percentage decreases in firing costs on the boundary labor levels once the economy is in regime 2. It confirms our previous assertion that a change in firing costs strongly affects the propensity to fire but not the propensity to hire.

Table 3: Effects of the change in regime on L_F and L_H (%)

	FRANCE		GERMANY		ITALY		U. KINGDOM	
	L_H	L_F	L_H	L_F	L_H	L_F	L_H	L_F
ϑ	-0.2	-10.4	-0.2	-5.8	-0.4	-11.0	-0.1	-2.1
σ	-2.1	0.1	0.3	-0.0	-2.8	0.1	-1.0	0.1
δ	0.8	-14.7	1.0	-19.2	1.0	-26.8	1.5	-12.4
ϑ_a	-0.2	-9.4	-0.2	-4.4	-0.3	-9.0	0.0	0.5
F	0.0	32.1	0.0	35.3	0.0	43.7	0.0	11.5
All	-4.3	-17.3	.8	-10.9	-6.8	-27.8	-0.2	-6.5

Table 4: Effects of the change in regime on MRP_F and MRP_H (%)

	MRP_H	MRP_F	MRP_H	MRP_F	MRP_H	MRP_F	MRP_H	MRP_F
ϑ	0.1	4.5	0.1	2.4	0.1	4.8	0.0	0.9
σ	0.9	-0.1	-0.1	0.0	1.1	-0.0	0.4	-0.0
δ	-0.3	6.6	-0.4	9.0	-0.4	13.3	-0.6	5.5
ϑ_a	0.1	4.0	0.1	1.8	0.1	3.8	-0.0	-0.2
F	0.0	-10.5	0.0	-11.4	0.0	-13.5	0.0	-4.3
All	1.8	7.9	-0.3	4.7	2.9	13.9	0.1	2.7

Table 5: Regime 2- Effects of a fall in firing costs on L_F and L_H (%)

	L_H	L_F	L_H	L_F	L_H	L_F	L_H	L_F
10%	0.0	-7.6	0.0	-8.3	0.0	-8.5	0.0	-3.2
20%	0.0	-14.5	0.0	-16.0	0.0	-16.2	0.0	-6.3
50%	0.0	-32.6	0.0	-35.7	0.0	-35.9	0.0	-15.5

The parameter values in the previous subsection are, at most, ballpark figures, and correspond to the whole industrial sector. There may be measurement errors, and individual sectors behave differently from the aggregate. In providing parameter values for our solutions, we have tried to get reasonable estimates of firm-level uncertainty and flexibility. The behavior of sectoral employment depends on these parameters in a somewhat loose way, but we leave the treatment of aggregation for future research. Also, while demand for industrial products has been growing on average, some industrial subsectors' production has been steadily declining. Therefore, while the average growth and attrition we find allow for infrequent firing, negative growth (which is usually accompanied by very low attrition) means that sectoral firing will occur, and this would not show up in our calculations above.

Sensitivity analysis was performed to check the robustness of the results. These are not reported in detail, given the roughness of the exercise. The main conclusions are: (a) The fall in L_H is very robust to most parameter values, but sensitive to the values of σ, δ and H; the fall in L_F is quite sensitive to all parameters, specially to F. (b) If volatility at the firm level is higher than we assume, L_H falls more and L_F falls less than in the baseline, which accords with the European experience (hiring freeze, not much firing).

b) - Implications for the European experience:

We now explore the implications of our model for the behavior of labor demand in Europe. There are two different issues: first, the change in the steady-state distribution of the MRP inside the corridor caused by the regime change and, second, the dynamics, both those caused by movements of z (the fundamental business conditions) within a regime, and those following the regime transition..

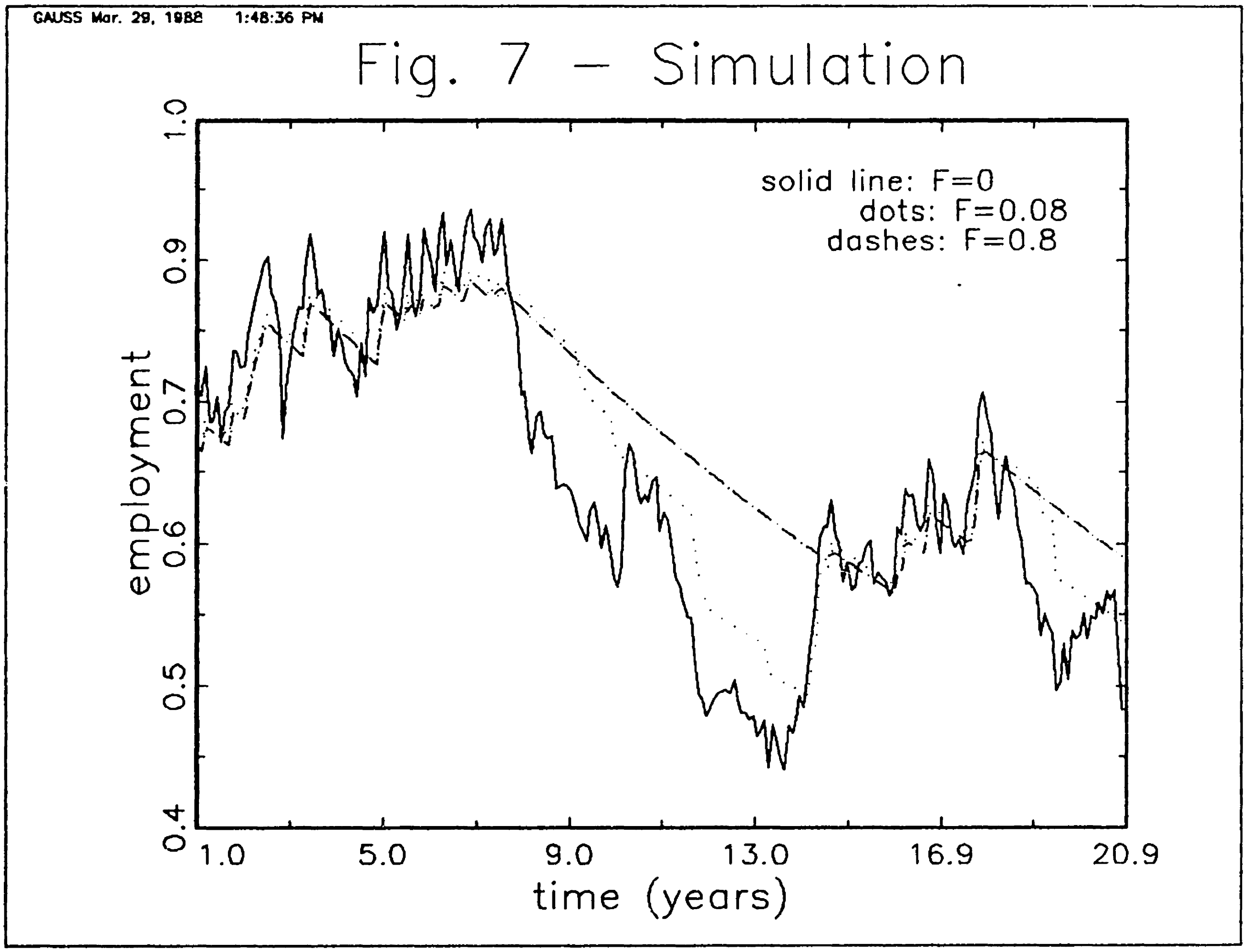
Section 3 derived the steady-state distribution of labor's marginal revenue product: with realistic parameter values, this distribution is skewed towards the hiring barrier. Out of the several parameter changes, let us first address the increase in F.

Average long-run labor demand, which is inversely related to the average MRP, is an increasing function of the magnitude of firing costs (although the increase is very shallow with realistic parameter values, it becomes very pronounced if the attrition rate is small or the demand growth rate is negative). This is at first sight quite surprising: we are raising firing costs and keeping the wage fixed, hence increasing the cost of labor, and we would expect the firm to use labor more sparingly. But firing costs prevent firing so much more than hiring that they increase average employment. There is a "ratchet effect" built in the optimal policy: the firm knows that marginal workers may one day have a low MRP and/or firing costs will have to be paid, but this possibility is heavily discounted since hiring occurs in good times, and bad times are far into the future. Ex-post, firing is less likely to occur if firing costs are large, and so average employment increases -which may be a rationale for workers organizations' support for severance-pay legislation.

The impact of higher firing costs on the steady-state average labor demand is almost negligible with the parameters in our Table 2 baselines (which are realistic for the industrial sector of those European economies). But the effect on the ergodic mean of the large fall in the attrition rate is high (from 1% in France to 2.1% in the U.K.): in the new environment, firms are much more reluctant to hire, but now hired workers quit much more rarely. The positive impact on the long-run mean is also magnified by the higher degree of uncertainty. In summary, all parameter shifts tended to lower the average MRP, i.e. raise the mean of labor demand. Again the interpretation is that firms would in steady state be likely to be closer to the firing barrier than before.

The latter point, of course, raises the issue of how long it takes to reach the new steady-state MRP distribution, which leads to the study of the dynamics implied by the model.

Consider first the dynamic behavior of employment within a regime: firing costs are clearly not always harmful for employment. They reduce labor demand in good times (as the firm requires a higher marginal revenue product to start hiring), but <u>increase</u> labor demand in bad times. To illustrate this point, in Figure 7 employment histories of economies with different firing costs are plotted for the same exogenous demand path. The implication is that adjustment of employment to exogenous shocks


GAUSS Mar. 29, 1988 1:48:36 PM
Fig. 7 – Simulation
solid line: F=0
dots: F=0.08
dashes: F=0.8
employment
0.4
0.5
0.6
0.7
0.9
1.0
1.0
5.0
9.0
13.0
16.9
20.9
time (years)

is more sluggish in an economy with high firing costs.

From our results it should be clear that large firing costs do not decrease long-run average levels of employment: rather, the employment performance of economies where firing costs are large should be dynamically different. The results in Gordon[1987] provide some empirical support for this view: the ranking of the four countries we analyze in terms of the size of the labor input response to output movements is inversely related to their ranking in terms of firing cost magnitude. Moreover, the slow speed of adjustment of labor repeatedly found by researchers (see Nickell[1987]) may not be reflecting a smooth path originated by quadratic adjustment costs (as usually assumed) but the inaction of firms inside the corridor.

We can then discuss the dynamics implied by a regime change: what happens if a large, aggregate negative shock occurs, and at the same time the parameters of the stochastic process and the firing costs are suddenly and unexpectedly changed, as we think was the case in the early 1970s?

Assuming that by that time firms were distributed between the barriers according to the steady-state distribution (shaped like the ones plotted in Figure 3), most firms would be close to the hiring barrier. In the aftermath of the first oil price shock, the negative change in fundamentals (demand, wages) would shift the whole distribution towards the firing barrier, but only a very small number of firms would immediately fire, given the concentration near the hiring barrier and the higher firing cost.

It is not possible to verify this implication since dismissals are not distinguished from quits in official data, but casual empiricism suggests that labor shedding by large European firms was in fact very small during the 1970s. Thus, given a slowly-adjusting real wage, employment would had certainly been lower in Europe had high firing costs not prevented firms from shedding redundant workers.

Since the bulk of firms that used to be close to the hiring point before the oil shock were carried far from it by the negative change in fundamentals, and the hiring barrier itself was shifted away, in the ensuing years most firms would find their labor force too high and would just let attrition reduce it. Even though idiosyncratic uncertainty would be smoothing the distribution of firms within the new barriers, given the sharply reduced quit rate it took a long time to bring any firms close to the hiring barrier. Some firms, on the other hand, were firing. Changes in aggregate business conditions (demand and real wages) were in the meantime moving the center of the distribution between the new hiring and firing barriers; but the mild recovery of the 1980s was, in this framework, not large enough to bring a significant number of firms to the hiring barrier.

Taking a more general view, are firing costs to blame for the European malaise of the last fifteen years? Without a general equilibrium model, we clearly cannot draw conclusions about the welfare effects of firing costs. But labor demand is more stable if firing costs are large (see Figure 7), and given that wage setters may actually trade off lower wages for higher firing costs, a case could probably be made that firing costs are beneficial. If demand fluctuations are not due to smoothly functioning markets but to Keynesian coordination failures, it is at least conceivable that high firing costs could improve workers' welfare in a second-best world: they tend to increase employment in bad times, without appreciably decreasing its long-run average level. Of course, the lower flexibility of production decreases profits and the value of the firm, introducing income distribution issues if the firms' owners are in fact, as a group, distinct from the workers.

On the other hand, the microeconomic impact of firing regulations should not be neglected: a rigid employment structure hampers microeconomic efficiency, as it inhibits reallocation of labor in response to idiosyncratic demand and productivity shifts; inasmuch as such intersectoral shifts are not transitory, desirable employment stabilization has in effect to be traded off against undesirable productive inefficiency.[21] We recognize that such idiosyncratic uncertainty is important when we assume, above, that uncertainty at the firm level is three times larger than at the industry level; but we do not have a good aggregative model, and therefore cannot, for now, address the issue of allocative

[21] However, Piore[1986] notes that the very existence of dismissal restrictions can induce firms to search for more flexible uses of its labor force, thus attaining a higher degree of "dynamic" efficiency: firms will respond to disturbances by retraining and redeploying, rather than shedding, their labor force.

efficiency. Firing costs also hamper productive efficiency when employers are imperfectly informed about the quality of individual workers: in particular, European laws make it hard for employers to dismiss incompetent workers, who are on the other hand the less likely to quit. This realistic feature should be taken into account, and it would probably be found that firing would be more frequent and firms' reluctance to hire would rise.

5- Conclusions

This Chapter proposes an analysis of labor demand in the presence of linear firing (and possibly hiring) costs, taking explicit account of dynamics and uncertainty. In particular, an attempt is made to characterize the effects of lower expected growth and higher uncertainty in post-oil shock Europe on a typical firm's employment policy. We find that such effects are nonnegligible. We do not provide a a complete macroeconomic model, but our model suggests that dynamics and uncertainty must be taken into consideration when modeling the European unemployment problem: the highly regulated nature of European labor markets constrains the flexibility of a firm's employment policies in such ways that hiring a worker is definitely a risky proposition, and the degree of uncertainty about the future is a crucial parameter in the firm's problem.

We also find that the magnitude of firing costs affects the firing policy of the firm much more dramatically than its hiring policy. The effect of firing costs depends on the environment. The fall in growth and productivity and the increase in volatility made these costs more important. But as Table 4 forcefully shows, a mere reduction in firing costs does not significantly increase firms' marginal propensity to hire while it strongly raises their willingness to fire. This implication would seem to be confirmed by the bout of firing in the U.K. in 1980-82, after flexibility measures were put up starting in 1979 and by the lack of an appreciable increase in hiring after similar measures were established in Germany (1985) and France (1986). Therefore, while it is reasonable to credit firing costs for the avoidance of mass firings after the first oil shock, it does not seem accurate to blame them for the European lackluster employment performance in the 1980s. If it is granted that employment stabilization may be desirable due to macroeconomic distortions, we find that large firing costs afford more employment stability than small (but positive) ones, <u>without</u> appreciably affecting the long run level of employment.

Our model provides, we think, a useful framework for discussing the effects of institutional constraints on a firm's employment policies under uncertainty. Much theoretical and empirical work will be needed to precisely pin down the relevance of this Chapter's insight, to provide a more realistic model of the firm, and to extrapolate our results on individual firms to the macroeconomy. The set of techniques proposed in this Chapter can be used, for example, to study the economics of marginal employment subsidies programs (i.e. a reduction in hiring costs). The research agenda includes modeling less flexible productive structures for the firm,[22] uncertainty about wages and firing costs, capital/labor substitution with interrelated employment and investment policies, aggregation (with attention to idiosyncratic uncertainty), and especially considering the supply of as well as the demand for labor, either using a search-theoretic framework or modeling wage- and firing-cost setting as a bargaining process.

[22]For example, labor could be used in fixed proportions with materials in production: such a Leontief production function would induce the firm to retain idle workers in its payroll during a demand slump.

APPENDIX 1: Expectation of controlled geometric Brownian motion

Consider a regulated geometric Brownian motion with control barriers at u and ℓ, i.e. a stochastic process $\{\eta_t\}$ defined by

$$\eta_t = \frac{\zeta_t L_t}{U_t}$$

where:

(i) $\{\zeta_t\}$ is a geometric Brownian motion, with stochastic differential

$$d\zeta_t = \zeta_t m\, dt + \zeta_t s\, dW_t$$

(m and s are real constants, W_t is a standard Brownian motion);

(ii) $\{U_t\}$ and $\{L_t\}$ are increasing and continuous processes, with $L_0=U_0=1$;

(iii) $\{L_t\}$ only increases when $\eta_t=\ell$, and $\{U_t\}$ only increases when $\eta_t=u$, where u and ℓ are given positive real numbers;

(iv) $\ell \le \eta_t \le u$ for all $t\ge0$

Harrison[1985]'s arguments can be adapted to show that these four properties uniquely identify $\{U_t\}$ and $\{L_t\}$; these two processes maintain η_t within the barriers using the minimum amount of control, in a well-defined sense.

Let f(.) be a twice continuously differentiable function. Note that $\{U_t\}$ and $\{L_t\}$ are processes of finite variation, and apply Ito's lemma to obtain, after using property (iii) above,

$$df(\eta_t)= \left[m\, f'(\eta_t)\eta_t + \frac{s^2}{2} f''(\eta_t)\eta_t^2 \right] dt + s\, f'(\eta_t)\eta_t\, dW_t + \ell\, f'(\ell)\, \frac{dL_t}{L_t} - u\, f'(u)\, \frac{dU_t}{U_t}$$

Now recall the Integration By Parts formula found in Harrison[1985], page 73, and apply it to $\{ f(\eta_t)\, e^{-\lambda t} \}$ to obtain:

$$\text{(A1)}\quad e^{-\lambda t}f(\eta_t) = f(\eta_0) + \int_0^t e^{-\lambda \nu}\left[mf'(\eta_\nu)\eta_\nu + \frac{s^2}{2} f''(\eta_\nu)\eta_\nu^2 - \lambda f(\eta_\nu)\right]d\nu$$

$$+ \int_0^t e^{-\lambda\nu} s f'(\eta_\nu)\eta_\nu dW_t + \ell\, f'(\ell)\int_0^t e^{-\lambda\nu}\, \frac{dL_t}{L_t} - u\, f'(u)\int_0^t e^{-\lambda\nu}\, \frac{dU_t}{U_t}$$

Take the expectation at time 0 of (A1) and let $t \to \infty$; provided that $f(\eta_t)$ is bounded and the following conditions hold:

[*] $\qquad \ell\, f'(\ell) = 0$

[**] $\qquad u\, f'(u) = 0$

the result is

$$\text{(A2)} \quad 0 = f(\eta_0) + E_0\left\{\int_0^{\infty} e^{-\lambda\nu}\left[mf'(\eta_\nu)\eta_\nu + \frac{s^2}{2} f''(\eta_\nu)\eta_\nu^2 - \lambda f(\eta_\nu)\right]d\nu\right\}$$

Now if a function $f(.)$ is found such that

$$\text{(A3)} \quad -\eta_\nu = mf'(\eta_\nu)\eta_\nu + \frac{s^2}{2} f''(\eta_\nu)\eta_\nu^2 - \lambda f(\eta_\nu)$$

and [*], [**] are satisfied, then rearranging (A2) gives

$$\text{(A4)} \quad f(x) = E_0\left\{ \int_0^{\infty} e^{-\lambda t}\, \eta_t \, dt \;\middle|\; \eta_0 = x \right\}$$

The general solution to differential equation (A3) is

$$\text{(A5)} \quad f(\eta) = \frac{1}{\lambda - m}\left[\eta + B_1\, \eta^{a_1} + B_2\, \eta^{a_2}\right]$$

with

$$a_1 \equiv \frac{-\left(m - \frac{s^2}{2}\right) + \sqrt{\left(m - \frac{s^2}{2}\right)^2 + 2\, s^2\, \lambda}}{s^2} \quad > 0$$

$$a_2 \equiv \frac{-\left(m - \frac{s^2}{2}\right) - \sqrt{\left(m - \frac{s^2}{2}\right)^2 + 2\, s^2\, \lambda}}{s^2} \quad < 0$$

where B_1 and B_2 are constants of integration to be determined by the boundary conditions. Conditions [*] and [**] form a system of two linear equations in B_1 and B_2, with solution

$$\text{(A6)} \quad B_1 = \frac{u^{a_2}\ell - u\, \ell^{a_2}}{a_1\left[u^{a_1}\ell^{a_2} - u^{a_2}\ell^{a_1}\right]} \quad , \quad B_1 = \frac{u\, \ell^{a_1} - u^{a_1}\ell}{a_2\left[u^{a_1}\ell^{a_2} - u^{a_2}\ell^{a_1}\right]}$$

Using these values in (A5), and recalling (A4), the result stated in the main text is proved.

Now we turn to the steady-state distribution of the MRP. Harrison [1985, page 90] derives the steady state distribution for <u>linear</u> regulated Brownian motion: if $\{\xi_t\}$ is a Brownian motion with drift M and standard deviation Σ, regulated at 0 and b>0, i.e.:

$$0 \le \xi_t \le b \qquad \begin{array}{ll} d\xi_t = M\,dt + \Sigma\,dW_t & \text{if } 0 < \xi_t < b \\ d\xi_t = 0 & \text{if } \xi_t = 0 \text{ or } \xi_t = b \end{array}$$

then in steady-state it has the following (truncated exponential) cumulative distribution function:

$$F\left(\xi\right) = \begin{cases} 0 & \xi<0 \\ \dfrac{\exp\left(\dfrac{2M}{\Sigma^2}\,\xi\right) - 1}{\exp\left(\dfrac{2M}{\Sigma^2}\,b\right) - 1} & 0\le\xi\le b \\ 1 & b<\xi \end{cases}$$

Noting now that $\ln(\eta_t/\ell)$ follows a linear regulated Brownian motion with drift $M=m-s^2/2$ and standard deviation s, regulated at 0 and at $b=\ln(u\,/\,\ell)$, it is easy to derive the steady-state distribution of η_t:

$$\mathrm{Prob}\left[\eta\le x\right] = \mathrm{Prob}\left[\ln(\eta_t/\ell)\le \ln(x/\ell)\right] =$$

$$= \begin{cases} 0 & x < \ell \\ \dfrac{(x/\ell)^{[(2m/s^2)-1]} - 1}{(u/\ell)^{[(2m/s^2)-1]} - 1} & 0 \le \xi \le b \\ 1 & u < x \end{cases}$$

The steady-state density of η, the marginal revenue product, is then

$$f(\eta) = \frac{\left(\frac{2m}{s^2} - 1\right) \frac{1}{\ell} \eta^{\left(\frac{2m}{s^2} - 2\right)}}{u^{\left(\frac{2m}{s^2} - 1\right)} - \ell^{\left(\frac{2m}{s^2} - 1\right)}} \; 1\left[\ell \le \eta \le u \right]$$

and simple integration gives the following expression for the ergodic mean of the marginal revenue product process:

$$\bar{\eta} = \left(\frac{m - s^2/2}{m}\right) \frac{u^{\left(\frac{2m}{s^2}\right)} - \ell^{\left(\frac{2m}{s^2}\right)}}{u^{\left(\frac{2m}{s^2} - 1\right)} - \ell^{\left(\frac{2m}{s^2} - 1\right)}}$$

APPENDIX 2: <u>Derivation of equation [4c']</u>

First rewrite equation [4c] in the text as:

$$\text{(A7)} \quad E_t\left\{\int_t^T \left[\mu A_\tau \left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-t)} \, d\tau \right\} + E_t\left\{ e^{-(r+\delta)(T-t)} \int_T^\infty \left[\mu A_\tau \left(\frac{z_\tau}{A_t L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-T)} \, d\tau \right\} = \frac{w}{r+\delta}$$

if $dX_t > 0$

For the next step, recall the Strong Markov Property of (controlled) Brownian motions (see e.g. Harrison[1985]):
<u>Strong Markov Property</u>: Let $\{z_t\}$ be a (regulated, geometric) Brownian motion process and let T be a stopping time (the Brownian motion attains for the first time a predetermined value at T). Then the random variable T and the stochastic process $\{z;t>T\}$ are independent.

It is then possible to separate the second term on the left hand side in (A7) into the product of two expectations; rearranging and taking iterated expectations (recall that t<T) we get:

$$\text{(A8)} \quad E_t\left\{\int_t^T \left[\mu A_\tau \left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-t)} \, d\tau \right\} =$$

$$= \frac{w}{r+\delta} - E_t\left\{e^{-(r+\delta)(T-t)}\right\} E_t\left\{E_T\left\{\int_T^\infty \left[\mu A_\tau \left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-T)} \, d\tau \right\}\right\}$$

Since T is by assumption a firing time, by eq. [4a] in the main text we get:

$$\text{(A9)} \quad E_T\left\{\int_T^\infty \left[\mu A_\tau \left(\frac{z_\tau}{A_\tau L_\tau}\right)^{1-\mu}\right] e^{-(r+\delta)(\tau-T)} \, d\tau \right\} = \frac{w}{r+\delta} - F$$

dX <0

Substituting the (nonstochastic) right hand side of (A9) into (A8), [4c] is finally reduced to [4c'] in the text.

APPENDIX 3: Parameter values for the solution baseline

Here we explain the procedures and sources for the baseline parameter values for the solution of the model in the text.

A) Parameters other than hiring and firing costs:

ϑ: Average percentage change in the first differences of the logarithm of the index of industrial production (IFS tape).

σ: Three times the average standard deviation of the first differences of the logarithm of the index of industrial production (International Financial Statistics (IFS) tape).

δ: France (manufacturing) and Germany (whole economy): Two thirds and one half of the separations rate (for the years available) in regimes 1 and 2, respectively (to capture the procyclical behavior of quits); from OECD[1986b]. Italy (industry; after 1976 firms with at least 50 employees): Dell'Aringa[1987]. U.K.(whole economy): Burgess and Nickell[1987].

ϑ_a: Average difference between the change in the first differences of the logarithm of the index of industrial production and the logarithm of the index of industrial employment (IFS tape, for the U.K. the source is OECD[1987b]).

r: Required rate of return on capital or profit rate. Set to 15% for both regimes.

μ: Inverse of the markup factor. Set to 0.6, which is the average of the estimates of this parameter in Burda[1987] for the four countries we are dealing with.

B) Hiring costs: We only have data for training costs. Calculations on data in Nollen[1987] give a maximum value of 6.6% of the average annual wage as the average training cost in Germany and 5.5% in the U.K., the latter not including on-the-job-training nor wages lost. We set H to 0.083 (1 months' pay).

C) Firing costs:

a) Notice period: The laws require a number of days per year of service (p.y.o.s.) with the firm. In terms of the cost, we equate one month of notice with one month of wage. Since the worker will be producing during the notice period this is an upper bound, equal to the payment *in lieu* of notice that can be made.

b) Expected dismissal cost: This is equal to:

$$F = N + (1-P_a)SP + P_a \{(1-P_u)(SP+LC) + P_u(UP+LC)\}$$

where N is the notice cost, P_a is the probability of the dismissal being appealed to the labor courts, P_u is the probability that the dismissal is ruled unfair, SP is the severance payment, LC are legal costs from going to court and UP is the payment for an unfair dismissal (we ignore the rare cases where reinstatement of the worker is mandated). We only have information on legal costs for the U.K.: Daniel and Stilgoe[1978] quote data equivalent to

1.8 to 5.6 months pay for these. We take 2 months' pay for all countries.

The estimates for the parameter values are given in Table A1, in the following page.

TABLE A1

	FRANCE	GERMANY	ITALY	UNITED KINGDOM
NP	2 months Indiv. dismiss. period= 2 months Collect. redund. consultation: 1-3 months	1.5 months Minimum for white collar workers with up to 5 years of service	3 months White collar: Consult.1/2-4 mo. Governm. 1 month	1.5 months 1 week p.y.o.s. Tenure profile for 1984 from from OECD[1986b] Collect.: 3 mths
SP	9 months' pay Law:0.675 of wage Improv.=10% mth's pay p.y.o.s. from Reynaud-C.[1986] Tenure profile in OECD[1987a]-1978	7 months' pay Mean of the distribution in Ochs[1976]	9.5 months' pay 1 month p.y.o.s. Tenure profile for 1978 from OECD[1986b]	1.5 months' pay Nickell[1979]: 4-5 weeks 1969-77
P_a	5%	6.6% Gennard[1985]: Prob[contested/ works council exists]=.1 (1977) Falke *et al*[1981] Workers covered by a works council = 66% (1978)	4.25% Gennard[1985] for 1978-80	9.23% Gennard[1985] for 1983
P_u	25%	40% Sengenberger [1985]:Cases where no compromise was reached (1981)	24% Gennard[1985]: Cases not dismissed:40% Dismissals declared unfair: 59% (1978-80)	11% EIRR[1977]: 8.18% in 1976 Gennard[1985]: 11.06% in 1983
UP	1 year	10 months' pay Emerson[1987]: 1 month p.y.o.s. Avg. tenure(1978) =10 years from OECD[1986b]	1 year's pay Emerson[1987]: not less than 5 months' pay	2 months' pay Nickell[1979]: 6.75 weeks'pay Mean for 1972-77

REFERENCES

Abel, Andrew B. and Olivier J. Blanchard[1986], "The Present Value of Profits and Cyclical Movements in Investment", Econometrica 1986, pp.249-273

Abel, Andrew B.[1983], "Optimal Investment Under Uncertainty", American Economic Review 1983, pp. 228-233

Abel, Andrew B.[1985], "A Stochastic Model of Investment, Marginal Q and the Market Value of the Firm", International Economic Review 1985, pp.305-322

Ando, Albert, Franco Modigliani, Robert Rasche and Stephen Turnovsky[1974], "On the Role of Expectations of Price and Technological Change in an Investment Function", International Economic Review 1974, pp. 384-414

Arrow, Kenneth 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 University Press 1968.

Baldwin, Richard and Paul Krugman[1986], "Persistent Trade Effects of Large Exchange Rate Shocks", NBER Working Paper #2017, 1986

Beneŝ, V.E., L.A. Shepp and H.S. Witsenhausen[1980], "Some Solvable Stochastic Control Problems", Stochastics 1980, pp. 39-83

Bentolila, S.[1987]: "The Labor Demand Side of European Unemployment: The Role of Adjustment Costs", mimeo, MIT, May.

Bernanke, Ben[1983], "Irreversibility, Uncertainty and Cyclical Investment", The Quarterly Journal of Economics 1983, pp. 85-106.

Blanchard, O., R. Dornbusch, J. Drèze, H. Giersch, R. Layard, and M. Monti: "Employment and Growth in Europe: A Two Handed Approach" in Blanchard, O. *et al.*(eds.): Restoring Europe's Prosperity, The MIT Press, Cambridge, Massachusetts.

Blanchard, O. and L. Summers[1986]: "Hysteresis and the European Unemployment Problem", in Fischer, S.(ed.): NBER Macroeconomics Annual 1986, The MIT Press, Cambridge, Massachusetts.

Burda, M.C.[1987]: "Monopolistic Competition, Costs of Adjustment and the Behavior of European Unemployment", unpublished Ph.D. dissertation, Harvard University.

Burgess, S.M. and J.J. Dolado[1987]: "Intertemporal Rules with Variable Speed of Adjustment: An Application to U.K. Manufacturing Employment", Applied Economics Discussion Paper no. 32, Institute of Economics and Statistics, Oxford University, September.

Burgess, S.M. and S.J. Nickell[1987]: "Turnover in U.K. Manufacturing", mimeo, Institute of Economics and Statistics, Oxford University.

Caplin, A. and K. Krishna[1986]: "A Simple Dynamic Model of Employment", mimeo. Harvard University.

Chow, Pao-Liu, José Luis Menaldi and Maurice Robin[1985], "Additive Control of Stochastic Flow Systems with Finite Horizon", SIAM Journal of Control and Optimization 1985, pp.858-899

Contini, B. and R. Revelli[1987]: "A Theoretical Model of Job Creation and Job Destruction", Consiglio Nazionale Delle Ricerche, Working Paper 7/RP, July.

Cox, D.R. and H.D. Miller[1965], The Theory of Stochastic Processes, Chapman and Hall 1965.

Daniel, W.W. and E. Stilgoe[1978]: The Impact of Employment Protection Laws, Policy Studies Institute, London.

Dell'Aringa, C.[1986]: "Labour Market Flexibility: The Case of Italy", International Symposium on Labour Market Flexibility, Geneva.

Dietrich, J.Kimball and Donald G.Heckerman[1980], "Uncertain Inflation and the Demand for Capital", Economic Inquiry 1980, pp.461-471

Dixit, Avinash[1987], "Entry and Exit Decisions of Firms under Fluctuating Real Exchange Rates", mimeo, 1987

Dornbusch, R.[1986]: "Unemployment: Europe's Challenge of the '80s", Challenge, September/October.

Dornbusch, R.[1987]: "Open Economy Macroeconomics: New Directions", National Bureau of Economic Research working paper no. 2372, August.

Emerson, M.[1987]: "Regulation or Deregulation of the Labour Market: Policy Regimes for the Recruitment and Dismissal of Employees in the Industrialised Countries", Commission of the European Communities, Economic Paper no. 55, June.

European Economy [1986]: "Employment Problems: Views of Businessment and the Workforce", contribution by G. Nerb, March.

European Industrial Relations Review (E.I.R.R.): 1977: "Unfair Dismissal Statistics", March.

Falke *et al.*[1981]: "Kundigungspraxis und Kundigungsschutz in der Bundesrepublik Deutschland", Forschungsbericht, No.47, Bd. 1 und 2, Der Bundesminister fur Arbeit und Sozialordnung, Bonn.

Flanagan, R.J.[1987]: "Labor Market Behavior and European Economic Growth", in Lawrence, R.Z. and Ch.L. Schultze (eds.): Barriers to European Growth. A Transatlantic View, The Brookings Institution, Washington D.C.

Gavin, M.K.[1986]: "Labor Market Rigidities and Unemployment. The Case of Severance Costs", International Finance Discussion Paper no. 184, Washington, D.C.

Gennard, J.[1985]: "Job Security: Redundancy Arrangements and Practices in Selected OECD Countries", mimeo, OECD, Paris, September.

Giersch, H.[1985]: "Eurosclerosis", Kiel Discussion Papers no. 112, Kiel.

Gordon, Robert J.[1987], "Wage Gaps vs. Output Gaps: Is There a Common Story for all of Europe?", National Bureau of Economic Research Working Paper No. 2454, December.

Hall, Robert E.[1977], "Investment, Interest Rates, and the

Effects of Stabilization Policies", Brookings Papers on Economic Activity 1:1977, pp.61-121

Hall, Robert E., "Investment under Uncertainty: Theory and Tests using Industry Data", mimeo 1987

Hamermesh, D.S.[1987]: "Labor Demand and the Structure of Adjustment Costs", mimeo, Michigan State University, October.

Harrison, J. Michael[1985], "Brownian Motion and Stochastic Flow Systems", John Wiley & Sons, 1985

Harrison, J.M., Th.M. Sellke and A.J. Taylor[1983]:"Impulse Control of Brownian Motion", Mathematics of Operations Research.

Hayashi, Fumio[1982], "Tobin's Marginal q and Average q: A Neoclassical Interpretation", Econometrica 1982, pp.213-224

Holt, Ch.C., F. Modigliani, J.F. Muth and H.A. Simon: Planning Production, Inventories, and Work Force, Prentice-Hall, Inc., New Jersey.

Hoshi, Takeo and Anil K Kashyap[1987], "Evidence on Q for Japanese Firms", mimeo MIT 1987

Ingersoll, Jonathan E. Jr, and Stephen A. Ross, "Waiting to Invest: Investment and Uncertainty", mimeo Yale 1987

Jorgenson, Dale W.[1963], "Capital Theory and Investment Behavior", American Economic Review Papers and Proceedings 1963, pp. 247-259

Karatzas, Ioannis and Steven E. Shreve[1984], "Connections between Optimal Stopping and Singular Stochastic Control I:

Monotone Follower Problem", SIAM Journal of Control and Optimization 1984, pp.856-877

Karatzas, Ioannis and Steven E. Shreve[1985], "Connections between Optimal Stopping and Singular Stochastic Control II: Reflected Follower Problems", SIAM Journal of Control and Optimization 1985, pp. 433-451

Karatzas, Ioannis and Steven E. Shreve[1988], "Brownian Motion and Stochastic Calculus", Springer-Verlag 1988

Kelsey, D.[1986]: "Labour Demand with Transactions Costs", mimeo, Churchill College, May.

Kemp, M.C. and H.Y. Wan, Jr.[1974]: "Hysteresis of Long-run Equilibrium from Realistic Adjustment Costs" in Horwich, G. and P.A. Samuelson (eds.): Trade, Stability and Macroeconomics. Essays in Honor of Lloyd A. Metzler, Academic Press, New York.

Kennan, J.[1979]: "The Estimation of Partial Adjustment Models with Rational Expectations", Econometrica, November.

Krugman, P.R.[1987]: "Slow Growth in Europe: Conceptual Issues", in Lawrence, R.Z. and Ch.L. Schultze (eds.): Barriers to European Growth. A Transatlantic View, The Brookings Institution, Washington D.C.

Lazear, E.P.[1987]: "Job Security and Unemployment", mimeo, University of Chicago, May.

Leban, R. and J. Lesourne[1980]: "The Firm's Investment and Employment Policy Through a Business Cycle", European Economic Review.

Leijonhufvud, Axel[1973], "Effective Demand Failures", Swedish Journal of Economics.

Lucas, Robert E. Jr and Edward C. Prescott[1971], "Investment Under Uncertainty", Econometrica 1971, pp. 659-681

McDonald, Robert and Daniel Siegel[1985], "Investment and the Valuation of Firms when there is an Option to Shut Down", International Economic Review 1985, pp. 331-349

McDonald, Robert and Daniel Siegel[1986], "The Value of Waiting to Invest", The Quarterly Journal of Economics 1986, pp.707-727

Majd, Saman and Robert S. Pindyck[1987], "Time to Build, Option Value, and Investment Decisions", Journal of Financial Economics 1987, pp.7-27

Merton, Robert C.[1973], "Theory of Rational Option Pricing", Bell Journal of Economics 1973, pp.141-183

Nelson, Daniel[1987], "Conditional Heteroskedasticity in Asset Returns: A New Approach", mimeo, 1987

Nickell, Stephen[1974], "On the Role of Expectations in the Pure Theory of Investment", Review of Economic Studies 1974, pp. 1-19

Nickell, S.J.[1978]: "Fixed Costs, Employment and Labor Demand Over the Cycle", Economica.

--------- [1979]: "Unemployment and the Structure of Labor Costs", Carnegie-Rochester Conference Series on Public Policy, Supplement to the Journal of Monetary Economics, no. 11.

--------- [1987]: "Dynamic Models of Labour Demand" in Ashenfelter

O. and R. Layard (eds.): Handbook of Labor Economics, vol. 2, North Holland.

Nollen, S.D.[1987]: "The Training Market in the United States, Britain and Germany", mimeo, Georgetown University, January.

Ochs[1976]: Analyse betrieblicher Sozialplane, Saarbrucken.

Oi, W.[1962]: "Labor as a Quasi-Fixed Factor", Journal of Political Economy.

OECD[1985]: Employment Outlook, Paris.

OECD[1986a]:Employment Outlook, Paris.

OECD[1986b]: Flexibility in the Labor Market. The Current Debate, Paris.

OECD[1987a]: Economic Outlook, Paris, December.

OECD[1987b]: Labor Force Statistics, Paris.

Pindyck R.S. and J.J. Rotemberg[1983]: "Dynamic Factor Demands and the Effect of Energy Price Shocks", American Economic Review, December.

Pindyck, Robert S.[1986], "Irreversible Investment, Capacity Choice, and the Value of the Firm", NBER Working Paper #1980, 1986

Piore, M.[1986]: "Perspectives on Labor Market Flexibility", Industrial Relations, Spring.

Prescott, Edward C.[1973], "Market Structure and Monopoly Profits:A Dynamic Theory", Journal of Economic Theory, 6 (1973), pp.546-557

Reynaud-Cressent, B.[1986]: "Diversite de la Relation Salariale de

Branches et Codification Des Conventions Collectives", in R. Salais et L. Thevenot (eds.): Le Travail. Marches, Regles et Conventions, Paris, Insee-Economica.

Romer, Ch.[1987]: "Changes in the Cyclical Behavior of Individual Production Series", National Bureau of Economic Research working paper no., September.

Sargent, Th.[1978]: "Estimation of Dynamic Labor Demand Schedules under Rational Expectations", Journal of Political Economy, December.

Sargent, Thomas J.[1979a], "Macroeconomic Theory", Academic Press 1979

Sargent, Thomas J.[1979b], "Tobin's Q and the Rate of Investment in General Equilibrium", in On the State of MAcroeconomics, edited by K.Brunner and A.H.Metzler, North Holland 1979

Sengenberger, W.[1985]: "Labor Market Flexibility in Germany", mimeo, OECD, Paris.

---------- and G. Loveman[1987]: "Smaller Units of Employment. A Synthesis Report on Industrial Reorganisation in Industrialised Countries", Discussion Paper 3/87, International Institute for Labor Studies.

Shapiro, M.[1986]: "The Dynamic Demand for Capital and Labor", Quarterly Journal of Economics, August.

Solow, R.M.[1968]: "Short run adjustment of Employment to Output" in J.N. Wolfe (ed): Value, Capital and Growth: Papers in Honor of Sir John Hicks, Edinburgh University Press.