ESSAYS ON FINANCIAL THEORY

by

Sudipto Bhattacharya

B.Sc. (Honors) Physics, University of Delhi
(1971)

P.G.D.B.A., Indian Institute of Management, Ahmedabad
(1973)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(September, 1977)

(i.e. February 1978)

Signature of Author ....
Department of Management, September, 1977

Certified by ....
Thesis Supervisor

Accepted by ....

ESSAYS ON FINANCIAL THEORY

by
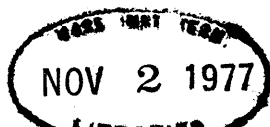
Sudipto Bhattacharya

Submitted to the Alfred P. Sloan School of Management, M.I.T.
on September 16, 1977 in partial fulfillment
of the requirements for the Degree of Doctor of Philosophy.

ABSTRACT

This dissertation consists of six essays on financial theory. The
first essay, "The Welfare and Savings Effects of Indexation," deals with
the welfare consequences of the introduction of indexed debt contracts and
its effects on economy-wide savings. The second essay, "Discrete and Con-
tinuous-Time Valuation of Mean Reverting Cash Flow Streams," extends the
literature on the pricing of multi-period cash flow streams using the Capi-
tal Asset Pricing Model to two mean-reverting cash flow processes, and
investigates the accuracy of traditional methods of asset valuation like
the "cost of capital" rule. The third essay, "Stock Price Jumps and Coven-
ant-Induced Safety Barriers: the Perpetuity Case" extends the pricing of
risky debt contracts to jump processes for the firm value when debt coven-
ants specify a 'safety barrier' - if firm value falls below such a barrier
then debtholders take over the firm. The fourth and fifth essays, "Imper-
fect Information, Dividend Policy, and the Bird in the Hand Fallacy" and
"Notes on a Non-Dissipative Signalling Structure" deal with models in which
imperfect information, in the form of investors' inability to distinguish
among firms of different profitability 'a priori', plays a crucial role.
The fourth essay develops a model in which the payment of cash dividends

serves as a 'signal' of project value - the signalling cost arises from the higher personal taxes on dividends compared to capital gains - and investigates the comparative statics of the dividend payout with respect to the personal tax rate , the rate of interest, and the investors' planning horizon. The fifth essay develops a model in which signalling occurs not through exogenously costly signals but through zero-sum incentive structures. The model is developed in the context of a labour market in which workers' mean productivities differ. Possible extensions to dividend policy in a world without taxes are investigated, and an interesting difficulty related to a conflict with the assumptions of rational expectations and freedom of liquidation time is pointed out. The sixth essay, "Diffusion of Technological Innovations: a Sequential Experimentation Approach" uses a model of sequential investment decision-making by firms to re-examine the empirical evidence on the characteristics of the diffusion of technological innovations. The model is shown to imply diffusion characteristics that have, in the past, been explained by a hypothesis of "imitative behavior." It also throws new light on the comparative statics of the diffusion process with respect to the parameters of the innovation, such as its mean profitability, the rate of resolution of uncertainty about profitability, and so forth.

Thesis Supervisor: Professor Stewart C. Myers
                    Professor of Finance

-4-

Acknowledgements

I wish to thank my adviser, Professor Stewart C. Myers and Professor Robert C. Merton for their help, advice and constructive criticism. Professor Richard Cohn is to be thanked for reading the dissertation and for his constructive comments.

I would like to thank all my teachers from whom I have tried to learn finance and economics. I have benefitted from discussions with Professors Carliss Baldwin and Fischer Black and from a roomfull of colleagues at E52-201 - the arguments, polemics and gossip exchanged in that office will be long remembered.

Stewart C. Myers is to be thanked for his unfailing graciousness and Maureen Myers for her warm hospitality and for my introduction to the pleasures of the mince pie! It was a pleasure to have worked with Professor Franco Modigliani on teaching a course. I thank my parents for their moral support. Chris, Vicki, Sharon, Diane, and Jean are to be thanked for their help in preparing the manuscript.

Most of all, I hope that this dissertation marks the begining of professional association with a group of colleagues at M.I.T. and elsewhere whose work it has been a pleasure to learn about and appreciate.
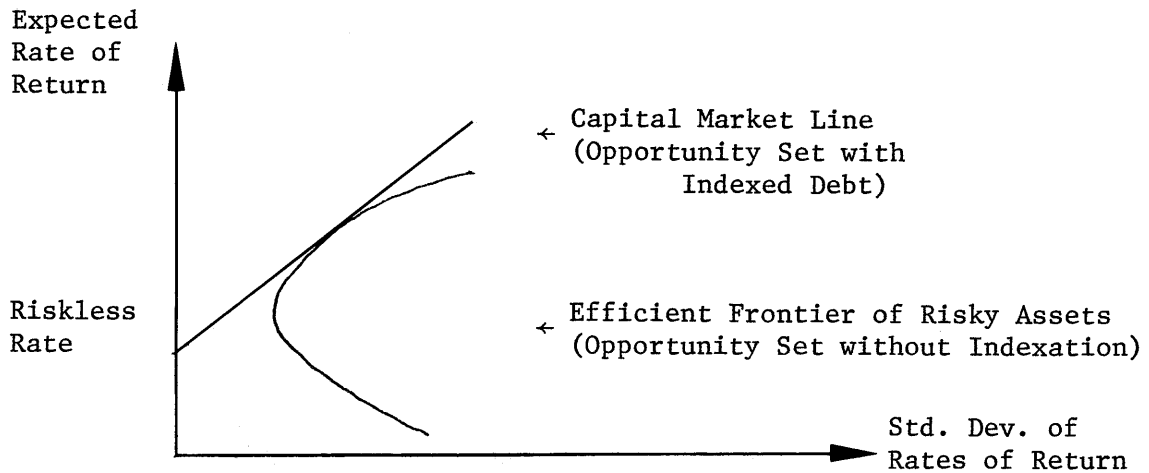
LIST OF CONTENTS .

THE WELFARE AND SAVINGS EFFECTS OF INDEXATION

Sudipto Bhattacharya

I.   The aims of this paper are twofold. First, we look at the welfare

implications of the introduction of indexed debt contracts and demonstrate

that the literature that claims to disprove the Pareto-superiority of the

introduction of indexed debt contracts is methodologically incorrect and

that the basis for any doubts about such Pareto-superiority is subtler.

Second, we investigate the impact of indexation of debt contracts on the

level of saving and investment in the economy and claim to provide a sim-

ple "general equilibrium" answer to this question. Our analysis is car-

ried out in the framework of a two-period mean-variance model (of portfo-

lio selection) where the inflation rate is taken to be exogenous and un-

certain, and the usual assumption about homogeneity of expectations is

made.

In the first part, the following are given. (a) The economy in

question is a monetized one, with a single real commodity. The economy is

in equilibrium with no indexed and, hence, certain debt contracts avail-

able. (b) Investors start introducing indexed debt contracts, and the

economy moves to a new equilibrium. It is assumed in the first part that

investment in real (stochastic) production technologies is kept unaltered.

The emphasis is on the private introduction of indexed debt contracts

since, as Levhari and Liviatan (4) point out, models wherein the govern-

ment or some other entity starts substituting indexed for non-indexed debt

and no other portfolio changes take place, run into problems (in the sense

of the action making any real economic difference) as soon as one consid-

ers the changed interest payment obligations that the government incurs.

In what follows we also assume that the net holding of indexed or non-

indexed debt, summed across all agents, is zero.

Sarnat (8) first demonstrated that the introduction of indexed debt would make every investor better off since it would expand the investment opportunity set of every investor as in the diagram below.

```
Expected
Rate of
Return      ↑
            │                          ⟋
            │                        ⟋    ← Capital Market Line
            │                      ⟋          (Opportunity Set with
            │                    ⟋⟍               Indexed Debt)
            │                  ⟋   ⟍
            │                ⟋      ⟍
            │              ⟋         ⟍
Riskless    │            ⟋            │
Rate        │          ⟋              │     ← Efficient Frontier of Risky Assets
            ├        ⟋               ⟋          (Opportunity Set without Indexation)
            │      ⟍              ⟋
            │        ⟍         ⟋
            │          ⟍     ⟋            Std. Dev. of
            │            ⟍ ⟋          ➤  Rates of Return
            └──────────────────────────────────────
```

Fischer (3) in a brief remark, and Subrahmanyam (9) pointed out that, in this form, the argument was false since in general the introduction of a new financial asset would change all rates of return, through changes in the prices for the same next-period cash flows, and since the new efficient frontier of risky assets could lie below the old one in parts, they argued that everyone would not be better off. Recently, Ragazzi and Bicksler and Hess (8a) have also commented on this problem. The following quote from Ragazzi provides an example of the "typical" contention of this strand of the literature. "In order to move from the original investment locus to the market line, . . . each investor would have to sell the combination originally held. The increased demand for asset N (the market portfolio) can not but push up its price and lower its rate of return. Hence investors who previously held portfolio N . . . would be worse off."

Obviously, an argument of the above type is itself incomplete, since the change in prices also produces wealth effects, and thus the
-8-

analysis of welfare effects cannot be done in the space of rates of return alone. One other analysis of this issue is also discussed, only to provide an illustration of an alternative approach to the problem. Turnbull (10) does an intertemporal, continuous-time analysis of the problem and concludes that, under some assumptions about partial homogeneity of utility functions across investors, the lenders would benefit and the borrowers would become worse off after indexation. However, though rates of return are endogenous in the model, the assumption made in doing the welfare analysis is that the value of the market portfolio is the same in the two equilibria. This assumption has no logical foundation. Although Subrahmanyam (9) claims to prove this, the proof is flawed by a simple circularity. We set up the notation and discuss this before proceeding to our own conclusions. Let,

$$\tilde{D}_j = \text{end-of-period cash flow of the j-th firm}$$

$$\sum_j \tilde{D}_j = \tilde{D}_M = \text{end-of-period cash flow of the market portfolio}$$

$$\sigma_{jM} = \text{Covariance}(D_j, D_M)$$

R_L = one plus the rate of return on indexed debt in the new equilibrium

R_B = one plus the nominal rate of interest on non-indexed debt in old equilibrium

$\tilde{I}$ = inflation rate, assumed to be exogenous

V_j = value of the j-th firm in the non-indexed equilibrium

V_j' = value of the j-th firm in the indexed equilibrium

$\overline{R}_Z$ = expected real rate of return on the minimum-variance zero-beta portfolio in the non-indexed equilibrium

All variables except R_B are measured in real terms. The capital asset pricing model and its extension by Black (2) give us the following two

relationships directly.

$$V_j' = \frac{1}{R_L} [\overline{D}_j - \left\{ \frac{\overline{D}_M - R_L V_M'}{\sigma_M^2} \right\} \sigma_{jM}]$$ (1a)

$$V_j = \frac{1}{R_Z} [\overline{D}_j - \left\{ \frac{\overline{D}_M - R_Z V_M}{\sigma_M^2} \right\} \sigma_{jM}]$$ (1b)

In his derivation of $(V_j' - V_j)$ Subrahmanyam (9) cancels $\frac{V_M' \sigma_{jM}}{\sigma_M^2}$ with

$\frac{V_M \sigma_{jM}}{\sigma_M^2}$ to arrive at a formula that, naturally enough, has the character-

istic that, when j is the market portfolio, the difference equals zero.

Of course, from a relative asset pricing model like the CAPM, one cannot

get an idea of how asset values change between equilibria in different

"asset economies."

The correct answer to the welfare question goes along the following

lines. Let investor i's holdings in the old equilibrium be a fraction C_i

of the market portfolio, and a fraction E_i of the minimum-variance zero-

beta portfolio which is, without loss of generality, scaled to have an

expected cash flow of unity in the next period. Of course, in the aggre-

gate,

$$\sum_i C_i = 1$$ (2a)

$$\sum_i E_i = 0$$ (2b)

(When there are more than two assets, such a characterization of holdings

is, of course, non trivial -- it follows from the model of Black (2).)

-10-

Let investors privately introduce an indexed debt instrument and proceed

to a new equilibrium, characterized by a new set of consumption-investment

programs consistent with the same aggregate investment, and a new set of

prices -- in particular P_z for the zero-beta portfolio and P_L for the in-

dexed debt, where $P_L = 1/R_L$ by definition. From the model of Mossin (5),

we know that $P_z = P_L$. One feasible strategy for any investor would have

been to exchange his zero-beta portfolio holdings for the indexed debt,

thus keeping his current and expected next-period cash flow the same and

reducing the variance of next-period cash flows by $E_k^2 \sigma_z^2$. Since his

actual holding in the new equilibrium is utility maximizing given the

prices and the resultant budget set, it must be true that (a) any investor

who held anything besides the market portfolio in the old equilibrium is

strictly better off in the new one, and (b) every other investor is at

least as well off. Given that a new financial asset in zero net supply is

being introduced, the result is hardly surprising.

The questions that remain about the welfare effect are two-fold.

First, we have not shown that if, starting with non-equilibrium alloca-

tions, investors proceed to equilibrium with and without the indexed debt,

then everyone is better off in the first equilibrium. Second, it is pos-

sible that, in an explicit mult-period model, the changes in asset prices

due to indexation would restrict subsequent portfolio revision opportuni-

ties. The first point does not seem possible to demonstrate, but its

ramifications are rather beside the point.[1] The second problem, unfortu-

---

[1] One cannot obtain a simple answer to the question of which group of in-
vestors becomes worse off, if any do, unlike in Ng's (6) analysis of the
effects of changes in (homogenous) beliefs in an exchange economy. The
essential difference between the two situations is that if indexation is
introduced at an equilibrium of the non-indexed economy, then, in general,
some people are strictly better off and others no worse off, whereas in

nately, cannot be analyzed with current models, since we have no multi-period equilibrium models that allow for all of (a) homogeneous expectations, (b) necessity of portfolio revision opportunities, (c) endogenized intertemporal price paths, (d) mean-variance based portfolio selection.

II.

In this section we deal with the effects of indexation on aggregate savings and investment. Other than making aggregate investment endogenous, all the other "perfect market type" assumptions of the previous section are retained. This analysis, therefore, does not apply to more bureaucratically controlled capital markets, where indexation seems to bring with it the liberalization of controls on real rates of interest allowed -- an example of which is provided by the study done in Barro (1).

We analyze a simple case in which the market portfolio consists of a single stochastic-constant-returns-to-scale technology and seek to explore sufficient conditions for aggregate investment to rise or decline. Levhari and Liviatan (4) have also attempted to look at this issue, but they carry out their analysis assuming quadratic utility functions and unchanged market prices and rates of return in the new equilibrium. As the authors acknowledge, their analysis does not provide a clear-cut answer to the question of what happens in general equilibrium when all prices are endogenous.

As it turns out, the general equilibrium analysis is fairly simple

---

the situation dealt with by Ng (6), changes in homogeneous beliefs at any equilibrium keep the allocations at the resulting equilibrium totally unchanged, and any gains or losses in expected utility are strictly "windfall gains" due to changes in subjective probabilities alone.

in this particular case. Let (A) represent the allocation of current con-
sumption, expected next-period consumption, and the variance of next-
period consumption in the old non-indexed equilibrium. Since investment
is endogenous, and there is only one (stochastic) technology, it follows
that at this allocation every investor is indifferent at the margin to an
increment in his savings and investment. In the previous section we
showed that with the introduction of indexation there exists a feasible
market allocation (B) at which every investor has a lower variance of con-
sumption than at allocation (A), with everything else remaining the same
-- if we assume that no investor held just the market portfolio (with no
borrowing or lending) in the old equilibrium. Of course, the (hypotheti-
cal) new equilibrium without any change in aggregate investment is not
necessarily allocation (B). Consider however a "thought experiment" in
which aggregate investment is increased by a very small amount I at allo-
cation (B) by taking $C_i^* I$ from the i-th investor -- where C_i, as defined
in the previous section, is his "share" of the market portfolio. We note
that (a) as long as I is sufficiently small, it does not make infeasible
any set of reallocations that represents the difference between allocation
(B) and the hypothetical new equilibrium without any change in aggregate
investment. Hence, if every investor prefers to have this increment to
aggregate investment take place at allocation (B), then by the Pareto-
optimality of the capital market with a riskless asset in a mean-variance
world, a marginal increase in aggregate investment must have taken place
in the indexed equilibrium. An exactly similar argument applies to dis-
investment.

Now a marginal increase in savings that is invested in the (market)
technology has the same incremental effect on the variance of second-

period consumption at allocations (A) and (B) -- but not the same incremental effect on the standard deviation of second-period consumption -- as may be easily verified. Let $U(C_0, \overline{C}_1, X)$ represent an individual investor's (derived) utility function in terms of current consumption, expected next-period consumption, and the variance of next-period consumption. Let P be the mean unit productivity and σ^2 the incremental variance of second-period consumption per unit of investment. Clearly, preference for investment is determined by the criterion,

$$\left[-\frac{\partial u}{\partial C_0} + P \frac{\partial u}{\partial \overline{C}_1} + \sigma^2 \frac{\partial u}{\partial X} \right] \tag{3}$$

whose value is zero at alocation (A). Sufficient conditions for more investment to be preferred at allocation (B) are that

$$\text{(i)} \quad \frac{\partial^2 U}{\partial X^2} \leq 0 \quad \text{and} \quad \text{(ii)} \quad \frac{\partial^2 U}{\partial \overline{C}_1 \partial X} \leq 0 \tag{4}$$

both hold, with at least one holding as a strict inequality. If the investor's "underlying" utility function of consumption is quadratic, then both of the above quantities are zero. Hence in this case it is ambiguous as to what effect indexation would have on aggregate savings and investment. More generally if the utility of consumption is temporarily additive, and returns are normally distributed, then we may write

$$U(C_0, \overline{C}_1, X) = V_0(C_0) + \int_{-\infty}^{\infty} V(\overline{C}_1 + \sqrt{x}S) f(S) dS \tag{5}$$

where S is the standard normal variate and f(S) is its density function. Then

$$\frac{\partial U}{\partial \overline{C}_1} = \int_{-\infty}^{\infty} V'(\overline{C}_1 + \sqrt{x}S)\ f(S)dS \quad (6a) \qquad \frac{\partial U}{\partial X} = \frac{1}{2\sqrt{X}}\int_{-\infty}^{\infty} SV'(\overline{C}_1 + \sqrt{x}S)f(S)dS \quad (6b)$$

$$\frac{\partial^2 U}{\partial \overline{C}_1 \partial X} = \frac{1}{2\sqrt{x}} \int_{-\infty}^{\infty} SV''\ f(S)dS \quad (6c)$$

$$\frac{\partial^2 U}{\partial x^2} = \frac{-1}{4x^{3/2}} \int_{-\infty}^{\infty} S\ V'\ f(S)\ dS + \frac{1}{4x} \int_{-\infty}^{\infty} S^2\ V''\ f(S)\ dS \qquad (6d)$$

The equation for $\frac{\partial^2 U}{\partial x^2}$ contains two terms. The first one is positive

because $V'' < 0$ and $f(S)$ is symmetric, and the second term is negative be-

cause $V'' < 0$. Its sign, therefore, depends on which term dominates. The

equation can be rewritten as

$$\frac{\partial^2 U}{\partial x^2} = \frac{-1}{4x^{3/2}} \int_{-\infty}^{\infty} S(V' - \sqrt{x} \cdot SV'')\ f(S)\ dS \qquad (7)$$

For a utility function that displays decreasing absolute risk aversion,

$V''' > 0$ and hence, since $f(S)$ is symmetric, equation (6c) implies that

$$\frac{\partial^2 U}{\partial \overline{C}_1 \partial x} > 0$$

Thus if $\frac{\partial^2 U}{\partial x^2} \geq 0$, then sufficient conditions for a <u>decrease</u> in invest-

ment would be met. However, it is not easy to predict the sign of $\frac{\partial^2 U}{\partial x^2}$.

For example

$$\frac{d}{dS}(V' - \sqrt{x} \cdot SV'') = \sqrt{x}V'' - \sqrt{x}\ V'' - xSV'''$$

$$= -x\ S\ V'''$$

$$\text{(for } V''' > 0) \quad < \quad 0 \quad \text{if} \quad S > 0$$

$$> \quad 0 \quad \text{if} \quad S < 0$$

and thus we cannot make use of arguments similar to those used to prove

$\dfrac{\partial^2 U}{\partial \overline{C}_1 \partial x}$ positive.

For particular cases, it can be shown that $\dfrac{\partial^2 U}{\partial x^2}$ is _negative_. For

example, if the utility function is logarithmic, then in equation (7)

$$V' - \sqrt{x}SV'' = \frac{1}{(\overline{C}_1 + \sqrt{x}S)} + \frac{\sqrt{x}S}{(\overline{C}_1 + \sqrt{x}S)^2}$$

$$= \frac{\overline{C}_1 + 2\sqrt{x}S}{(\overline{C}_1 + \sqrt{x}S)^2}$$

This term is higher at S than at -S. This is simply proved. Suppose to

the contrary, and let $\overline{C}_1 = m$ and $\sqrt{x}S = n$. Then the claim is that

$$\frac{m - 2n}{(m - n)^2} \quad > \quad \frac{m + 2n}{(m + n)^2}$$

or

$$(m - 2n)(m + n)^2 \quad > \quad (m + 2n)(m - n)^2$$

or

$$m^3 + 2m^2n + mn^2 - 2m^2n - 4mn^2 - 2n^3 \quad > \quad m^3 - 2m^2n + mn^2 + 2m^2n$$

$$- 4mn^2 + 2n^3$$

or

$$- 2n^3 \quad > \quad 2n^3$$

which is not possible. Hence, given the symmetry of $f(S)$, $\dfrac{\partial^2 U}{\partial x^2} < 0$ for

logarithmic utility. In general, the savings and investment effects of

indexation seem to be ambiguous.

The situation we have here is similar to the "income uncertainty" case of the savings and uncertainty literature, as examplified by Sandmo (7). In Sandmo (7), it is shown that when the investment asset has a certain rate of return, and the individual's second-period income from other sources is uncertain, then a reduction in the uncertainty of second-period income leaves (optimal) savings unchanged in the case of quadratic utility and decreases savings for utility functions with declining absolute risk aversion. The additional, and necessary, complication in our case is that the return on the investment technology is itself uncertain -- which is what gives us the resultant ambiguity.

REFERENCES

1.   Barro, R. "Monetary Correction in Columbia," unpublished manuscript, University of Chicago, 1975.

2.   Black, F. "Capital Market Equilibrium with Restricted Borrowing," Journal of Business, September 1972.

3.   Fischer, S. "The Demand for Index-Linked Bonds," Journal of Political Economy, June 1975.

4.   Levhari, D. and N. Liviatan.

     (a) "Government Intermediation in the Indexed Bond Market," Hebrew University Working Paper (1975).

     (b) "Government Intermediation in the Indexed Bond Market," American Economic Review, May 1976.

5.   Mossin, J. "Equilibrium in the Capital Asset Market," Econometrica, October 1966.

6.   Ng, D. S. "Information Accuracy and Social Welfare Under Homogeneous Beliefs," Journal of Financial Economics, March 1975.

7.   Sandmo, A. "The Effect of Uncertainty on Savings Decisions," Review of Economic Studies, 1970.

8.   Sarnat, J. "Purchasing Power Risk, Portfolio Analysis, and the Demand for Index-Linked Bonds," Journal of Money, Credit and Banking, August 1973.

8a. Comments on (8).

     (i)   Ragazzi, G. "Index-Linking and General Welfare," Journal of Money, Credit, and Banking, May 1976.

     (ii) Bicksler, J. L. and P. J. Hess. "More on Purchasing Power Risk, Portfolio Analysis, and the Case for Index-Linked Bonds," Journal of Money, Credit, and Banking, May 1976.

9.   Subrahmanyam, M. G. "The Optimal Introduction of Index-Linked Bonds," in Ph.D. Thesis, "Essays on the Optimal Allocation of Capital," 1974, Sloan School of Management, M.I.T.

10. Turnbull, S. "Inflation, Indexation, and the Structure of Returns," Manuscript, Boston University, January 1976.

Discrete and Continuous-Time Valuation

of Mean-Reverting Cash Flow Streams

Sudipto Bhattacharya

I. Introduction

The work done so far on the valuation of multi-period cash-flow streams has involved essentially positing some intertemporal stochastic process for cash flows and deriving the corresponding valuation using an intertemporal expected return-risk relationship like the capital asset pricing model. Applications to specific stochastic processes which provide tractable, simple and "intuitively interpretable" valuation formulas have been made primarily by Myers (7), Myers and Turnbull (8), Rubenstein (9), and Treynor and Black (13), although many others, including Bogue and Roll (2), Constantanides (4), Fama (5), Hamada (6), Brennan (3) and Turnbull (10), have worked on this type of problem. All applications to date have been to processes which have no mean-reverting properties. One of the processes used by Treynor and Black (13) and that used by Myers and Turnbull (8), which are essentially random walks, are processes that are "purely indeterministic," where the cash flow distributions do not converge to stationary distribution with finite second moments--essentially because of no tendency to revert to a mean value. In this paper, we augment this literature by deriving simple, "intuitively interpretable" valuation formulas for two prototypical mean-reverting processes, one in discrete and one in continuous time.

The primary justification for mean-reverting processes
is an economic one. Investment projects have positive net present
values if rents are earned in the product or factor markets.
In a competitive economy, we should expect some tendencies, at least
in the long-run, for these extra-ordinary rents to decline toward
levels that make firms indifferent about investment in that type
of investment opportunity. It is not reasonable, therefore, to
postulate that cash flows from particular assets "wander" forever.
In this paper the specific "long-run mean" that the cash flows tend
to revert to represents two alternative possibilities. It may be
thought of as that level of cash flows at which the net present
value of the project is zero if current cash flow is at that level,
and it follows the same stochastic process. Alternatively, along the
lines of a "product life-cycle" model, we may think of it as
representing expected profits at the ultimate "saturation" level
that the expected sales asymptotically tend to.

We concentrate largely on the valuation and capital budgeting
implications of mean-reverting processes and in particular on the
following two issues.

(a) How project risk   and expected return respond to
changes in the maturity, the strength of mean-reversion, and the
excess/deficit of transient cash flows over the "long-run mean"
level.

(b) Whether or not the "true" valuation can be approximated by discounting the expected cash flows at a risk-adjusted cost-of-capital. In this way, we extend the analysis of and the questions asked by Myers and Turnbull (8) to the case of mean-reverting cash flow processes. Our answers to question (b) -- the adequacy of "cost of capital" valuation -- also differs from theirs.

To provide the intertemporal recursion relationship for valuation, we use the discrete and continuous-time versions of the simple CAPM of Sharpe, Lintner and Mossin for discrete time and Merton's (11) (with intertemporally independent investment opportunities), for continuous-time. A problem with mean-reverting cash flows is that the resulting valuation formulas produce expected returns that are not intertemporally stationary--not unexpectedly, they depend on the relative values of the current, transient cash flow level and the long-run mean that it reverts to. This is a problem that has to be skirted around. From the work of many competent researchers, we know that the determination of multi-period equilibrium pricing models in the presence of non-stationary opportunity sets (for investors) is an extremely difficult and messy endeavor--and certainly the "hedging" approach of Merton's intertemporal CAPM (11) cannot be applied without a very messy general equilibrium linking-up of firm return non-stationarities with possibly existing intertemporal hedging portfolios. We, there-

fore, apply the simple CAPM pricing relationship--and our assumption that it is an adequate one is bolstered more by its intuitive appeal than its literal accuracy in all environments. If one wants to be "rigorous" one can assume investors have logarithmic utility, which would make the continuous-time CAPM quite valid. The discrete-time CAPM has further problems associated with the conflict between the required assumption of Gaussian return distributions and limited liability--but, like many before, we shall appeal to approximate validity for approximately Guassian distributions in order to obtain the advantages of using its intuitively appealing risk-return trade-off features.[1] Similar arguments apply to two other assumptions that we make--that the riskless rate of interest and the market price of risk are constant over time.

II

Discrete-Time Valuation

In discrete-time, we provide a valuation formula for a fairly general mean-reverting process, which is the auto-regressive model with an observation error. Specifically, suppose the variable X_t, the t-th period cash flow, is related to past values by the relationship:

$$x_t = (1-\alpha-V)[x_{t-1} + \alpha x_{t-2} + \alpha^2 x_{t-3} \cdots] + \tilde{\varepsilon}_t$$

where

$$x_t = (X_t - \overline{X})$$

and it is assumed that $0 \leq \alpha$, $V \leq 1$, $0 \leq (\alpha+V) \leq 1$ and $\{\varepsilon_t\}$ are independently identically distributed normal variables with zero mean, and $\overline{X}$ is a long-run mean that the process (as we shall see) tends to revert to. Special cases of this process are:

(a) $\qquad \alpha = 0 \qquad x_t = (1-V)x_{t-1} + \varepsilon_t$

which is the simple auto-regressive process with correlation parameter $\rho = (1-V)$

(b) $\qquad \alpha = 0$, $V = 0 \qquad x_t = x_{t-1} + \varepsilon_t$

which is the simple random walk

(c) $\qquad V = 0 \qquad\qquad x_t = (1-\alpha)[x_{t-1} + \alpha x_{t-2} + \alpha^2 x_{t-3} \cdots] + \varepsilon_t$

which is the random walk with an error of observation--and we shall soon see why it gets that name.

Treynor and Black (13), and Myers and Turnbull (8) used modified

versions of processes (b) and (c), respectively--with the further

difference that ε_t was proportional to the value of x_t anticipated

at (t-1). The former used a continuous-time version. In these

valuation models, there is one good reason to make ε_t proportional

to some cash flow value, since we want cash flows bounded by zero

on the lower side--otherwise complicated problems of optimal abandon-

ment time must be solved as part of the valuation exercise. The

proportional error model of Treynor and Black does this rigorously,

and Myers and Turnbull (8) do so if the proportionate error term

is strictly bounded away from -1, and this clearly has to make

its Gaussian character approximate. We deal for the moment with the

simple error model since it is easier to discuss some characteristics

of these processes, but for valuation we too switch to a proportionate

error model.

Let us deal first with the period to period changes in

levels and anticipations implied by the process in (1). Let $x_t{}^*, X_t{}^*$

be the anticipated t-th period values of x_t, X_t -- anticipated at t-1.

Then, for rational expectations,

$$x_t{}^* = (1-\alpha-V)[x_{t-1} + \alpha x_{t-2} + \ldots]$$

therefore $x_t{}^* - \alpha x_{t-1}^* = (1-\alpha-V)x_{t-1}$

or $\qquad x_t{}^* - x_{t-1}^* = (1-\alpha-V)(x_{t-1} - x_{t-1}^*) - Vx_{t-1}^*$

or $\qquad X_t{}^* - X_{t-1}^* = (1-\alpha-V)(X_{t-1} - X_{t-1}^*) + V(\overline{X} - X_{t-1}^*)$

$$= (1-\alpha-V)\varepsilon_{t-1} + V(\overline{X} - X_{t-1}^*)$$

$$= a\,\varepsilon_{t-1} + b(\overline{X} - X_{t-1}^*) \qquad\qquad (2)$$

Again, for the special cases

(i) If $\alpha = 0$, a pure autoregressive process, then $b = (1-a)$; if
further $b = 1$ then cash flows are intertemporally identically
distributed;

(ii) If $V = 0$, simple random walk or random walk with error or
observation. In this case, because $b = 0$, there is no
mean-reverting tendency, and the (rational) revision of
expectations is purely "adaptive." If further $\alpha = 0$, then
$a = 1$, the "elasticity of expectations" is unity.

The X_t's themselves follow a similar relationship, viz.

$$X_t - X_{t-1} = V(\overline{X} - X_{t-1}) + \varepsilon_t - \alpha\varepsilon_{t-1} \qquad\qquad (3)$$

Some further insight may be gained by looking at the moving-average representation of the process--in terms of a moving average of i.i.d. random variables of zero mean. It can be easily shown that the process (1) satisfies the moving average representation.

$$x_t = (X_t - \overline{X}) = Z_t + \gamma_t + \sum_{i=1}^{\infty}(1-V)^{i-1}Z_{t-i} \tag{4a}$$

where Z_t, γ_t are serially uncorrelated zero mean random variables satisfying the relationship

$$\gamma_t = \frac{(\alpha+V)}{(1-\alpha-V)} Z_t \tag{4b}$$

and, in line with the previous auto-regressive representation,

$$\varepsilon_t = Z_t + \gamma_t = \frac{1}{(1-\alpha-V)} Z_t \tag{4c}$$

It is instructive to examine the special cases again:

(i) $\alpha = 0$, pure auto-regressive

$$x_t = \sum_{i=0}^{\infty}(1-V)^{i-1}Z_{t-i} = \sum_{i=0}^{\infty} (1-V)^{i}\varepsilon_{t-i}$$

With $V > 0$, it is clear that a stochastic deviation at some $(t-i)$

has a smaller "impact" on x_t as i gets larger.

(ii) V = 0, random walk or random walk with error

$$x_t = \frac{Z_t}{(1-\alpha)} + \sum_{i=1}^{\infty} Z_{t-i}$$

which obviously has no (finite-variance) stationary distribution.
It also illustrates that, when $\alpha > 0$, the lower than unit "elasticity"
of expectations" $(1-\alpha)$ associated with the random walk with an
observation error is somewhat "illusory"--in the sense that it arises
not from a diminishing impact of distant-past stochasticity but from
a simple error-of-observation type additional error in the contempo-
raneous period alone.

Let us now proceed to derive a multi-period valuation
formula for a project with maturity T and cash flows given by
X_t, which satisfy the stochastic relationships

$$X_t = X^*_t(1 + \delta_t) \tag{5}$$

where X^*_t is the (rationally) anticipated value of X_t at t-1,
and $\{\delta_t\}$ are i.i.d., zero-mean, "almost Gaussian" random variables
bounded away from -1 and, further,

$$X^*_t = X^*_{t-1}(1+a\delta_{t-1}) + b(\overline{X} - X^*_{t-1}) \tag{6}$$

As is easily appreciated, this process is a simple modification

of that described by equations (1) and (2), where the error term

ε_t is now proportional to X^*_t (i.e., $\varepsilon_t = X^*_t \delta_t$ and δ_t are i.i.d.).

Since δ_t are bounded away from -1, cash flows are always positive.

The "almost-Gaussian" assumption requires that $|-1| = 1$ is large

compared to σ_δ the standard deviation of each $\{\delta_t\}$ so that the

approximation involved in using the discrete-time CAPM to value

the cash flows is small. At any time τ let $\tilde{Y}_{\tau+1} = \tilde{X}_{\tau+1} + \tilde{P}_{\tau+1}$

be the (uncertain) cash flows plus value (of the remaining stream)

at $(\tau+1)$. Then the discrete-time CAPM leads to the recursion

relationship

$$P_\tau = \frac{1}{1+r} [E_\tau(\tilde{Y}_{\tau+1}) - \lambda\rho\sigma(\tilde{Y}_{\tau+1})] \tag{7}$$

where E_τ denotes the expectations operator as of τ, r is the riskless

rate of interest, λ is the market price of risk, $\lambda = \dfrac{E(\tilde{R}_M)-r}{\sigma(\tilde{R}_M)}$, where

$\tilde{R}_M$ is the market rate of return, $\sigma(\tilde{Y})$, $\sigma(\tilde{R}_M)$, denote

the standard deviations of these variables, and ρ is the (assumed

stationary) <u>correlation</u> coefficient between $\tilde{\delta}$ and $\tilde{R}_M$.

Since the project ends at T, clearly at (T-1)

$$\tilde{Y}_T = \tilde{X}_T = X^*_T(1 + \tilde{\delta}_T)$$

and hence

$$P_{T-1} = \frac{X^*_T}{(1+r)} [1 - \lambda\rho\sigma_\delta] = X^*_T \cdot \phi \qquad (8a)$$

where

$$\phi = \frac{(1-\lambda\rho\sigma_\delta)}{(1+r)}$$

At T-2,

$$\tilde{Y}_{T-1} = \tilde{X}_{T-1} + \tilde{P}_{T-1}$$

$$= X^*_{T-1}(1+\tilde{\delta}_{T-1}) + \phi[X^*_{T-1}(1+a\tilde{\delta}_{T-1}) + b(\overline{X} - X^*_{T-1})]$$

Hence

$$F_{T-2}(\tilde{Y}_{T-1}) = X^*_{T-1}(1+\phi(1-b)) + \phi b\overline{X}$$

$$\sigma_{T-2}(\tilde{Y}_{T-1}) = X^*_{T-1}\sigma_\delta(1 + \phi a)$$

Hence

$$P_{T-2} = X^*_{T-1}[\frac{(1-\lambda\rho\sigma_\delta)}{(1+r)} + \phi(\frac{(1-b) - a\lambda\rho\sigma_\delta}{(1+r)})]$$

$$+ \frac{\phi b}{(1+r)} \cdot \overline{X}$$

Let

$$Z = [\ \frac{(1-b)\ -\ a\lambda\rho\sigma_\delta}{(1+r)}\ ] \tag{8b}$$

Then

$$P_{T-2} = X^*_{T-1}\phi(1+Z) + \frac{\phi b\overline{X}}{(1+r)}$$

Similarly, it can be easily shown that, for example,

$$P_{T-3} = X^*_{T-2}\phi(1+Z+Z^2) + \phi b\overline{X}[\ \frac{(1+Z)}{(1+r)} + \frac{1}{(1+r)^2}\ ]$$

and, in general, for n > 1

$$P_{T-n} = X^*_{T-n+1}\phi(1+Z+Z^2 + \ldots Z^{n-1})$$

$$+ \phi b\overline{X}[\ \frac{1}{(1+r)^{n-1}} + \frac{(1+Z)}{(1+r)^{n-2}} + \ldots + \frac{(1+Z+Z^2+Z^{n-2})}{(1+r)}\ ]$$

$$\tag{9}$$

When b = 0, the formula becomes the same as that of Myers-Turnbull
(8) and Myers (7), as expected.

The formula for a perpetuity is of some interest, and
this is found by letting n → ∞ in (9). Let us first consider

the term in the bracket multiplying $\phi b \overline{X}$ in (9) and call it S_n for the term in P_{t-n}. Then S_n satisfies

$$S_{n+1} = \frac{S_n}{(1+r)} + \frac{(1+Z+Z^2+\ldots+Z^{n-1})}{(1+r)}$$

As $n \to \infty$, the fixed point of this mapping is given by:

$$S_\infty = \frac{S_\infty}{(1+r)} + \frac{1}{(1-Z)(1+r)}$$

or

$$S_\infty = \frac{1}{r(1-Z)} \tag{10}$$

$\{S_n\}$ is a convergent sequence as long as the series $(1+Z+Z^2\ldots)$ converges, i.e., for $|Z| < 1$, which is also needed for the first term of (9) to converge. Sufficient conditions for $|Z| < 1$ are $r > 0$, $\rho\sigma_\delta \geq 0$, i.e., a positive riskless rate and a non-negative "beta," but the necessary conditions are less stringent. We can then write the full valuation for a perpetuity as $n \to \infty$ as

$$P_0 = X^*_1 \frac{\phi}{1-Z} + \frac{\overline{X}\phi b}{r(1-Z)} \tag{11a}$$

which, on utilizing the definition ϕ, Z from (8a), (8b), becomes

$$P_0 = [X^*_1 \frac{(1+r)}{(r+b+\lambda\rho a\sigma_\delta)} + \frac{\overline{X}b(1+r)}{r(r+b+\lambda\rho a\sigma_\delta)}]$$

$$= (1-\lambda\rho\sigma_\delta) [\frac{rX^*_1 + b\overline{X}}{r(r+b+\lambda\rho a\sigma_\delta)}] \tag{11b}$$

Note that the relative weights of the transient and long-run means X^*_1 and $\overline{X}$ depend on r and b, respectively--which is sensible since r going up represents a higher weight for nearby cash flows, whereas b going up implies a stronger tendency to revert to the mean value $\overline{X}$ per unit of time. The simple first-order auto-regressive case is obtained by setting b = (1-a). With intertemporally independent cash flows, b = 1, a = 0, $X^*_1 = \overline{X}$ and $P_0 = \frac{\overline{X}(1-\lambda\rho\sigma_\delta)}{r}$, i.e., a pure "certainty equivalent" adjustment.

We may also examine the behavior of the period to period rate of return on the perpetual project--in particular its relation to the ratio of the transient X^*_1 to $\overline{X}$. The gross period to period rate of return is given by:

$$\widetilde{R}_1 = \frac{X^*_1(1+\widetilde{\delta}_1) + k[r\{X^*_1(1+a\widetilde{\delta}_1) + b(\overline{X}-X^*_1)\} + b\overline{X}]}{k[rX^*_1 + b\overline{X}]}$$

where

$$k = \frac{(1-\lambda\rho\sigma_\delta)}{r(r+b+\lambda\rho a\sigma_\delta)} \tag{11c}$$

On rearrangement we get

$$\tilde{R}_1 = 1 + \frac{[X^*_1 + krb(\overline{X}-X^*_1)] + X^*_1 \tilde{\delta}_1 (1+kra)}{kb(\overline{X}-X^*_1) + k(r+b)X^*_1}$$

The expected net rate of return is therefore

$$E_0(\tilde{R}_1-1) = \frac{1}{k(r+b)} \left[\frac{1 + krb(\overline{X}-X^*_1)/X^*_1}{1 + \frac{b}{(r+b)} \frac{(\overline{X}-X^*_1)}{X^*_1}} \right]$$

The expected return <u>decreases</u> with an increase in $\left(\frac{\overline{X}-X^*_1}{X^*_1} \right)$ and hence, <u>increases</u> with an increase in $\frac{X^*_1}{\overline{X}}$ if

$$krb < b/(r+b)$$

or, since $b > 0$, if $kr < \frac{1}{(r+b)}$ which, from (11c) is seen to be the case if $(\rho\sigma_\delta) > 0$, i.e., the "beta" is positive. If $\rho\sigma_\delta > 0$, it also is readily seen that the beta, i.e., $\frac{\rho\sigma(\tilde{R}_1)}{\sigma(\tilde{R}_M)}$ decreases with an increase in $\frac{(\overline{X}-X^*_1)}{X^*_1}$ and thus increases with $X^*_1/\overline{X}$. If convergence (of the perpetuity valuation) holds, then this behavior is exactly reversed for negative beta, i.e., $\rho\sigma_\delta < 0$. This comparative static, incidentally, is just the reverse of that empirically found by Black (l), who finds that common stock volatility declines with an increase in stock value. Thus other possible reasons mentioned in

in Black (1) have to be invoked to explain that behavior--
mean reversion in cash flows is not a possible explanation.
Intuitively, this behavior of expected return and beta is easily
explained. Since the total unanticipated variability of next
period cash flows plus value is proportional to X^*_1, the expected
next-period cash flow level, when the relative proportion of the
value contributed by the term proportional to X^*_1 rises (due to a
rise in X^*_1, ceteris paribus), the unanticipated variability of the
rate of return rises. This causes a rise in expected return if
beta is positive and vice-versa.

III

In this section we provide a simple and intuitively
interpretable valuation formula for the continuous-time case for
a project of maturity T with the cash flow rate per unit time
X(t) following the mean reverting stochastic process.

$$dX = (B-AX)dt + VXdZ \qquad (12a)$$

V,B,A > 0 and dZ is a Wiener process. We can alternatively write

$$dX = A(\frac{B}{A} - X)dt + VXdZ \qquad (12b)$$

Clearly the cash flows tend to revert down to $\frac{B}{A}$ when $X > B/A$

and vice-versa. By referencing equation (3), it is seen that

this is the continuous time analog to the first-order auto-regressive

process (with $\alpha = 0$) with an error-term proportional to X_t; note,

however, that our discrete time derivation is carried out by

assuming that ε_t is proportional to X^*_t, and thus there is a

difference in detail.[2]

Let $F(X,\tau)$ be the value of the project when remaining

maturity is τ and the current cash flow rate is X. From $\hat{I}$to's

Lemma we get the expressions for the expected returns and the

standard deviation of return of this project to be $E(R_\tau)$ =

$$\text{Expected return} = \frac{\frac{1}{2}V^2 X^2 F_{XX} + (B-AX)F_X - F_\tau + X}{F}$$

$$\sigma(R_\tau) = \text{standard deviation of return} = \frac{VXF_X}{F}$$

From the continuous-time CAPM we have

$$E(R_\tau) = r + \lambda\rho\sigma(R_\tau) = r + a\sigma(R_\tau) \quad \text{where } a = \lambda\rho \tag{13}$$

where λ, again, is the market price of risk, r is the riskless rate

of interest and ρ is the (assumed stationary) correlation between

dZ and the rate of return on the market portfolio.

So we have,

$$\frac{\frac{1}{2}V^2X^2F_{XX} + (B-AX)F_X - F_\tau + X}{F} = r + \frac{aVXF_X}{F}$$

or,

$$\frac{1}{2}V^2X^2F_{XX} + [B - (A+aV)X]F_X - F_\tau - rF + X = 0 \qquad (14a)$$

with the terminal condition $F(X,0) = 0$ (14b)

i.e., value is zero at maturity.

Note that there is no boundary condition at $X = 0$, since it is

a reflecting barrier for the process (12a)--a reflecting barrier

that is not reached with probability one in this case.[3]

The path to the solution is seen more transparently by

substituting with the variable

$$Y = B - (A+aV)X \qquad (15a)$$

to get

$$F_X = F_Y(A+aV) \qquad F_{XX} = F_{YY}(A+aV)^2 \qquad (15b)$$

and, finally, from (14)

$$\tfrac{1}{2} V^2(B-Y)^2 F_{YY} - Y(A+aV)F_Y - rF - F_\tau + \frac{(B-Y)}{(A+aV)} = 0 \qquad (16a)$$

$$F(Y,0) = 0 \qquad (16b)$$

We try a solution of the form

$$F(Y,\tau) = g(\tau) + Q(Y,\tau) \qquad (17)$$

Suppose $g(\tau)$ satisfies

$$rg + g_\tau = B/(A+aV) \quad \text{and} \quad g(0) = 0$$

This is clearly satisfied by

$$g_\tau = \frac{B}{(A+aV)\cdot r} [1 - e^{-r\tau}] \qquad (18a)$$

Then $Q(Y,\tau)$ must satisfy

$$\tfrac{1}{2} V^2(B-Y)^2 Q_{YY} - Y(A+aV)Q_Y - rQ - Q_\tau - \frac{Y}{(A+aV)} = 0$$

Let us try the solution $\qquad Q(Y,\tau) = KY[1 - e^{-d\tau}]$

which clearly satisfies the terminal condition.

K,d are undetermined parameters. On substitution, we get

$$Q_Y = K[1 - e^{-d\tau}], \quad Q_\tau = KYde^{-d\tau}, \quad Q_{YY} = 0$$

Hence, for the equation to be satisfied, we must have (by substituting into the differential equation and cancelling out)

$$-(A+aV)K(1 - e^{-d\tau}) - rK(1 - e^{-d\tau}) - Kde^{-d\tau} - \frac{1}{(A+av)} = 0$$

which implies that

$$-(A+aV)K - rK - \frac{1}{(A+aV)} = 0$$

or $\quad K = -(1/(A+aV))/(r+A+aV)$

and further

$$(A+aV)K + rK - Kd = 0$$

or $\quad d = (r+A+aV)$

So we have

$$F(Y,\tau) = g(\tau) + Q(Y,\tau)$$

$$= \frac{B}{r(A+aV)}[1 - e^{-r\tau}] + \frac{(X-B/(A+aV))}{(r+A+aV)}[1 - e^{-(r+A+aV)\tau}] \quad (19)^4$$

Equation (19), the valuation formula, has a clear "intuitive"

interpretation. Given the reverting value, B/A, $\frac{B}{(A+aV)}$ is

the "certainty equivalent" of it. The valuation formula contains

a cash flow rate equal to the certainty equivalent discounted at

the riskless rate, plus a term in which the transient excess of

X over this certainty equivalent is discounted at the higher

"risk-adjusted" cost of capital (r+A+aV). A more striking "intuitive"

analogy can be given by considering the valuation of the

following certain cash flow streams: Suppose

$$dX = A(\frac{B}{A} - X)dt$$

Then if "now" is t = 0, X at some t > 0 satisfies

$$(\frac{B}{A} - X(t)) = (B/A - X(0))e^{-At}$$

i.e., X(t) approaches B/A (with certainty) exponentially--it

never crosses B/A though. Clearly this stream of maturity τ would

be valued at

$$V(\tau) = \int_{0}^{\tau} X(t)e^{-rt}dt$$

$$= \int_{0}^{\tau} e^{-rt}[\frac{B}{A} + X(t) - \frac{B}{A}]dt$$

$$= \frac{B}{A \cdot r}[1 - e^{-r\tau}] + \int_0^\tau (\frac{B}{A} - X(0))e^{-(r+A)t}dt$$

$$= \frac{B}{A \cdot r}[1 - e^{-r\tau}] + \frac{(\frac{B}{A} - X(0))}{(r+A)} [1 - e^{-(r+A)\tau}] \qquad (20)$$

In our uncertainty case the difference is that there is only a mean reverting tendency towards B/A -- the associated random movements would lead to crossing B/A from both sides. This uncertainty results in two changes in (19) as compared to (20). The "certainty equivalent" for B/A becomes $\frac{B}{(A+aV)}$ and the (r+A) discount cum trend adjustment rate of (20) is adjusted to (r+A+aV). Valuation is done in analogy with the valuation for the "corresponding" certain evolution of cash flows--which is the kind of analogy that the "certainty equivalent" and "cost of capital" notions try to capture, but succeed in doing so only in special cases like here!

Note that if, holding B/A constant, A -- which is a measure of the strength of mean-reversion -- increases, then $\frac{B}{(A+aV)}$ moves closer to B/A and (r+A+aV) moves closer to (r+A) (in proportionate terms), i.e., the effect of the uncertainty on valuation is decreased, which is sensible. The simplicity of the valuation formula, which is rather striking, is not likely to be replicated for other mean-reverting, bounded above zero dynamics for X(τ).

As regards instantaneous rates of return on the project, we have that

Standard Deviation of Rate of Return

$$\sigma(R_\tau) = \frac{VXF_X}{F} = \frac{\frac{VX}{(r+A+av)}[1 - e^{-(r+A+av)\tau}]}{\frac{(X-\frac{B}{A+av})}{(r+A+av)}\left(1 - e^{-(r+A+av)\tau}\right) + \frac{B}{(A+av)r}\left(1 - e^{-r\tau}\right)}$$

$$= \frac{V}{1 + \frac{1}{X} \cdot \frac{B}{(A+av)}\left(\frac{\frac{1}{r}(1-e^{-r\tau})}{\frac{1}{(r+A+av)}}(1-e^{-(r+A+av)\tau}) - 1\right)}$$

(21)

It is clear that if av is positive (i.e., positive beta), then

standard deviation and beta (and hence expected return) decrease as

$\frac{1}{X}\frac{B}{(A+av)}$ goes up and, hence, increase as X goes up. For negative

beta (av < 0) the same behavior is true for standard deviation, but

beta (algebraically) and hence expected return behaves oppositely.

These results are similar to those found for the discrete-time model,

and have a similar intuitive explanation.

IV

Some Numerical Examples

In this section we deal with some other important
comparative statics of the valuation formulas derived, asking
questions regarding both direction and magnitude of effects.
We also investigate, numerically, as to how well a "naive"
cost-of-capital type discount rule applied to expected future cash
flows works out compared to the theoretical valuation formula--
when the discount rate is chosen to reflect the beta of the
project at some "average" level of cash flows and maturity.

We stick to the continuous-time valuation formula for
this whole exercise because of its simplicity. From equation (21)
(and the intertemporal CAPM that was used) it is clear that the
expected return on the (value of) the project when current cash
flow rate is X is given by:

$$Z = r + \cfrac{av}{1 + \frac{1}{X} \cdot (\frac{B}{A+av}) \left[\cfrac{\frac{1}{r}(1 - e^{-r\tau})}{\frac{1}{(r+A+av)}(1 - e^{-(r+A+av)\tau})} - 1 \right]}$$

(22)

It is clear for positive beta (av > 0) that, holding B/A constant,

the expected return (and the beta) are (a) decreasing in A, and

(b) decreasing in τ, the maturity (left).[5] Table 1 provides some

results on the magnitudes involved. It is also interesting

to inquire as to how Z varies when τ is varied but, instead of

keeping A constant, we vary it so that something like the

"half-life" of approach to the "mean value" B/A is kept constant as

a proportion of maturity. Since this aspect of the exercise is

also important for calculating the expected levels for future

cash flow rates (given the current level X), we deal with this in

a little more detail below.

Let E(x,t) be the expected value of the cash flow rate

t   time away -- given that the current cash flow rate is X. Then

E(x,t) satisfies the Kolmogorov backward equation for this process:

$$\tfrac{1}{2} \, V^2 X^2 E_{XX} + (B-AX)E_X - E_t = 0 \qquad (23a)$$

and the boundary condition

$$E(X,0) = X$$

It is straightforward to show that the solution to this is given by

$$(\frac{B}{A} - E(x,t)) = (\frac{B}{A} - X)e^{-At} \tag{23b}$$

Given this, a project of maturity τ has "half-life" τ^* if

$$\frac{(B/A - E(x,\tau^*))}{(B/A - X)} = e^{-A\tau^*} = \frac{1}{2}$$

$$\text{or} \quad A = \frac{\text{Log }(2)}{\tau^*}$$

In our numerical simulation, we have considered $\tau^* = \tau$, $\frac{\tau}{\sqrt{5}}$, and $\frac{\tau}{5}$
i.e., equal to, $1/\sqrt{5}$ of, and $1/5$ of, maturity, respectively.
For the cost of capital discounting exercise we have somewhat
arbitrarily taken Z to be that which results when $X = B/A$ and
τ = half the maturity of the projects. Given the structure of
expectations about future cash flows, the "cost of capital
valuation" (with "cost of capital" Z) is given by

$$P = \int_0^\tau E(x,t)e^{-Zt}dt = \int_0^\tau (\frac{B}{A} + E(x,t) - \frac{B}{A})e^{-Zt}dt$$

$$= (\frac{B}{A}) \frac{1}{Z}[1 - e^{-Z\tau}] + \frac{(X - \frac{B}{A})}{(Z+A)} [1 - e^{-(Z+A)\tau}] \tag{24}$$

The formula, in its basic structure, is so similar to the correct
valuation formula of equation (19) that one must admit to some
doubts as to why the "simplification" of the cost of capital formula

is really worthwhile (and would be followed by corporate management)

in this case -- the only strong counter-argument arises in a setting

in which cash flow expectations generation and project evaluation

are done at different levels, and where the theory approximately

captures the full reality of the stochastic processes, etc., involved.

In Table 2 below, we summarize the results of our numerical

simulations. Besides confirming the comparative statics that

have been discussed before, the table provides two interesting

results. First, it is interesting that if maturity T is varied

and A is also varied to keep the ratio of "half-life" of deviations

to maturity constant, then there is remarkably little, almost

negligible, effect on project beta and expected return at $X = \frac{B}{A}$

and mid-point of life. The effect due to the decline of A is counter-

acted, almost exactly, by the effect due to the increase of τ

alongside. Second, the percentage error in valuation arising from

the "heuristic" cost of capital method can be quite sizable. The

error, which goes up with maturity (holding half-life as a fraction

of maturity constant) is about 8 per cent at $\frac{X}{(B/A)} = .75$ for

T = 20 and 10 per cent for the same X at T = 40 -- though the errors

are quire small, of the order of 2-3 per cent for small maturities,

e.g., T = 5. Thus valuation using a "standardized," "risk-adjusted"

cost of capital may not be so robust after all -- and this stands

in some contrast to the results of Myers and Turnbull (8) for their

process.

The magnitude (in percentage terms) of the bias caused by the cost of capital valuation increases as X decreases relative to (B/A). The reason for this may be seen by examining the perpetuity case of the valuation formula. In that case,

$$\text{Correct Valuation} \quad F = \frac{B}{(A+av)} \left[\frac{1}{r} - \frac{1}{(r+A+av)} \right] + \frac{X}{(r+A+av)} \tag{25}$$

$$= \frac{B}{r(r+A+av)} + \frac{X}{(r+A+av)}$$

Expected return (on value of the project) at X = B/A

$$Z = r + \frac{av}{1 + \frac{A}{(A+av)} \left[\frac{(r+A+av)}{r} - 1 \right]}$$

$$= \frac{(r+A+av)}{1 + \frac{A}{r}}$$

$$= \frac{r(r+A+av)}{(r+A)}$$

$$\text{and} \quad Z+A = \frac{A(r+A) + r(r+A+av)}{(r+A)} = \frac{(r+A)^2 + rav}{(r+A)}$$

The "cost of capital valuation" is given by

$$P = \frac{B}{A \cdot Z} + \frac{(X - \frac{B}{A})}{(Z+A)}$$

$$= \frac{B}{A} \cdot \frac{(r+A)}{r(r+A+av)} + (X - \frac{B}{A}) \cdot \frac{(r+A)}{(r+A)^2 + rav}$$

$$= \frac{B}{A} \cdot (r+A) \left[\frac{(r+A)^2 + rav - r(r+A+av)}{r(r+A+av)\{(r+A)^2 + rav\}} \right] + \frac{X}{(r+A) + \frac{rav}{(r+A)}}$$

$$= \frac{B(r+A)^2}{r(r+A+av)[(r+A)^2 + rav]} + \frac{X}{(r+A) + \frac{rav}{(r+A)}} \qquad (26)$$

By comparing (26) with (25) it is seen that for positive beta (av > 0) the "cost of capital valuation" puts a relatively higher weight on the transient X and a relatively lower weight on B/A. This explains why its proportionate downward bias is greater when X is lower compared to $(\frac{B}{A})$, ceteris paribus. The existence of any bias at all in the perpetuity case is another result that stands in contrast to those of Myers and Turnbull (8).

V

Conclusions

We have attempted to extend the literature on the valuation
of multiperiod income streams to two examples of mean-reverting
processes, and to ask some questions regarding (a) the effect
of various parameters, and (b) the adequacy of approximating by
a "cost of capital" type discounting. Answers have been
obtained that extend and throw some new light on the issues
addressed in the existing literature in this area.

Table 1

Expected Return at Different Points of Project
Life and Levels of Cash Flow

Assumption: r = .05   av = .1   (B/A) = 1

Project Life T	Half-life $H=\dfrac{\log 2}{A}$	Fraction of Project-Life Left(τ/T)	Levels of Cash Flow $X/(B/A)=X$ 1.5	1.25	1.0	.75	.5
5	$5/\sqrt{5}$	1	.111	.107	.101	.094	.084
		1/2	.127	.124	.119	.113	.103
		1/3	.134	.132	.128	.123	.114
		1/4	.138	.136	.133	.128	.121
		1/5	.140	.138	.136	.132	.125
10	$10/\sqrt{5}$	1	.112	.108	.102	.095	.085
		1/2	.127	.124	.119	.113	.103
		1/3	.134	.131	.128	.123	.114
		1/4	.138	.136	.133	.128	.121
		1/5	.140	.138	.136	.132	.125
20	$20/\sqrt{5}$	1	.114	.110	.105	.097	.088
		1/2	.128	.124	.120	.113	.103
		1/3	.134	.131	.128	.122	.114
		1/4	.138	.136	.133	.128	.120
		1/5	.140	.138	.136	.132	.125
40	$40/\sqrt{5}$	1	.120	.116	.111	.104	.094
		1/2	.129	.126	.121	.115	.105
		1/3	.134	.132	.128	.123	.114
		1/4	.138	.136	.133	.128	.120
		1/5	.140	.138	.136	.132	.125

Table 2

Accuracy of "Cost of Capital" Valuation

Assumptions: r = .05   av = .1   (B/A) = 1

Project Maturity	Half-life	A	"Cost of Capital"	Cash Flow Level	Ratio of Cost of Capital Value to Correct Valuation
"Years"	H	A	Z	X	
5	5	.1343	.1386	1.5	1.001
				1.25	.996
				1.0	.990
				0.75	.981
				0.5	.967
	$5/\sqrt{5}$	.310	.1194	1.5	.999
				1.25	.993
				1.0	.985
				.75	.975
				.5	.962
	5/5	.693	.09768	1.5	.9952
				1.25	.9907
				1.0	.9855
				.75	.9796
				.5	.9727
10	10	.069	.1342	1.5	1.000
				1.25	.992
				1.0	.980
				.75	.964
				.5	.938
	$10/\sqrt{5}$	.155	.1194	1.5	.996
				1.25	.985
				1.0	.971
				.75	.953
				.5	.929
	$10/\sqrt{5}$	.347	.0980	1.5	.991
				1.25	.982
				1.0	.973
				.75	.961
				.5	.947

Table 2 continued

"Years"	H	A	Z	X	Ratio...
20	20	.0347	.1343	1.5	.998
				1.25	.985
				1.0	.967
				.75	.941
				.5	.900
	$20/\sqrt{5}$	.0775	.1197	1.5	.993
				1.25	.975
				1.0	.952
				.75	.922
				.5	.881
	20/5	.173	.0989	1.5	.986
				1.25	.971
				1.0	.953
				.75	.931
				.5	.905
40	40	.0173	.135	1.5	1.002
				1.25	.986
				1.0	.965
				.75	.933
				.5	.880
	$40/\sqrt{5}$	.0387	.1213	1.5	1.000
				1.25	.977
				1.0	.946
				.75	.904
				.5	.844
	40/5	.0866	.1016	1.5	.994
				1.25	.969
				1.0	.940
				.75	.903
				.5	.858

Footnotes

1. Alternatively, we may take recourse to the Ross (12) linear
factor arbitrage model of capital asset pricing.

2. This case is also of special interest because any response of
the rate of return "volatility" (and the expected return) to the
transient cash flow level arises only from the cash flow to value
"transformation," since the cash flows themselves show constant
standard deviation to scale.

3. The claims of Constantanides (4) to the contrary, the formulation
of a valuation equation like (14a), for the most general Wiener
process for X, does not constitute a fundamental discovery. Treynor
and Black (13), among others, used this device a long time back.
Nor is solving these equations by the "method of Cox and Ross"
anything new. Equations of the type of (14a) are trivial transforms
of Kolmogorov equations; but only in some cases are the solutions
simple, "intuitive," and interesting to interpret.

Footnotes continued

4. This is, of course, a particular solution. That this is the
unique solution for equation (14a) with the terminal condition
$F(X,0) = 0$ is easily shown.

5. From here on, we deal almost exclusively with the positive
beta (av > 0) case.

References

1. Black, F., "Studies of Stock Price Volatility Changes," Proceedings
        of the 1976 Meetings of the American Statistical Association,
        Business and Economic Statistics Division.

2. Bogue, M. and R. Roll, "Capital Budgeting of Risky Projects
        With 'Imperfect' Markets for Physical Capital,"
        Journal of Finance, May 1974.

3. Brennan, M. J., "An Approach to the Valuation of Uncertain
        Income Streams," Journal of Finance, June 1973.

4. Constantanides, G. M., "Market Risk Adjustment in Project
        Valuation," Working Paper, Carnegie-Mellon University,
        September 1976.

5. Fama, E. F., "Risk-Adjusted Discount Rates and the Cost of
        Capital in a Two-Parameter World," Working Paper,
        European Institute for Advanced Studies in Management,
        April 1976.

6. Hamada, R. S., "Multi-Period Capital Asset Prices in an
        Efficient and Perfect Market," unpublished ms., University
        of Chicago, 1972.

7. Myers, S. C., "The Relationship Between Real and Financial
        Measures of Risk and Return," in J. Bicksler and I. Friend,
        eds., Studies in Risk and Return, Ballinger, 1976.

8. Myers, S. C. and S. M. Turnbull, "Capital Budgeting and the Capital
        Asset Pricing Model: Good News and Bad News," Journal of
        Finance, May 1977.

9. Rubenstein, M., "Valuation of Uncertain Income Streams and the
        Pricing of Options," Bell Journal of Economics, Autumn 1976.

10. Turnbull, S. M., "Market Value and Systematic Risk," Journal
        of Finance, forthcoming.

11. Merton, R. C., "An Intertemporal Capital Asset Pricing Model,"
     Econometrica, September, 1973.

12. Ross, S. A., "The Arbitrage Model of Capital Asset Pricing,"
     Journal of Economic Theory, December 1976.

13. Treynor, J. and F. Black, "Corporate Investment Policy," in
     S. C. Myers, ed., Modern Developments in Financial
     Management, Praeger, 1976.

Stock-Price Jumps and Covenant-Induced Safety Barriers:

The Perpetuity Case

Sudipto Bhattacharya

I

In their work on the effects of bond indenture provisions
Black and Cox (1) dealt with a case in which bond indentures
required the stockholders to give up the firm to the bondholders
when firm value fell below a certain floor, even though no
actual defaults on repayments due had occurred. The analysis
was done assuming that the firm value followed a "lognormal"
diffusion process so that changes in stock prices were continuous
with probability one, and one of the questions posed (and left
unanswered) was whether or not bond value solutions would have
similar behavior if the firm value could have large, discontinuous
jumps, particularly in the neighborhood of the "safety barrier,"
so that firm value, and hence the pyaoff to bondholders, could
indeed fall below the covenant-induced safety barrier. In this
paper, we analyze a simple case of safety barriers with jumps in
stock prices possible -- the case where the debt is a perpetual
consol bond. While no great independent interest attaches to
this case, it does illustrate some of the technical issues relating
to the solution of problems of this kind. In section II we set
out the model, and discuss some of the difficulties associated
with the general problem. In section III a particular case with

a simple structure, the case in which total payout by the firm is
proportional to its value, is discussed in greater detail.

II

The value of the firm is assumed to follow a combination
of a lognormal diffusion and a Poisson jump process. We shall
only consider jumps that cause known, constant proportionate
changes in firm value -- since that turns out to be difficult
enough, and amply serves to illustrate the technical issues
involved. Let the variance of the lognormal diffusion process be
σ^2, and let the jump component have frequency parameter λ and
proportionate jump magnitude $Y < 1$, and let $K = (Y-1)$. Let
the bond be a consol with promised payments C per unit time,
paid continuously, and, initially, assume that there are no other
payments into or out of the firm. Let $B < C/r$, where r is
the riskless rate of interest, be the safety barrier. Then,
proceeding along the lines of Merton (3), and, in particular,
assuming that the jump risk is diversifiable ("non-systemmatic"),
it is easy to obtain the following differential equation for the
bond value function $F(V)$, as a function of firm value V. In
region II, where a jump would not lead to crossing the safety
barrier,

$$1/2 \ \sigma^2 V^2 F'' + ((r-\lambda K)V - C)F' - rF + C + \lambda[F(YV) - F(V)] = 0 \quad (1)$$

Obviously, this region is defined by $V > B/Y$. In region I, where a jump would lead to crossing the safety barrier, the equation satisfied is

$$1/2 \ \sigma^2 V^2 F'' + ((r-\lambda K)V - C)F' - rF + C + \lambda[YV - F(V)] = 0 \quad (2)$$

Of course, $F(B) = B$, and $\lim\limits_{V \to \infty} F' = 0$.

Further, to rule out "almost certain" arbitrage, subject to non-systemmatic jump risk only, the two solutions and their first derivatives must coincide at $V = B/Y$.

We run immediately into the first technical problem. Consider equation (1) and a point in region II. It is perfectly possible that $V' = YV$ is in region (I) and thus methods similar to the elegant induction argument of Merton ((3), equation A(6)) cannot be made use of to solve the mixed differential-difference equation. That is the first interesting difficulty that one runs into, and for the time being we choose to wriggle out of it by making the following somewhat artificial assumption, that jumps in stock prices take place, because of arrivals of large blocks of information, only when the firm value is low. In particular we assume that this only happens in region I, and in

all of region I, so that whenever jumps take place the safety

barrier is crossed with certainty. Clearly, "only" the latter

part of the assumption is important. The solution in region

II is to the equation corresponding to pure lognormal diffusion

$$1/2 \ \sigma^2 V^2 F'' + [rV-C]F' - rF + C = 0 \qquad (3)$$

An acceptable solution, which satisfies $\lim_{V \to \infty} F' = 0$ is given by

$$F(V) = \frac{C}{r} + A \left[\frac{C}{\sigma^2 V} \right]^{\frac{2r}{\sigma^2}} M \left[\frac{2r}{\sigma^2}, \ 2 + \frac{2r}{\sigma^2}, \ \frac{-2C}{\sigma^2 V} \right]$$

where M(....) is the confluent hypergeometric function, and A is

an arbitrary constant. Let us denote this by $F_2(V)$. In the

lower region $V < B/Y$ the solution is to the equation

$$1/2 \ \sigma^2 V^2 F'' + [(r-\lambda K)V - C]F' - (r+\lambda)F + C + \lambda YV + 0 \qquad (4)$$

A particular solution is V since

$$[(r-\lambda K)V - C] - (r+\lambda)V + C + \lambda YV = -\lambda(K+1)V + \lambda YV = 0$$

We are left with the homogeneous part

$$1/2 \ \sigma^2 V^2 F'' + [(r-\lambda K)V - C]F' - (r+\lambda)F = 0$$

and this can be transformed to Kummer's equation. Let us consider

transforming an equation of the form

$$\frac{1}{2} \sigma^2 V^2 F'' + [xV - C]F' - yF = 0$$

to Kummer's equation for $h(Z)$

$$Zh'' + (b-Z)h' - ah = 0$$

by the transformation $Z = m/V$ and $F(V) = Z^n \exp(-Z)h(Z)$. We get

$$F' = -e^{-Z} \frac{m^2}{V^2} [\{nZ^{n-1} - Z^n\}h + Z^n h'] - \frac{e^{-Z}}{m} [\{nZ^{n+1} - Z^{n+2}\}h + Z^{n+2}h']$$

and,

$$F'' = e^{-Z} \frac{Z^2}{m^2} [\{n(n+1)Z^n - (n+2)Z^{n+1} - nZ^{n+1} + Z^{n+2}\}h +$$

$$\{(2n+2)Z^{n+1} - 2Z^{n+2}\}h' + Z^{n+2}h'']$$

Hence, collecting terms in h, h', h'' we have

$$h''(\frac{Z^2}{m^2} \cdot \frac{\sigma^2 V^2}{Z} \cdot Z^{n+2}) + h'(\frac{Z^2}{m^2} (2n+2)Z^{n+1} - 2Z^{n+2})\frac{\sigma^2 V^2}{2} + (xV-C) \cdot \frac{-1}{m} \cdot Z^{n+2})$$

$$+ h(\frac{\sigma^2 V^2}{2} \cdot \frac{Z^2}{m^2}\{n(n+1)Z^n - (2n+2)Z^{n+1} + Z^{n+2}\} + (xV-C) \cdot \frac{-1}{m}(nZ^{n+1}-Z^{n+2}) - yZ^n)$$

$$= 0$$

Dividing through by ($\frac{\sigma^2}{2} \cdot Z^{n+1}$) we get that in the bracket

multiplying h, for (i) terms in Z^{-1} to cancel to zero we must have

$$n(n+1) - \frac{2x}{\sigma^2} \cdot n - \frac{2y}{\sigma^2} = 0$$

or

$$n = \frac{-(\sigma^2-2x) \pm \sqrt{(\sigma^2-2x)^2+8y\sigma^2}}{2\sigma^2}$$

in our case x = (r-λK) and y = (r+λ), and (ii) for terms in Z' to
cancel

$$1 - \frac{2C}{\sigma^2 m} = 0 \qquad \text{or} \qquad m = (\frac{2C}{\sigma^2})$$

and thus the equation becomes

$$Zh'' + [(2n+2 - \frac{2x}{\sigma^2}) - Z]h' - [2n+2 - \frac{2x}{\sigma^2} - n]h = 0 \qquad (5)$$

The usual solution (e.g., Merton (4)) uses only Kummer functions

of the first kind, since that is sufficient to satisfy the boundary

conditions. Here, however, we have three conditions to satisfy

and hence the solution in the lower bounded region I involves

terms in Kummer functions of both the first and second kind. Thus

the solution in the lower region can be written as

$$F_1(V) = V + \left(\frac{2C}{\sigma^2 V}\right)^n \exp\left(\frac{-2C}{\sigma^2 V}\right) \left[D \cdot M(n + 2 - \frac{2x}{\sigma^2}, \; 2n + 2 - \frac{2x}{\sigma^2}, \; \frac{2C}{\sigma^2 V}) \right.$$
$$\left. + F\left(\frac{2C}{\sigma^2 V}\right)^{(-1-2n+\frac{2x}{\sigma^2})} M\left(1-n, \; -2n + \frac{2x}{\sigma^2}, \; \frac{2C}{\sigma^2}\right) \right]$$

(6)

where D and F are arbitrary constants. The formulas are simplified by making use of the wellknown recurrence relationship for M viz.,

$$M(a,b,Z) = e^Z M(b-a,b,-Z)$$

to give

$$F_1(V) = V + D\left(\frac{2C}{\sigma^2 V}\right)^n M\left(n, \; 2n+2 - \frac{2x}{\sigma^2}, \; \frac{-2C}{\sigma^2 V}\right)$$
$$+ F\left(\frac{2C}{\sigma^2 V}\right)^{(-1-n + \frac{2x}{\sigma^2})} M\left(-1-n + \frac{2x}{\sigma^2}, \; -2n + \frac{2x}{\sigma^2}, \; \frac{-2C}{\sigma^2 V}\right) \quad (6a)$$

the arbitrary constants A, D and F are determined by the relationships

$$F_1(B) = B \tag{7a}$$

$$F_1(B/Y) = F_2(B/Y) \tag{7b}$$

$$F_1'(B/Y) = F_2'(B/Y) \tag{7c}$$

to give the complete solution. The mechanics of satisfying these
is simplified by making use of the recurrence relationship

$$M'(a,b,Z) = \frac{a}{b} M(a+1,b+1,Z)$$

In the next section we go on to a consideration of how
the overlapping regions difficulty (alluded to above) may be
handled, manageably, in the case where total payouts from the
firm are proportional to firm value.

III

Here we consider the case where jumps can occur in every
region -- although the magnitude of the jump is still known with
certainty. The overlapping regions problem mentioned in the
previous section has to be faced, and the way to proceed is
iteratively, by breaking up the positive half of the real line
$V > B$ into regions defined by

Region   0 $B \leq V \leq B/Y$

            1     $B/Y \leq V \leq B/Y^2$

            2     $B/Y^2 \leq V \leq B/Y^3$

and so on, so that, conditional on a jump in region i, the firm
value always ends up in region (i-1), and crosses the safety barrier
if i = 0. To vastly simplify the mechanics of the calculations
involved, and also probably in greater accord with real life
practice, we assume that total payouts (to debt and equity
holders) is a proportion P of firm value. We then proceed with
the iterative procedure. In region 0, the differential equation
for F(V) is given by

$$1/2 \ \sigma^2 V^2 F_{VV} + aVF_V - rF + C + \lambda[YV - F] = 0 \tag{8}$$

where $a = (r-\lambda K-P)$. A particular solution is $F_0 = b_0 V + \dfrac{C}{(r+\lambda)}$ where

b_0 is such that $ab_0 - (r+\lambda)b_0 + \lambda Y = 0$, or $b_0 = \dfrac{\lambda Y}{r+\lambda-a} = \dfrac{\lambda Y}{\lambda Y+P} \tag{8a}$

The solution to the homogeneous part is $A_0 V^m$ where A_0 is an
arbitrary constant, and m is the root of the quadratic equation

$$1/2 \ \sigma^2 m(m-1) + am - (r+\lambda) = 0$$

or

$$m = \frac{-(a-\frac{1}{2}\sigma^2) \pm \sqrt{(a-\frac{1}{2}\sigma^2)^2 + 2\sigma^2(r+\lambda)}}{\sigma^2} \tag{8b}$$

In general different values of A_0 attach to the different roots

of m, if both roots are admissible.

Consider the equation in region I now

$$\frac{1}{2} \sigma^2 V^2 F_{VV} + aVF_V - rF + C + \lambda[F_0(YV) - F] = 0$$

or,

$$\frac{1}{2} \sigma^2 V^2 F_{VV} + aVF_V - (r+\lambda)F + C(1+\frac{\lambda}{r+\lambda}) + \lambda b_0 YV + \lambda A_0 (YV)^m = 0 \quad (9)$$

Consider a particular solution

$$F_1 = b_1 V + \frac{C}{(r+\lambda)} (1 + \frac{\lambda}{r+\lambda}) + d_1 (\log V) V^m$$

b_1 must satisfy $\quad ab_1 - (r+\lambda)b_1 + \lambda b_0 Y = 0$

or

$$b_1 = \frac{b_0(\lambda Y)}{r+\lambda-a} = b_0^2 \quad (9a)$$

Now,

$$(\log V \cdot V^m)' = m \log V \cdot V^{m-1} + V^{m-1}$$

$$(\log V \cdot V^m)'' = m(m-1)\log V \cdot V^{m-2} + mV^{m-2} + (m-1)V^{m-2}$$

Hence, d_1 must satisfy

$$d_1[1/2\ \sigma^2(2m-1) + a] + \lambda A_0 Y^m = 0$$

or,

$$d_1 = \frac{-\lambda A_0 Y^m}{[\frac{1}{2}\sigma^2(2m-1)+a]} \tag{9b}$$

Homogeneous solutions are, of course, still given by $A_1 V^m$, where

A_1 is the arbitrary constant(s) for region I, again different

for the two different roots of m, in general.

Before we move on to solutions in regions "further out" it

is useful to see if any simplification can be achieved by having

to consider only one of the roots of m -- for, otherwise, the

exercise of "boundary matching" is likely to be sheer tedium[F].

For this purpose, we consider the problem of matching the solutions

F_0 and F_1 at the boundary of regions zero and one viz., $V = B/Y$.

We also assume that P is large enough so that we can afford to

ignore another type of phenomenon that occurs in proportioate

payout models viz. that stockholders (optimally) give up the firm

to the bondholders at some stage rather than pouring in more money.

This is ensured by having P high enough so that $P \cdot B > C$, so that

stockholders are never actually pouring in money into the firm, but total payout is proportional to firm value.

Suppose (the solution corresponding to using only one of the roots of m is sufficient. Then we have

$$F_0(B) = B = b_0 B + \frac{C}{(r+\lambda)} + A_0 B^m$$

Hence

$$A_0 = \frac{(1-b_0)B - C/(r+\lambda)}{B^m} \qquad (10)$$

Now,

$$F_1 = b_o^2 V + \frac{C}{r+\lambda}(1 + \frac{\lambda}{r+\lambda}) + \left(\frac{-\lambda A_0 Y^m}{\frac{1}{2}\sigma^2(2m-1)+a} \right) \log V \cdot V^m + A_1 V^m$$

For the two solutions to be equal at $V = B/Y$ we have to have

$$F_0(\frac{B}{Y}) = b_0(\frac{B}{Y}) + \frac{C}{r+\lambda} + A_0(\frac{B}{Y})^m = F_1(\frac{B}{Y}) = b_0^2(\frac{B}{Y}) + \frac{C}{r+\lambda}(1+\frac{\lambda}{r+\lambda})$$

$$- \frac{\lambda A_0 Y^m}{\frac{1}{2}\sigma^2(2m-1)+a} \log(B/Y)(B/Y)^m + A_1(B/Y)^m$$

Hence

Hence

$$A_1 = \frac{b_0(1-b_0) \cdot \frac{B}{Y} - \frac{C}{r+\lambda}(\frac{\lambda}{r+\lambda}) + A_0(\frac{B}{Y})^m\left[1 + \frac{\lambda Y^m \log(B/Y)}{\{\frac{1}{2}\sigma^2(2m-1)+a\}}\right]}{(B/Y)^m} \tag{11}$$

Can this A_1 also produce equality of first derivatives (of F_0 and F_1) at B/Y? We have,

$$F_0' = b_0 + A_0 \cdot m \cdot V^{m-1} \qquad F_1' = b_0^2 - \left(\frac{\lambda A_0 Y^m}{\frac{1}{2}\sigma^2(2m-1)+a}\right)\{V^{m-1} +$$

$$m \cdot \log V \cdot V^{m-1}\} + A_1 \cdot m \cdot V^{m-1}$$

At V = B/Y

$$F_0' = b_0 + A_0 \cdot m(B/Y)^{m-1} = F_1' = b_0^2 + \left(\frac{-\lambda A_0 Y^m}{\frac{1}{2}\sigma^2(2m-1)+a}\right)\{(B/Y)^{m-1} +$$

$$m \log (B/Y)(B/Y)^{m-1}\} + A_1 \cdot m(B/Y)^{m-1}$$

or

$$A_1 = \frac{b_0(1-b_0) + A_0(B/Y)^{m-1}\left[m + \frac{\lambda Y^m}{\{\frac{1}{2}\sigma^2(2m-1)+a\}} (1+m \log(B/Y))\right]}{m(B/Y)^{m-1}} \tag{12}$$

Equations (11) and (12) for A_1 are reconciliable if and only if

$$b_0(1-b_0) - \frac{C}{r+\lambda} \cdot \frac{\lambda}{r+\lambda} \cdot \frac{Y}{B} + A_0 \left(\frac{B}{Y}\right)^{m-1} \left[1 + \frac{\lambda Y^m}{(\frac{1}{2}\sigma^2(2m-1)+a)} \log B/Y\right]$$

$$= \frac{b_0(1-b_0)}{m} + A_0(B/Y)^{m-1} \left[1 + \frac{\lambda Y^m}{(\frac{1}{2}\sigma^2(2m-1)+a)} \left(\frac{1}{m} + \log(B/Y)\right)\right]$$

or,

$$b_0(1-b_0)\left(1 - \frac{1}{m}\right) - \frac{C}{r+\lambda} \cdot \frac{\lambda}{r+\lambda} \cdot \frac{Y}{B} = A_0\left(\frac{B}{Y}\right)^{m-1} \cdot \frac{\lambda Y^m}{(\frac{1}{2}\sigma^2(2m-1)+a)} \cdot \frac{1}{m} \quad (13)$$

This is consistent with the value of A_0 defined by equation (10) if (13) satisfies (10). Let us check this for a simple case in which $P = 0$ and $b_0 = \frac{\lambda Y}{\lambda Y + P} = 1$ and $(1-b_0) = 0$. For consistency, we require

$$- \frac{C}{r+\lambda} \cdot \frac{\lambda}{r+\lambda} \cdot \frac{Y}{B} = A_0\left(\frac{B}{Y}\right)^{m-1} \cdot \frac{\lambda Y^m}{(\frac{1}{2}\sigma^2(2m-1)+a)} \cdot \frac{1}{m} \quad (13a)$$

or if

$$A_0 = \left(- \frac{C}{r+\lambda} \cdot \frac{1}{B^m}\right) \cdot \frac{1}{r+\lambda} \left[\frac{1}{2}\sigma^2 m(2m-1)+ma\right] = \frac{(-C/(r+\lambda))}{B^m} \text{ from (10).}$$

Things would be perfect if m satisfied $1/2\ \sigma^2 m(2m-1)+ma = r+\lambda$, but from equation (8b) m satisfies $1/2\ \sigma^2 m(m-1)+am = (r+\lambda)$. Thus

the two do not coincide unless $\sigma^2 = 0$, and except in this case

the solution to our problem has to involve both the roots of m,

and a forbiddingly tedious boundary matching operation has to

be carried out for solutions for alternative stages -- the

constants multiplying the homogeneous solutions (e.g., A_0, A_1)

can no longer be found sequentially, a simultaneous solution

becoming necessary.

In what follows, we consider the case where $\sigma^2 = 0$,

so that there is no diffusion but only jumps. This is, of course,

oversimplifying the problem, but it is somewhat instructive to

go through the solution for this case. The differential equation

for this case could also be derived by the Cox & Ross (2)

exact hedge for jumps -- and the only difference would be the

following. The λ in the equation would be that λ^* which makes

$P+\mu+\lambda^*K = r$, where μ is the drift term on the firm's dynamics --

and this λ^* would be equal to the true λ of the firm's dynamics

if and only if the total expected return is r, i.e., jump risk is

"non-systematic." In the perpetual case, with pure fixed-amplitude

jumps only, the solution can be readily written down. We are only

going to write down the solution explicitly up to region II, and

then provide a recursion relationship for solutions "further

out." In region zero the solution is

$$F_0 = b_0 V + \frac{C}{r+\lambda} + A_0 V^m \quad \text{where } b_0 = \frac{\lambda Y}{(r+\lambda)-a} = \frac{\lambda Y}{\lambda Y+P}$$

(14)

$$m = \frac{(r+\lambda)}{a} = \frac{(r+\lambda)}{(r-\lambda K-P)} \quad \text{and} \quad A_0 = \frac{(1-b_0) \cdot B - C/(r+\lambda)}{B^m}$$

In region I the solution is

$$F_1 = b_0^2 V + \frac{C}{(r+\lambda)} \left(1 + \frac{\lambda}{r+\lambda}\right) + d_1 \cdot \log V \cdot V^m + A_1 \cdot V^m$$

where

$$d_1 = \frac{-\lambda A_0 Y^m}{a} \quad \text{and} \quad A_1 = \frac{b_0(1-b_0) \cdot \frac{B}{Y} - \frac{C}{r+\lambda} \cdot \frac{\lambda}{r+\lambda} + A_0 (\frac{B}{Y})^m [1 + \frac{\lambda Y^m \log(\frac{B}{Y})}{a}]}{(B/Y)^m}$$

(15)

(In this case it is clear that A_0, A_1 can satisfy the demands for
equalization of both solution values and first derivatives at
the boundary of regions I and 0, as is easily demonstrated below.
From (13) we have,

$$b_0(1-b_0)(1-\frac{1}{m}) - \frac{C}{r+\lambda} \cdot \frac{\lambda}{r+\lambda} \cdot \frac{Y}{B} = A_0 (\frac{B}{Y})^{m-1} \frac{\lambda Y^m}{a} \cdot \frac{1}{m}$$

Substituting for m and A_0 in terms of relationships derived above
we have,

$$b_0(1-b_0)[\frac{r+\lambda-a}{r+\lambda}] - \frac{C}{r+\lambda} \cdot \frac{\lambda Y}{(r+\lambda)B} = \frac{\{(1-b_0)B - \frac{C}{r+\lambda}\}}{B} \cdot \frac{\lambda Y}{a} \cdot \frac{a}{r+\lambda}$$

or,

$$(1-b_0)\frac{\lambda Y}{r+\lambda} - \frac{C}{r+\lambda} \cdot \frac{\lambda Y}{(r+\lambda)B} = \frac{(1-b_0)\lambda Y}{r+\lambda} - \frac{C}{(r+\lambda)^2 B}\lambda Y \quad)$$

Finally, the solution in region II can be written down. In this region the equation is

$$aVF_V - rF + C + \lambda[b_0^2(YV) + \frac{C}{r+\lambda}(1 + \frac{\lambda}{r+\lambda}) + A_1(YV)^m$$

$$+ d_1 \log(YV)(YV)^m - F] = 0$$

A particular solution is

$$F_2 = \frac{C}{r+\lambda}(1 + \frac{\lambda}{r+\lambda} + \frac{\lambda^2}{(r+\lambda)}) + b_0^3 V + d_2 \log V \cdot V^m + e_2(\log YV)^2 \cdot V^m$$

where

$$d_2 = -\frac{\lambda A_1 Y^m}{a}$$

and

$$e_2 = -\frac{\lambda d_1 Y^m}{2a}$$

The homogeneous solution is still given by $A_2 V^m$ and an expression

for A_2 is easily determined by setting $F_2(\frac{B}{Y^z}) = F_1(\frac{B}{Y^z})$. From

the solutions emerging for the successive stages, it is clear

that the solution for the general n^{th} stage (i.e., for $B/(Y^n) \leqq$

$V \leqq B/(Y^{n+1})$) satisfies the following recursion relationship. Let

us write

$$F_n(V) = b_0^{(n+1)} V + \frac{C}{r+\lambda} (1 + \frac{\lambda}{r+\lambda} + \ldots \{\frac{\lambda}{r+\lambda}\}^n)$$

$$+ \sum_{i=1}^{n} d_i^n (\log Y^{i-1} \cdot V)^i V^m + A_n V^m$$

with the understanding that, for any n, $d_0^n = A_n$ and $d_i^n = 0$

for $i > n$. Then the d_i^n's are linked by the relationship.

For $1 \leq i \leq n$, $n \geq 1$

$$i\, a\, d_i^n + \lambda d_{i-1}^{n-1} Y^m = 0$$

or

$$d_i^n = \frac{-\lambda Y^m d_{i-1}^{n-1}}{ia}$$

and

$$d_0^n = A_n =$$

$$\frac{\{b_0^n(1-b_0)\frac{B}{Y^n} - \frac{C}{r+\lambda}(\frac{\lambda}{r+\lambda})^n + A_{n-1}(\frac{B}{Y^n})^m + \sum_{i=1}^{n} (d_i^{n-1} - d_i^n)(\log Y^{i-1} \frac{B}{Y^n})^i (\frac{B}{Y^n})^m\}}{(B/Y^n)^m}$$

We notice that the "key" constant term converges, as we proceed upward through the stages to $\frac{C}{r+\lambda} \cdot \frac{1}{\{1- \frac{\lambda}{r+\lambda}\}} = \frac{C}{r}$, which is comforting, since the debt becomes nearly certain as V tends to infinity.

References

1. Black, F. and J. C. Cox, "Valueing Corporate Securities:
   Some Effects of Bond Indenture Conditions," Journal
   of Finance, May 1976.

2. Cox, J.C. and S. A. Ross, "The Valuation of Options for
   Alternative Stochastic Processes," Journal of Financial
   Economics, January 1976.

3. Merton, R. C., "Option Pricing When Underlying Stock Returns
   are Discontinuous," Journal of Financial Economics, January
   1976.

4. Merton, R. C., "The Theory of Rational Option Pricing,"
   Bell Journal of Economics, Fall 1973.

IMPERFECT INFORMATION, DIVIDEND POLICY

AND THE "BIRD IN THE HAND FALLACY"

Sudipto Bhattacharya

I. <u>INTRODUCTION</u>

In this paper, we develop a model in which cash dividends function as a signal of expected earnings in an imperfect information world. The major part of the signalling cost that allows dividends to function as a signal arises because dividend payments are taxed at the ordinary income-tax rate, whereas capital gains are taxed at a lower rate. Within this framework, this paper explains why firms may pay dividends despite their tax disadvantage.

An interesting result that is suggested by the model is that the strength of motivations that are related to the desires of finite-lived agents to 'realize' their wealth (for consumption purposes) over a finite span of time plays a significant role in determining the equilibrium dividend payout. It is assumed that the productive assets in which the agents' wealth is invested stay 'in place' for a time longer than the agents' lifetime, assumed to be in perpetuity here, their ownership being transferred, over time, to other (groups of) agents. Specifically, we find that the shorter the horizons over which the (finite-lived) agents have to realize their wealth, the higher is the equilibrium proportion of dividends to earnings in the signalling equilibrium.[1/] We also explore other important comparative statics of the dividend-signalling equilibrium with respect to major variables like the personal income-tax rate and the rate of discount.

To keep our formulation and analysis manageable, and to highlight the essential characteristics, we carry out the analysis making the two major analytical simplifications that

(i)   the valuation of cash flow streams is done in a risk-neutral world.

(ii)   that, especially given (i), the 'urgency' of agents' need to real-
ize their wealth can be parameterized by the length of the planning
horizon over which the agents maximize their expected discounted
realized wealth, and that the detailed intertemporal pattern of
asset disposal can be ignored.

The meaning, restrictiveness and necessity of these assumptions are further
discussed in the next and concluding sections, after the model is set-up and
its main properties explored in the next section.

Recently, Leland & Pyle [ ] and Ross [ ] have used the paradigm of
signalling models to examine financial market phenomena related to (i) unsys-
tematic risk borne by entrepreneurs, and (ii) firm debt-equity choice deci-
sions, respectively. Besides being an application to a different issue and
a different cost-structure, our model adds some element of time-structure
and the implications and issues that arise from it to the class of financial
signalling models. In the course of this exercise, we shall have occasion
to discuss, at various points, the similarities and dissimilarities regarding
the assumptions and structure among these different applications.

II. THE MODEL

Consider a cross-section of projects owned by different firms. The
projects are all perpetuities and their mean cash flows differ, but
(outside) investors can not distinguish among them. The only people who
know the true cash flow distribution of each project are the "insiders"
for each project, whom we shall term the "management." The management's
task is to maximize the interests of the existing group of shareholders,

according to criteria that we shall discuss below. The management(s) are agents of the shareholders, and they accept this task, possibly because their own incentive compensation is tied to how well the shareholders do according to some well-defined (after tax) criterion.

Assume for the time being that the project cash flows of each project are intertemporally identically distributed. With opportunities for reinvestment of earnings available or proceeds from share repurchase taxed at the capital gains rate, the payment of dividends is costly, since dividends are taxed at a higher rate - which we take to be $(1-\alpha)$. For simplicity, it is assumed that capital gains are not taxed at all. It is assumed that a firm's promising to pay dividends D has the following implications.

(a) If earnings X are above D then D is paid (shareholders receive αD) and $(X-D)$ is invested in other projects. When we say $(X-D)$ 'is invested,' we mean $(X-D)$ of the current shareholders' money replaces (part of) what would have been other sources of financing for the firm's (other) investment opportunities. (Dividend policy is considered holding investment policy fixed.) Thus the present value - the discounted expected value - of such investment at that time is exactly $(X-D)$. For analytical simplicity, we assume that these investments are made (and can be made) in projects which are either (i) of the perfect information variety, so that rationally no dividend decision vis à vis them is needed, or (ii) the dividend decision on them is taken as given outside this model, and whatever it is, the incremental present value generated for the current shareholders is $(X-D)$. If proceeds that shareholders receive from repurchase of stock by the firm are not taxed, then we can also consider $(X-D)$ as being paid out

through stock repurchase.

(b) If earnings X are below D then dividends D are still paid, and the (D-X) deficit is raised by either

(i) selling assets in the secondary market to other firms, or

(ii) admitting new claimants to the firm.

It is assumed that either recourse has some information, organization and transaction costs that are finite, but possibly small, so that the cost to the firm is $(1+\beta)(D-X)$ where $\beta > 0$.

The cost structure has been chosen for its realism vis à vis actual dividend payment practices. Other analytical possibilities that allow dividends to function as a signal are discussed in the concluding section. We assume that transaction costs for transferring ownership claims by shareholders in the stock market are zero - and this difference with the transactions costs for firms is an important asymmetry that reflects some costs of either the

(i) transfer of operations of a real, physical asset to a different organization, or

(ii) incremental negotiation and transactions costs of obtaining immediate, "unanticipated" new financing.

The cost structure seeks to reflect the realistic notion that while earnings above dividends may be used to replace outside financing in the firm's "growth investments," the immediate nature of the deficit of earnings over dividends forces some additional, frictional costs of making up the deficit. The case for such a dissipative, frictional cost is made stronger if we assume, realistically, that part of the market conventions that surround dividend signalling disallows recourse to external financing to make

up dividend deficits, and thus the firm has to incur additional costs of either (a) selling real assets, incurring some transaction costs, or (b) postponing some investment/replacement plans, or (c) keeping liquid reserves, earning less then the discount rate, whose magnitude is some proportion of the average deficit. The basic notion is that financiers, bankers, for example, look at the dividend promise as a commitment that the firm 'should' be able to meet, and there is either non-recourse for or extra costs of financing deficits of earnings compared to dividends. This aspect of the cost-structure plays an important role. (A similar punitive cost of meeting the deficit of liquid assets compared to deposit withdrawal demands plays a critical role in bank asset management models – see e.g. Pyle [8]) (It is also clear that if there are transactions costs associated with immediate investment of the surplus of earnings over dividends or penalties associated with the retirement of external financing with the surplus, then there is a similar effect on the cost structure of creating a higher cost of a deficit compared to the benefit of a surplus of the same size.)

As we shall see, the existence of such frictional costs is necessary for a signalling equilibrium. However, these costs are likely to be small compared to the tax-loss costs of dividend payments, i.e. β is probably small compared to $(1-\alpha)$. Though these frictional costs are essential, we shall demonstrate that the tax-loss costs of dividends also play a critical role in making for a feasible signalling equilibrium.

To derive the model structure we begin by considering the case in which the current shareholders' planning horizon is one period. They plan to

realize the one-period cash flow and sell off their shares at the end of

the period, because the planning horizon corresponds to their remaining

life. Suppose V(D) is the signaled value of the project next period as a

function of dividends paid. Then if the insiders set a policy of paying

dividends D, then the total expected gain of the current shareholders

(taking gains from other parts of the firm as given) in present value terms

is

$$E = \frac{1}{1+r} \left[\int_D^{\bar{X}} [\alpha D + (X-D) + V(D)] f(x) dx + \int_{\underline{X}}^D [\alpha D + (1+\beta)(X-D) + V(D)] f(x) dx \right] \quad (1a)$$

where $\underline{X}$, $\bar{X}$, $f(x)$ represent the lower limit, upper limit, and the density

function of cash flows $\tilde{X}$ and r is the one-period riskless rate of inter-

est. Integrating by parts, (1a) can be written as

$$E = \frac{1}{1+r} [M + V(D) - (1-\alpha)D - \beta \int_{\underline{X}}^D F(x) dx] \quad (1b)$$

where $F(x)$ is the distribution function of $\tilde{X}$ and M is the mean of $\tilde{X}$.

If, cross-sectionally, the project cash flows are such that for any Z,

$F(Z)$ has the same ordering across projects, possibly weak if $F(Z)=1$, then

it is seen that the marginal signalling cost at any given level of D

$$\frac{\partial}{\partial D} [(1-\alpha)D + \beta \int_{\underline{X}}^D F(x) dx] = (1-\alpha) + \beta F(D)$$

is negatively related to the mean M of the particular cash flow stream,

and thus the potential for a signalling equilibrium exists (Spence (11))

provided that $D < \bar{X}$ in equilibrium.

This proved, consider the characteristics of the signalling equilibrium

for a particular, simple example. Let project cash flows be distributed

uniformly over $(0,t)$ with mean $\frac{t}{2}$. Although t differs in the cross-

section, between tmin = 0 and some tmax, investors cannot distinguish

that. Given a market signalling value function V(D), the current share-

holders' agents act so as to maximize

$$E = \frac{1}{(1+r)} [\frac{t}{2} + V(D) - (1-\alpha)D - \beta\frac{D^2}{2t}] \tag{2}$$

which implies the first order condition

$$V'(\overset{*}{D}) - (1-\alpha) - \beta\frac{D^*}{t} = 0 \tag{3}$$

at the optimum D* conditional on t.

A market signalling value function survives in equilibrium only if

expectations are fulfilled, i.e. that $V(\overset{*}{D})$ is the true value of future

cash flows for the firm that chooses D* as its dividend payout. In order

to impose this criterion, future levels of dividend to be paid by the firm

must be specified - because $V(\overset{*}{D})$ should only reflect the value that is

not dissipated by the tax and other loss of future dividends. In a model

with genuine time structure, this is a difficult issue to decide. On the

one hand, dividend signalling will probably have to be carried on for more

than one period to produce a feasible signalling equilibrium, i.e. one in

which the equilibrium dividends are less than t, so that marginal signal-

ling cost is still negatively related to t.[2/] The tax loss of one period

is not likely to be sufficient for "telling the truth" i.e. to produce an

interior optimum D*(t) which also satisfies the consistency condition

mentioned above. (On this point, see the detailed derivation below.) In

conjunction with the perpetuity structure of the model, which makes the

future look the same from any particular point in time, this suggests a

constant dividend policy. Against this is the argument that in a model
with genuine time structure (as opposed to idealized replications of a sta-
tic model), there should be some learning about t (from ex-post realiza-
tions) as time goes on, and thus the ability to discriminate should improve.
There appears to be no simple way to incorporate dynamic learning phenomena
into an imperfect information signalling model. However, the following
kind of realistic extension of our 'story' would make the assumption of a
cost-structure based on continued dividend payments more plausible.

Suppose we relax the condition that cash flows of projects are inter-
temporally identically distributed and assume instead that, for any project,
t follows a random walk without drift over time, so that a continuing divi-
dend signal is required in the future. Given the random walk, the expected
value of future t's is just today's t. Now consider the following con-
vergence argument. Assume for a moment that in equilibrium V(D) is linear
in D and that equilibrium D*(t) is proportional to t. Given that we
have, taking expectations with respect to any future period, that

$$\varepsilon\,[V(D)] \;=\; V(E[D]) \;=\; V(D)$$

$$\varepsilon\,[(1-\alpha)D] \;=\; (1-\alpha)D$$

$$\varepsilon\,[\beta\frac{D^2}{2t}] \;=\; \beta\frac{D^2}{2t}$$

where ε is the expectations operator (with respect to any future period)
and the right hand side terms refer to current values of the variables.
Given this, the current shareholders' objective function remains the same
as in equation (2) and the consistency condition is

$$V(D*(t)) = K[\frac{t}{2} - (1-\alpha)D*(t) - \beta\frac{D*(t)^2}{2t}] \tag{4}$$

where $K = 1/r$, and r is the riskless rate of interest. Equation (4) must hold along the equilibrium schedule $D*(t)$. Suppose we now show that, given the structure provided by equations (2) and (4), the equilibrium signal $D*(t)$ is indeed proportional to t and $V(D)$ is linear in D. Then we have an internally consistent signalling equilibrium when $t(s)$ follows a random walk. We investigate the characteristics of that equilibrium below.[3]

Equations (3) and (4) together provide us with enough information to solve for the equilibrium $V(D)$ and $D*(t)$ schedules - and to check that the second order condition for the maximization problem of equation (2) is satisfied. In the context of this model, the economic mechanism by which the consistency condition is reached is an interesting issue, whose discussion we postpone to the next section. Here we work out the detailed implications of our model.

Totally differentiating equation (4), and substituting for $V'(D)$ from equation (3) we have

$$(K+1)[(1-\alpha) + \frac{\beta D}{t}]\frac{dD}{dt} = K[\frac{1}{2} + \frac{\beta D^2}{2t^2}] \tag{5}$$

as the equation that must be satisfied along the equilibrium schedule. Since we have assumed that, cross-sectionally, $tmin = 0$, the boundary condition to equation (5) for the surviving Pareto-superiod signalling schedule (as in Riley [9]) is $D*(0) = 0$, i.e. the "lowest member has no requirement for dissipative signalling.[4] To solve (5), try a solution of the form

$$D^*(t) = At \tag{6a}$$

Substituting from (6a) into (5), we derive the quadratic equation for A

$$\frac{(K+2)}{(K+1)} A^2 + \frac{2(1-\alpha)}{\beta} A - \frac{K}{\beta(K+1)} = 0$$

Solving for A, we have that

$$A = \frac{\frac{-2(1-\alpha)}{\beta} \frac{(K+1)}{(K+2)} \pm \sqrt{\frac{4(1-\alpha)^2}{\beta^2} \frac{(K+1)^2}{(K+2)^2} + \frac{4K}{(K+2)}}}{2}$$

or, since only the positive root for A is relevant

$$A = \frac{-(1-\alpha)}{\beta} \frac{(K+1)}{(K+2)} + \frac{(1-\alpha)}{\beta} \frac{(K+1)}{(K+2)} \sqrt{1 + \frac{\beta K(K+2)}{(1-\alpha)^2 (K+1)^2}} \tag{6b}$$

Given equation (3), together with the boundary condition $V(0) = 0$, equilibrium $V(D)$ is then given by

$$V(D) = ((1-\alpha) + \beta A)D \tag{7}$$

This low response of $V(D)$ to D occurs because in the one-period horizon case a tremendous dissipative loss is required for signalling. This low response is less pronounced in the multiperiod horizon case considered later. It is easy to check that, given this $V(D)$, the first order condition is satisfied at $\frac{D}{t} = A$ and that the second order condition for a maxima is satisfied too for $(\beta, A) > 0$. If $\beta << (1-\alpha)^2$ then the solution for A can be approximated through the usual Binomial expansion as

$$A \simeq \frac{(1-\alpha)}{\beta} \frac{(K+1)}{(K+2)} \frac{1}{2} \frac{\beta K(K+2)}{(1-\alpha)^2 (K+1)^2} = \frac{1}{2} \frac{K}{(K+1)(1-\alpha)} \tag{6c}$$

As is to be expected from the continuity of solutions in parameters, the equation (6c) solution for A is a 'candidate solution' when $\beta=0$ which satisfies the first order and consistency conditions, but it does not satisfy strict maxima conditions for the firms - given the 'equilibrium' $V(D) = (1-\alpha)D$ firms would be indifferent about the levels of dividends they signal. Thus the dissipative cost of financing 'deficits' is essential for a valid solution. In practice, β is likely to be small and the resulting convexity is likely to be 'weak'. The implications of this are further discussed in the next section.

Several things can be noted about the solution. First, no matter how large K is, signalling is feasible (i.e. A<1) as long as tax rate $(1-\alpha)$ $> \frac{1}{2}$. This is a sufficient condition, and the exact magnitude of the tax rate needed for feasible signalling will depend on the distributional characteristics of the cross-section of projects. Second, the equilibrium A - the fraction of t promised as dividends - is a decreasing function of the personal tax rate $(1-\alpha)$. This is seen trivially for the approximate solution (6c). For the exact solution (6b), note that with parameters $C > 0, d > 0$

$$\frac{\partial}{\partial(1-\alpha)}\left[(1-\alpha)C\sqrt{1 + \frac{d}{(1-\alpha)^2}} - (1-\alpha)C\right]$$

$$= C\sqrt{1 + \frac{d}{(1-\alpha)^2}} - C - \frac{1}{2}\frac{(1-\alpha)C}{\sqrt{1 + \frac{d}{(1-\alpha)}}}\frac{2d}{(1-\alpha)^3}$$

$$= \frac{C + \frac{Cd}{(1-\alpha)^2} - C\sqrt{1 + \frac{d}{(1-\alpha)^2}} - \frac{Cd}{(1-\alpha)^2}}{\sqrt{1 + \frac{d}{(1-\alpha)^2}}} < 0$$

Third, from equation (6b) it follows that, because both $\frac{K+1}{K+2}$ and $\frac{K(K+2)}{(K+1)^2}$

are increasing functions of K, A is an increasing function of K. There

is one straight-forward way to interprete this viz. that A, the 'payout

fraction' is a decreasing function of the interest rate r. On a note of

casual empiricism, both these comparative statics are in accord with the

empirical results of Brittain's [ ] comprehensive study of dividends. The

response of dividend payments to a change in the interest rate, has not to

our knowledge received strong theoretical support in previous research

(See e.g. Pye [7]). Attempts to 'explain' it by arguing that debt is more

'expensive' when interest rates are high, and thus internal financing is

increased, are in conflict with the Modigliani-Miller [ ] theorem on lever-

age indifference - or possible strict preference for debt in an after-tax

world.$\underline{5/}$ In the context of our model, the intuition behind these results

is as follows. In the first case, a lower proportionate payout has a suf-

ficient dissipative cost to "enforce telling the truth" about value because

the tax rate is higher. In the second case, a higher relative weight on

the (value of) future cash flows compared to the dissipative tax loss of

one-period dividends that arises from a rise in K requires a higher (equi-

librium) payout "to tell the truth." The intuition behind these two results

seems robust enough, and the same characteristics are displayed by the

(asymptotic) solutions of the examples that we discuss in the Appendix. It

can also be shown that $\partial A/\partial \beta < 0$, as would be expected from the similar

comparative static with respect to $(1-\alpha)$.

The comparative static with respect to K does, further provide some

support for the effects (as opposed to the reasoning) of the older "bird in

the hand" notions. The essence of the "bird in the hand" notion is that the investors' desire to realize their wealth (for consumption) makes dividends constructively "different" from capital gains. What our model suggests is that if, ceteris paribus, the "urgency" of the investors' need to realize their wealth diminishes (as measured by an increase in the investors' planning horizon), then in the signalling equilibrium the equilibrium proportion of dividends paid out (and thus the dissipative tax loss) diminishes. Mathematically, this is seen as follows. Suppose investors have an (n+1) period planning horizon so their (and their agents') objective function is given by

$$(8) \quad E = \frac{1}{1+r} \left[(1 + \frac{1}{1+r} + \ldots \frac{1}{(1+r)^n}) (\frac{t}{2} - (1-\alpha)D - \frac{\beta D}{2t}) + \frac{1}{(1+r)^n} V(D) \right]$$

Now, let $\mu = (1 + (1+r)^2 + \ldots (1+r)^n)$

Then, the equation corresponding to equation (5) for the equilibrium $D*(t)$ schedule can be written as

$$(K+\mu) \left[(1-\alpha) + \frac{\beta D}{t} \right] \frac{dD}{dt} = K \left[\frac{1}{2} + \frac{\beta}{2} \frac{D^2}{t^2} \right]$$

Now $\frac{K}{(K+\mu)} = \frac{K*}{K*+1}$ for some $K*<K$ if $\mu>1$, and $K*$ declines with a rise in μ caused by an increase in n. Thus the effect of a longer horizon is to lower the "effective K" of the equation, and thus to reduce <A> of the equilibrium $D*(t) = At$ solution. The intuition is as follows: As the planning horizon becomes longer, the relative weight of the interim cash flows (paid out or reinvested) goes up and that of the end-of-horizon 'return of capital' goes down, in the investors' objective function. Given this, a given fractional dividend and thus dissipative loss hurts the

investors more. Therefore, a lower fractional dividend (and tax loss) is required in order to have current shareholders' agents accurately signal the true expected cash flows and thus the true value of the firm. The intuition behind the result, again, seems to be quite robust. It is also straight-forward to show that in this case, in equilibrium

$$V(D) = \mu[(1-\alpha) + \beta A]D \qquad (9)$$

Comparison of equation (9) with equation (7) demonstrates that the low response of $V(D)$ to D in the one-period planning horizon model disappears in the multi-period case. This makes it more plausible that deficits $(D-X)$ can be raised by selling assets. It can also be shown that, when $\beta \longrightarrow 0$, the equilibrium value of the firm is given by

$$V(D) = \frac{Kt}{2}\left[1 - \frac{K}{(K+\mu)}\right] \qquad (10)$$

$$= \frac{Kt}{2}\left[\frac{\mu}{K+\mu}\right]$$

For large horizons, μ may be larger then $K=1/r$, and thus less than half the value may be lost dissipatively.[6/] Our conclusions on these points are sensitive to the assumptions about the cash flow distributions and the cross-section of "indistinguishable" projects. For example, the dissipative loss needed for signalling would decrease as $tmin$ increases from zero towards $tmax$.

However, multi-period horizons raise many new issues, especially if this structure is incorporated into a general equilibrium model of the overlapping generations type. To catalog some of these, we note the following points.

(a) It is not clear how effective the length of the planning horizon
is in capturing the (lack of) "urgency" of the investors' need to realize
their wealth. It works literally only if we assume that either investors
care only about their terminal wealth, or, if interim consumption withdraw-
als are to be permitted, that investors can finance their interim consump-
tion with consumption loans. These loans have to be repaid at the end of
their planning horizon, and lenders in these markets use the information
about the (signalled) values of the investors' assets to set the competitive
promised rate of interest on the loans. If, instead, we assume that inves-
tors dispose of portions of their assets over the horizon, then some com-
plexities arise. It is clear that if we assume a fixed pattern of asset
disposal (e.g. $\frac{1}{n}$ fraction over each period), then a result of the same type
is going to be obtained. However, it is likely that the pattern of disposal
of the project would respond to ex-post capital gains and dividends.

(b) It is difficult to accommodate individuals with different horizons
simultaneously. Shareholders with different horizons would prefer different
dividend decisions when faced with the same V(D). It is not clear what the
optimal decision rule for the management is. Further, with different inves-
tor horizons it is not easy to interpret the comparative static with respect
to the horizon as applying across firms that have different horizon 'clien-
teles' unless different clienteles are irrevocably attached to different
types of firms. The comparative static for the planning horizon is thus
better interpreted in the context of two (alternate) economies with agents
having different (but homogeneous) time-lengths of asset accumulation and
use for consumption, with all other things like rate of time-preference

kept constant.

(c) If shareholders have multi-period horizons, then these are some
conceptual problems if security markets are to be active at 'intermediate'
points. As calendar time progresses, the 'time left' will be less than the
shareholders' full horizon. It could be argued that dividend setting
should take this into account and, conditional on the same t, D should
change as calendar time passes. However, suppose markets are open and
there are other people around who have freshly come into the asset owning
phases of their life, and thus they have longer horizons. For projects that
they start with savings from the accumulation phase of their lifetimes, they
would, for the same level of t, react with a lower D to the same V(D)
function, and no V(D) function can simultaneously satisfy the first-order
and consistency conditions for them and the 'older' people with 'whittled
down' horizons. This is resolved only if either

    (i) 'market conventions' do not allow changes in dividend policy in
or   the interim, and only allow 'stationary' dividend policies

    (ii) the notion of 'finite life-time' of current shareholders is cap-
        tured not through truly finite life-times, but through exponenti-
        ally decreasing holdings of the project - and the comparative
        statics are done across different rates of decline.

(d) A more mundane complication that arises is the following. We have
assumed that one way a dividend 'deficit' is made up is through the sale of
the project assets. To keep this consistent with equation (8) one must
assume that either

(i) the firm has sufficient other assets to sell if necessary in the interim, or

(ii) assets are sold in the secondary market to other firms which (have surplus cash and) can appraise and will buy the project (or any fraction thereof) for the true worth over the remaining part of the horizon, and not necessarily at $V(D)$, (dissipative costs $\beta(D-X)$ aside). In equilibrium, this proviso is redundant, since $V(D)$ is consistent.

While many such problems of detail arise, the basic intuition of our result relating the length of planning horizon to the equilibrium proportion of dividends should be robust. But sufficient resolution of these problems to permit incorporation of this model into a richer general equilibrium model is beyond the scope of this paper.

In the next section we go on to consider two important points. First, we discuss what mechanism might lead to the equilibrium consistency condition [Eqn. (4)] being realized. Second, this model is compared with other financial signalling models in the literature.

III. CONCLUDING REMARKS

The primary issue we tackle here is that of the economic mechanism that allows the consistency condition of equation (4) to be reached. This is also related to the issue of the "stability" of the signalling equilibrium despite the fact that the convexity provided by β, which arises from the dissipative costs of making up dividend deficits, is likely to be 'weak'. The crux of the matter lies in the fact that there is one critical difference in the time-structure of events between financial signalling models of this type and the job market signalling models (Spence [11]). In the latter,

the actions that result in the signalling cost (e.g. education) are taken

before the rewards (wages). In the former, the actions resulting in the

signalling costs are promises (e.g. dividends, "agreed upon" bankruptcy

penalties) that have results that get reflected in market values now. (The

model of Leland & Pyle [4] is an exception to this.) Let us take a concrete

example. In the model of the previous section, conditional on the equili-

brium $V(D)$, the current shareholders, or their agents, maximize their one-

or multi-period objective by setting $D = At$. However, what mechanism is

there to prevent them from setting a higher D (dividend promise), gener-

ating a $V(D)$ that is higher than the sustainable value of the firm given

that level of continued dividends, and selling out now? In the context of

a tâtonement model of reaching equilibrium, there is an answer to this ques-

tion. Given the communication of a higher D and an intention to sell out

now, the market will lower its $V(D)$ schedule, since it is aware that such

a bid can only arise from an overvalued firm. A far-sighted management will

recognize that such bids can not help its objective function for the one-

(or multi-) period problem and thus optimize and bid with respect to its

'true' horizon. In a realistic non-tâtonement model that allows trading

(at possibly disequilibrium prices) at all times, we have to give up the

'convenient fiction' of price-taking, and will have to assume that an

attempt to sell out (an overvalued firm) or an attempt by insiders to trade

more than what is warranted by normal levels of 'retiring' shareholders

will have an effect on market prices, and specifically on the market $V(D)$

schedule.[7] Thus it is likely that the time-structure of 'confirming' the

consistency condition will also be reversed from that in the labor market

signalling model, where such convergence is reached by employers revising

the estimates of productivity conditional on level of education ex-post - in

financial signalling models such an approach to equilibrium and 'confirma-

tion' will likely arise more from prior bids and attempts to sell. This is

a reason why the model may be "stable" even with the low degree of convexity

provided by β - since an attempt to set higher dividends D will also

trigger earlier selling by insiders. These issues need further study,

however.

About our simplifying assumption of valuation in a risk-neutral world,

we have little to say. The assumption simplifies the tasks of valuing the

"truncated" part of the signalling costs. It makes it possible to decide

on the optimality of the objective function (of realized expected discounted

wealth), and it makes it somewhat easier to ignore the role of intra-horizon

trading by shareholders in the multi-period set-up. Difficult problems

arise in the resolution of these three issues (and possibly others that we

have ignored) in a more realistic risk-averse world.

Lastly, a word on the cost-structure adopted here. Analytically, other

possibilities exist. For example, a 'truncated' signalling cost-structure

in which earnings X are paid as dividends if X is below D, but no

dividends are paid if X is above, would provide a signalling equilibrium

with the same qualitative properties as here.[8/] Ross [10] works out a sig-

nalling model of the debt-equity decision which employs a truncated cost

structure based on significant "bankruptcy penalties" for managers. The

problem with such a structure is that when enforceable penalties of similar

magnitude (relative to the rewards of non-bankruptcy) are not available for

shareholders, assuming that signalling can be obtained through managers ignores the fact that shareholders have a (large) incentive to make side-payments to managers to signal falsely about value by employing higher levels of debt. In the case of dividends, where the tax costs of signalling "are there," such a truncated structure is certainly feasible. One reason it may not be employed is that it diminishes the effectiveness of dividends as a 'verifier' of ex-post earnings (!) - in our structure, the fact that firms can meet the dividend payments, either from earnings or through sales of some fraction of assets in the secondary market, is some indication of the earnings. In our model, the truncated part of the cost-structure, that due to dissipative costs of making up dividend deficits, provides the requisite convexity - since marginal signalling costs are negatively related to mean project cash flow. But the non-truncated part, due to tax losses of dividend payments, contributes to the existence of an equilibrium signalling schedule i.e. one where $A<1$ in $D^*(t) = At$. In addition, other examples of "non-dissipative" signalling cost-structures fit a somewhat modified framework in a world without taxes in which the signalling cost arises not from 'deadweight costs' like taxes, but from market accepted expectation revision structures, conditional on payment or non-payment of dividends, from ex-post earnings alone, for example.$\underline{9/}$

The reader is warned that many of the conclusions are still speculative and much remains to be resolved about models in this area. Clearly, imperfect information assumptions are interesting and worth exploring in financial research. It is to be hoped that this paper contributes some fuel to that discussion.

APPENDIX

In this section we derive the signalling equilibrium solution for an example in which, cross-sectionally, the cash flow distributions are bounded by the same upper and lower bounds but their means differ because of differences in the distribution function between the bounds.

Consider the family projects, with distribution functions given by

$$F(x) = x - \delta \sin(\pi x) \qquad\qquad x\varepsilon(0,1)$$

so that density is

$$f(x) = 1 - \pi\delta \cos(\pi x)$$

and $0 \le \delta \le \frac{1}{\pi}$, where δ is the parameter that varies across members of the family.

The mean cash flows are given by

$$M(\delta) = \int_0^1 x[1 - \pi\delta \cos(\pi x)] \, dx$$

$$= \frac{1}{2} + \int_0^1 \delta \sin \pi x \, dx$$

$$= \frac{1}{2} - \frac{\delta}{\pi} \cos \pi x \left.\vphantom{\int}\right|_0^1$$

$$= \frac{1}{2} + \frac{2\delta}{\pi}$$

For the one-period signalling problem, the insiders' objective function is given by

$$E = V(D) \left[\frac{1}{2} + \frac{2\delta}{\pi}\right] - (1-\alpha)D - \beta \int_0^D [x - \delta\sin\pi x]dx$$

$$= V(D) + \frac{1}{2} + \frac{2\delta}{\pi} - (1-\alpha)D - \beta\left[\frac{D^2}{2} + \frac{\delta}{\pi}(\cos\pi D - 1)\right]$$

So the first order condition for the optimum D is

$$E_D = V' - (1-\alpha) - \beta[D - \delta\sin\pi D] = 0$$

The consistency condition is

$$V(D) = K\left[\frac{1}{2} + \frac{2\delta}{\pi} - (1-\alpha)D - \beta\left\{\frac{D^2}{2} + \frac{\delta}{\pi}(\cos\pi D - 1)\right\}\right]$$

Totally differentiating the consistency condition and substituting the first order condition we get

$$(K+1)[(1-\alpha) + \beta(D \delta\sin\pi D)] \frac{dD}{d\delta} = K\left[\frac{2}{\pi} - \frac{\beta}{\pi}(\cos\pi D - 1)\right]$$

with the boundary condition $D(\delta=0) = 0$ i.e. no dissipative signalling for the lowest mean member. An exact solution is difficult here, but we can look at the solution for $\beta=0$ (which, as in the text, does not satisfy the second order conditions), knowing that similar comparative statics go through for β in some neighborhood of zero. For $\beta=0$ the solution is

$$D(\delta) = \frac{2}{\pi} \frac{K}{(K+1)(1-\alpha)} \delta$$

and

$$V(D) = \frac{K}{2} + (1-\alpha)D$$

We see, again, that $D(\delta)$ is increasing in K and decreasing in $(1-\alpha)$. The proportionate dissipative loss needed for signalling is much lower here, because $Mmin = \frac{1}{2}$.

FOOTNOTES

1. The older 'bird in the hand' argument that (essentially) agents have to realize their wealth, and that somehow, dividends are 'superior' to capital gains (which overcomes their tax disadvantage) is, of course, fallacious in a perfect information, competitive financial market - even under uncertainty. For a proof, refer to Miller & Modigliani [6].

2. The concept of a 'feasible' dividend signal is much more clearcut when, say, cash flows are bounded by the same limits in the cross-section, but the means differ because of differences in the distribution function. Learning is also likely to be slower in such a case. In the Appendix we provide an example of such a case, and show that it has similar characteristics as the solution for this case.

3. The specification of the consistency condition would be easier in a model in which the project has a finite life, and the current share-holders are the only ones who sell out before project maturity - the next set of shareholders need not signal. It would be somewhat messier to allow multiple generations of retiring shareholders for a finite lived project - the consistency condition for current shareholders would have to be defined recursively. With a finite-life project, the dynamic learning issue is easier to ignore.

4. If, cross-sectionally, $t_{min} > 0$ then the exact solution for a non-zero β is somewhat more difficult. But we can still take the asymptotic solutions for the $\beta = 0$ case which will be of the form $D^*(t) = -b + At$ (with $b = A t_{min}$) and show that A has similar comparative statics with respect to K and $(1-\alpha)$. By the continuity of solutions in parameters, these same comparative statics remain valid for β in some neighborhood of zero.

5. The justification of this empirical result through our model involves some important implicit assumptions viz. that it applies to changes in the interest rate, ceteris paribus, i.e. changes that have no implications for future cash flow levels. As such, it is likely to be valid only for changes in the real rate of interest. This makes the use of the results of our model to "explain" Brittain's empirical results somewhat speculative.

6. For example, with $r = .10$ per 'year,' $K = 10$, $\mu = 18.3$ for $n = 10$, or a 11 'year' horizon. Note also that if, more realistically (but with more messiness) we assumed that cash flows are distributed over a smaller range (e.g. uniformly between $t/4$ and $3t/4$) then the maximum deficit can be more easily met by selling assets. (It can be shown that the $D^*(t)$ solution is linear in t in this case too.)

7. Informational effects of prices and bids have been noted, among others, by Grossman and Stiglitz [3].

8. The sentence means just what it says! Other possibilities including a structure in which all earnings X are paid as dividend if X is below D, and D in dividends is paid otherwise or a structure in which the signal is the payout ratio - the proportion of ex-post earnings paid out—do not satisfy the critical requirement for equilibrium signalling, that (here) marginal signalling cost must be negatively related to the mean cash flow level.

9. The distinction is somewhat similar to that between 'signals' and 'contingent contracts' that is discussed in Spence [12]. The time-structure of financial signalling models often makes it difficult to have non-dissipative signalling equilibria be consistent with rational expectations about firm value at all times and contingencies. However if feasible 'contingent contracts' are available to sort out firms, we should expect the market participants to go over to that mode of distinguishing rather than do it through the dissipative signalling costs of the non-cooperative signalling equilibrium. For a discussion of these issues, see Bhattacharya [1].

REFERENCES

1. Bhattacharya, S., "Notes on a Non-Dissipative Signalling Structure," in Ph.D. Thesis, "Essays on Financial Theory," Sloan School, M.I.T. (1977).

2. Brittain, John A., "Corporate Dividend Policy," Washington, D.C.: Brookings Institution, (1966).

3. Grossman, S.J., and J.E. Stiglitz, "information and Competitive Price Systems," American Economic Review, (May 1976).

4. Leland, H., and Pyle, D.H., "Informational Asymmetries, Financial Structure, and Financial Intermediation," Graduate School of Business, University of California, Berkeley, Working Paper no. 41, (September 1976).

5. Modigliani, F., and M.H. Miller, "Cost of Capital, Corporation Finance, and the Theory of Investment," American Economic Review (June 1958).

6. Miller, M.H., and F. Modigliani, "Dividend Policy, Growth and the Valuation of Shares," Journal of Business (October 1961).

7. Pye, G., "Preferential Tax Treatment of Capital Gains, Optimal Dividend Policy and Capital Budgeting," Quarterly Journal of Economics (May 1972).

8. Pyle, D.H., "Descriptive Theories of Financial Institutions Under Uncertainty," Journal of Financial and Quantitative Analysis (December 1972).

9. Riley, J., "Competitive Signalling," Journal of Economic Theory (April 1975).

10. Ross, S.A., "The Determination of Financial Structure: the Incentive Signalling Approach," Bell Journal of Economics (Spring 1977).

11. Spence, A.M., "Competitive and Optimal Responses to Signals: Analysis of Efficiency and Distribution," Journal of Economic Theory (March 1974).

12. Spence, A.M., "Signals, Contingent Contracts, and the Divergence of Earnings from Productivity," Harvard Institute of Economic Research, Discussion Paper no. 540, (March 1977).

NOTES ON A NON-DISSIPATIVE SIGNALLING STRUCTURE[*]

Sudipto Bhattacharya

We discuss below a non-dissipative signalling structure which has the effect, in the current labour market context, of sorting out workers of different mean productivity. By a non-dissipative structure we refer to a cost-structure in which the signalling cost is not dissipative, or a dead-weight cost. The cost structure discussed here differs from that considered by Spence (6) in his recent discussion of 'contingent contracts' in three major ways. (A) in our model, workers' productivity is random and differences in mean productivity are not (independently) discovered in the second period of a two-period model. (B) The level of expected productivity, in our model, is not affected by the level of the signal (contingent contract variable) - unlike in Spence's contingent contracts model in which such an effect is essential.

(C) As a result of (A) and (B), the analytical structure of our model is quite different from Spence's 'contingent contracts' model (6); indeed it is closer to Spence's (dissipative) signalling model (5). Because of the reasons mentioned in (B), while Spence's (6) structure is non-dissipative in the sense that investment in the signal is efficient (given its productivity benefits and costs), ours is non-dissipative in the sense that the signalling 'cost' contributes 'value' to the employer, in a sense that will be made precise below.

Let us motivate the discussion with an example. Workers have productivities that are (in each period) random - distributed uniformly between 0 and t with mean $\frac{t}{2}$; t differs among workers, but employers cannot distinguish that. The following kind of contingent contract is available. A worker 'commits' a productivity of at least D to get a wage commitment

W(D). If ex-post productivity X is less than D then $\alpha(D-X)$ is deducted from W(D), either in this period's wages in the context of a one-period model or from next-period's wages in the context of a two-period model. (In the context of the two-period model it is assumed that one-period's realizations result in no significant learning about mean productivity.) If ex-post productivity is greater than D then no adjustment is made.[1/] It is assumed that workers are risk-neutral and, solely for convenience, the discount rate is assumed to be zero.

In the context of the one-period model the workers maximize

$$E(D;t) = W(D) - \int_o^D \frac{\alpha}{t} (D-X)dx = W(D) - \frac{\alpha D^2}{2t} \tag{1}$$

This implies the first-order condition

$$E_D = W_D - \frac{\alpha D}{t} = 0 \tag{2}$$

where subscripts denote (partial) derivatives.

The competitive consistency condition (that on average employees receive their mean productivity) requires that, along the equilibrium schedule,

$$W(D) - \frac{\alpha D^2}{2t} = \frac{t}{2} \tag{3}$$

Totally differentiating (3) and substituting (2) we get that along the equilibrium schedule $D^*(t)$

$$0 = \left(\frac{1}{2} - \frac{\alpha D^2}{2t^2} \right) \cdot \frac{dt}{dD}$$

which means that, for a separating schedule

$$0 = \left(\frac{1}{2} - \frac{\alpha D^2}{2t^2} \right)$$

$$\frac{D}{t} = \sqrt{\frac{1}{\alpha}} \tag{4}$$

For a (necessary) interior (optimum) signalling equilibrium one must have $\alpha > 1$—$\alpha > 2$ in the two-period case.[2/] Feeding (4) back into the consistency condition (3) we get

$$W(D) = \frac{1}{2} (\sqrt{\alpha} + \sqrt{\alpha})D$$

$$= \sqrt{\alpha} D$$

Given this $W(D)$ the first order condition for maxima is satisfied when (using (2))

$$\sqrt{\alpha} - \alpha \frac{D}{t} = 0$$

or $\frac{D}{t} = \sqrt{\frac{1}{\alpha}}$ which checks with (4).

The second order condition is also satisfied since

$$E_{DD} = -\frac{\alpha}{t} < 0$$

The only "inconsistency" in this example is the violation of the (reasonable) requirement that {W(D) − Maximum Penalty} be non-negative in equilibrium. Such problems are eliminated if, for example, we consider the slightly more messy example with X distributed uniformly between t/4

and 3t/4, with mean t/2, i.e. when the range is narrowed. This is demonstrated in the Appendix.

Let us now formally describe the general structure of and the conditions for the existence of (this type of) non-dissipative signalling. In the notation of Spence's original signalling paper [5] let (with n = ability, y = level of signal)

productivity = $S(n,y)$, $S_y = 0$ here, $S_n > 0$, $\frac{t}{2}$ in our example

signalling cost = $C(n,y)$, $C_y > 0$ and having other characteristics we later discuss, $\frac{\alpha D^2}{2t}$ in our example

wage schedule - $W(y)$, $W(D)$ in our example

The worker maximizes

$$W(y) - C(n,y)$$

giving us the first-order condition

$$W_y - C_y = 0 \qquad (5)$$

and the second order condition

$$W_{yy} - C_{yy} < 0 \qquad (6)$$

The consistency condition (unlike in Spence (5) and (6)) is

$$W(y) = S(n) + C(n,y) \qquad (7)$$

which, on differentiating totally, implies,

$$W_y = C_y + (S_n + C_n) \frac{dn}{dy} \qquad (7a)$$

Hence, using (5) for a separating equilibrium one must have

$$(S_n + C_n) = 0 \qquad\qquad (8)$$

Since $S_n > 0$, we need $C_n < 0$. Let us further assume that $C_{ny} < 0$ (i.e. both total and marginal signalling costs are negatively correlated with productivity as was true for our example. In Spence's (5) dissipative signalling structure, $C_{ny} < 0$ and $C_y > S_y$ for $y >$ some $y°(n)$ (with $S_n > 0$) is sufficient to satisfy the second order conditions. We shall see that such is not the case here.

Totally differentiating (8) we get

$$(S_{nn} + C_{nn}) \frac{dn}{dy} + C_{ny} = 0 \qquad\qquad (9a)$$

From equations (5) and (9a) we have

$$W_{yy} - C_{yy} = C_{ny} \cdot \frac{dn}{dy} \qquad\qquad (9b)$$

$$= \frac{-(C_{ny})^2}{(S_{nn} + C_{nn})}$$

Since $W_{yy} - C_{yy} < 0$ from (6), we must have $S_{nn} + C_{nn} > 0$ as a necessary condition. Thus, for a separating equilibrium to exist, $S(n) + C(n,y)$ must, for each value of y (in the relevant domain), attain a unique global minima with respect to n.[3/]

We should point out that for an example in which the cost-structure is just the 'reverse' of our example -- where wages are increased by $\alpha(x-D)$ for $x > D$ but not decreased for $x < D$, the condition $(S_{nn} + C_{nn}) > 0$

(at $Sn + Cn = 0$) is not satisfied (with n standing for t and y for D) for the equivalent signalling cost function $\left[- \frac{\alpha}{2t} (t-D)^2 \right]$ and mean productivity $\frac{t}{2}$. Neither is $C_{ny} < 0$ satisfied (i.e. $C_{tD} < 0$).

But, $(C_t)C_n < 0$ is satisfied (for $D < t$) and we have to remember that $C_{ny} < 0$ is strictly not necessary for our model. Thus the $(S_{nn} + C_{nn}) > 0$ condition is an 'important' one. As straightforward calculations would verify, no (non-dissipative) signalling equilibrium satisfying the second-order condition exists in this case.

It is also useful to examine if the type of structure discussed in our example can work for more general distribution functions. Let productivity x be distributed with distribution function $F(x)$, density $f(x)$ in $(0,1)$, without loss of generality for bounded random variables. The employees would then maximize the objective function

$$E(D,t) = W(D) - \alpha \int_0^D (D-x)f(x)dx = W(D) - \alpha \int_0^D F(x)dx \qquad (10)$$

The signalling cost is

$$C(D,t) = \alpha \int_0^D F(x)dx$$

$\left(\frac{t}{2} \right)$ now refers to the <u>mean</u> of the random variable and, as before, true mean productivity $S(t) = \frac{t}{2}$. $C_t < 0$ if the distribution functions satisfy the Rothschild-Stiglitz [3] increasing risk ordering (without preserving means). $C_{Dt} < 0$ if the distribution functions $(F(z)s)$ are strictly ordered in the same ordering for all z (possibly weak only if $F(z) = 1$). But whether $[S(t) + C(t,D)]$ attains a unique minima in t for each D (in the relevent domain) is something that has to be verified in each case.

[In our discussion and derivation of the equilibrium schedule we have
not (unlike Spence (6)) dealt with incentives for workers to leave at the
beginning of the second-period in a two-period model. In Spence's model
(6) true productivity is known in the second period (and will be rewarded
in the outside market) and thus it is very important to deal with this is-
sue -- and the analytics are also well defined. In our two-stage model
this issue may be avoided by assuming that the two stages (time-periods)
correspond to life-cycle states ('young' vs 'old' workers). If a (then
'old') worker leaves the firm at the beginning of the second period to join
another firm he will still get $[W(D) - \alpha(D-x)]$ i.e. the wage he would
have gotten in his old job. His average productivity is not known, and nor
can he, unlike a 'young' worker get a 'fresh trial' starting at $W(D)$. We
have not verified if such a custom can survive competitive entry with one-
period contracts. In any case, our model can always be thought of as a
one-period model with 'quotas' -- with the penalty for not meeting quotas
being deducted from wages that are paid at the end of the period of working.]

The generalization minded reader might notice that a structure of the
same type as our example has some potential for application to a vastly
different phenomenon -- that of dividend policy in a world without differ-
ential taxation of dividends and capital gains.[4/] Think of x as i.i.d.
earnings in the two periods of a two-period model, distributed uniformly
between zero and t in each period. The current shareholders wish to
liquidate after the first period. $W(D)$ is the 'tentative' liquidation
value conditional on promised dividends D. The structure specifies a "mar-
ket accepted" revision of expectations about earnings and thus a reduction

in liquidation value of second period earnings by $\alpha(D-x)$ if earnings x

fall short of promised dividends D. Then, on replacing W(D) by $(W(D)+\frac{t}{2})$

and expressing true "productivity" as $t(2t/2)$ everywhere, our example

adapts to this new problem. But there is one major difficulty. We know

that (since the penalty structure is truncated) the 'equilibrium' (calcu-

lated) $W(D^*(t))$ will be greater than $\frac{t}{2}$. What is the current market val-

ue of the firm then -- 2 W(D)? But if that is so, then there are incen-

tives to sell out now. In general, the valuation results produced by the

non-dissipative signalling structure are not consistent with rational expec-

tation either ex-ante or, in general, ex-post (i.e. after the first period)

-- they are only consistent in an expected value sense at the end of the

first period. This type of inconsistency with ex-ante rational expectations

is likely to be an important difficulty in all non-dissipative financial

signalling models in which the signaller has a choice with respect to li-

quidation time. This is discussed further in the Appendix. Development

along these lines might, however, prove beneficial to our understanding of

such capital market phenomena.

In conclusion, our purpose has been to present a small extension to

the signalling structure of Spence (4) and illustrate it with a labor mar-

ket example. Other potential but more speculative possible applications of

such a non-dissipative signalling structure have also been noted.

APPENDIX

Proof that the Non-Negativity of Total Wage is Satisfied

When X is Distributed Uniformly Between t/4 and 3t/4 With Mean t/2

In the context of a one-period model, where $\alpha(D-X)$ is deducted from promised wages $W(D)$ if $X < D$, we have that

$$C(D,t) = \text{Signalling Cost} = \int_{\underline{X}}^{D} \alpha(D-X)f(x)dx$$

$$= \alpha\int_{\underline{X}}^{D} F(x)dx$$

$$= \frac{2\alpha}{t} \int_{\frac{t}{4}}^{D} \left(x - \frac{t}{4}\right) dx$$

$$= \frac{\alpha}{t} \left[\left(x - \frac{t}{4}\right)^2\right]_{\frac{t}{4}}^{D}$$

$$= \frac{\alpha}{t} \left[D^2 + \frac{t^2}{16} - \frac{Dt}{2}\right]$$

Hence, Equation (8) ($S_n + C_n = 0$) implies that, along the equilibrium schedule $D^*(t)$

$$\frac{1}{2} - \frac{\alpha D^2}{t^2} + \frac{\alpha}{16} = 0 \quad \text{Let } \alpha = 2 \quad \text{Then } \frac{D^*(t)}{t} = \sqrt{\frac{5}{16}} \simeq \frac{2 \cdot 25}{4}$$

Given that, equation (7) the consistency condition implies that, in equilibrium

$$W(D^*(t)) = \frac{t}{2} + 2t\left[\frac{5}{16} + \frac{1}{16} - \frac{1}{2} \cdot \sqrt{\frac{5}{16}}\right]$$

$$\sim \frac{t}{2} + 2t\left[\frac{6}{16} - \frac{4 \cdot 5}{16}\right] = \frac{t}{2} + \frac{3}{16}\,t \quad = \frac{11}{16}\,t$$

$$\text{Maximum Penalty} = \alpha(D^{*}(t) - \frac{t}{4}) = 2t\left(\sqrt{\frac{5}{16}} - \frac{1}{4}\right) = t\left(\frac{2 \cdot 5}{4}\right) = \frac{10}{16} \cdot t$$

$$\therefore \text{Minimum after-penalty wage} = \frac{t}{16} > 0$$

Experiments with the uniform distribution lead us to believe that, as the range of the productivity distribution(s) narrows - and $\frac{t}{4}$ to $\frac{3t}{4}$ is still a very large range by real world standards - then, by choosing larger α, the equilibrium $W(D^{*}(t))$ can be brought closer to $\frac{t}{2}$ and thus the problem with ex-ante rational expectations discussed in connection with the dividend policy example is 'reduced'. Of course, this problem can not be eliminated as long as the signalling cost structure is 'truncated' below D. Numerical experiments with non-truncated cost-structures, in which there is a penalty for performance below the quota and a reward for performance above it, have not been very successful in producing signalling 'equilibria' that provide interior maxima (solutions) which are continuous in the parameters.

FOOTNOTES

* Needless to say, this discussion owes much to the pioneering work of Spence (5). His recent work on Contingent Contracts (6), which is related to the discussion here, was discovered by me soon after I obtained the results of this paper. Salop and Salop (4) were one of the first to consider contingent contracts in a labour market context.

1. We restrict our attention to signalling in a (potential) adverse selection situation. Problems of moral hazard are assumed away - the uncertainty in productivity is assumed to be purely "technological."

2. A separating equilibrium - as opposed to a non-separating equilibrium in which, for example, workers are paid their ex-post productivity - is helpful for job assignment. A worker's true productivity (distribution) is implicitly conditional on optimal job assignment.

3. An intuitive explanation of this requirement goes as follows. Suppose this requirement is not met. Let $n_1(y_1)$ and $n_2(y_2)$ be two different points on the (hypothetical) 'equilibrium' schedule (derived from equation (8)) and suppose $(Snn + Cnn) \lessgtr 0$ at $n_1(y_1)$. Then it follows that

$$S(n_2) + C(n_2,y_1) < S(n_1) + C(n_1,y_1) = W(y_1)$$

Hence

$$W(y_1) - C(n_2,y_1) > S(n_2)$$

But, by the definition of the 'equilibrium' schedule

$$W(y_2) - C(n_2,y_2) = S(n_2)$$

Hence the agent with "ability" n_2 is better off by signalling y_1 rather than y_2. Thus the hypothetical 'equilibrium' schedule is not a valid one.

4. A (dissipative) signalling structure in which dividends function as a signal because of their tax costs - dividends are taxed at a higher rate than capital gains - is discussed in Bhattacharya [1].

REFERENCES

1. Bhattacharya, S., "Imperfect Information, Dividend Policy, and the 'Bird in the Hand Fallacy'," in "Essays on Financial Theory," Ph.D. Thesis, Sloan School, M.I.T., (1977).

2. Riley, J., "Competitive Signalling," Journal of Economic Theory, (April 1975).

3. Rothschild, M. and J. Stiglitz, "Increasing Risk: A Definition," Journal of Economic Theory, (1970).

4. Salop, S. and J. Salop, "Self-Selection and Turnover in the Labor Market," Quarterly Journal of Economics, (November 1976).

5. Spence, A.M., "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution," Journal of Economic Theory, (March 1974).

6. _____, "Signals, Contingent Contracts and the Divergence of Earnings from Productivity," Harvard Institute of Economic Research, Discussion Paper No. 540, (March 1977).

Diffusion of Technological Innovations: A Sequential

Experimentation Approach

Sudipto Bhattacharya

I

Introduction:

There exists a large literature on the problem of the rate

of diffusion of new technical innovations in industry. As Rosenberg

(10) points out, the regularities that are observed and sought to

be explained are those having to do with (A) the general slowness

of the diffusion process, and (B) wide variations in rates of

acceptance for different innovations. Rosenberg himself discusses

some aspects of the process, like secondary innovations that increase

the applicability of the original one, complementarity with the

firms' existing technologies, improvements in the old technologies,

etc., that relate to category (A) above. His mode of analysis is

largely historical, but it makes one aware of the following fundamental,

if obvious, feature of the problem. Any theory or explanation of

the diffusion process has to start with some premises about the

heterogeneity of adopting units, or firms, along dimensions that

are relevant to the innovation. One very important dimension is the

range of related but not identical technological and industrial

activities, or any economic activity for that matter, to which

the innovation is potentially relevant. Other important dimensions

of possible heterogeneity are, for example, the economic agents'

degree of pessimism or optimism regarding the innovation, their

risk-return trade offs, their differences in access to the relevant

information about innovation, their access to capital markets for

financing investment in the new technology, and so on. An explana-

tion of the diffusion process has to come to grips with the issue

of the relative importance of difference dimensions of heterogeneity

in determining the characteristics of the diffusion process and the

relationship of these characteristics to the parameters of the

innovation itself. The exploration of some aspects of this issue

is the purpose of this paper.

Significant and methodical investigation of some of these

questions have been carried out by Mansfield ((5) and (6)), in his

path-breaking work. An understanding of his approach is critical

to an appreciation of the somewhat unusual exploration that we

carry out later. The essence of Mansfield's approach is to view

the diffusion process as one of "imitation." The likelihood of the

hold-out firms' adoption is fundamentally related to the number

for firms who have already adopted the innovation -- because "as

more information and experience accumulate, it becomes less of a

risk to begin using it ... competitive pressures mount and "bandwagon"

effects occur. Where the profitability of using the innovation is

very difficult to estimate, the mere fact that a large proportion

of its competitors have introduced it may prompt a firm to consider

it more favorably." ((5) , pp. 137-138). In more specific form,

this hypothesis was tested in the following form. Let N be the

the total number of firms in the industry to which the innovation is

relevant, and let M(t) be the total number to have adopted by time

t. Then it was hypothesized that

$$\frac{dM(t)}{dt} = [N - M(t)][Q + \phi\frac{M(t)}{N}] \qquad (1)$$

where ϕ is the "adoption parameter." It was further hypothesized

that ϕ was related to the parameters of the innovation itself by

the relationship

$$\phi = a_0 + a_{1*}\Pi + a_{2*}I$$

$$a_1 > 0, \quad a_2 < 0 \qquad (2)$$

Where Π is a measure of profitability, and I of investment size [1]

required for any significant commercial use. To test the relationship,

logistic "S-shaped" diffusion curves

$$\text{Log } [\frac{M}{N-M}] = \ell + \phi \cdot t \qquad (3)$$

implied by equation (1) were fitted to the empirical evidence of

diffusion of twelve innovations in four industries, and the diffusion

parameters obtained were regressed cross-sectionally using the speci-

fication of equation (2). [2] The hypotheses were well supported by

the data. ((5), chapter 7) Mansfield was careful to apply the

framework to technological, cost-reducing innovations, and not to
new products where other phenomena, like a rush to introduce to
exploit initially high profits in a limited market may be involved,
thus making profitability of introduction strongly related to
timing. He also pointed out that this simplified framework could not
accommodate other important parameters of the innovation, like its
initial degree of uncertainty, and its resolution over time. The
innovations chosen were also those in which the importance of
secondary innovations, in the sense discussed by Rosenberg (10),
was not overwhelming. Regarding the effects of Π and I on the
diffusion parameter ϕ, the reasoning advanced was that (a) firms
note and respond more quickly when the profitability difference is
higher, and that (b) firms tend to be more cautious with large
investments, and often have difficulty raising the money. Other
variables explored by Mansfield are discussed later.

The effect of firm level parameters on speed of adoption
was also explored by Mansfield, and it was found that except
for firm size, there were no significant effects due to the firms'
profitability, growth rate or liquidity. Larger firms tended to
be somewhat quicker, but not out of proportion to their dominance
of the industry's size. There was no significant relationship between
a firm's quickness of response to one innovation to its response to
a succeeding one.

Despite its good empirical performance, one of the shortcomings

of the "bandwagon" model has been its fuzziness on precisely what

aspects of inter-firm heterogeneity contribute to the bandwagon

effect, and how. Mansfield himself concludes with a somwhat cryptic

remark about the dependence of ϕ on Π -- "...it would suggest that

there exists an important economic analogue to the psychological

laws relating reaction time to the intensity of the stimulus."

((5), pp. 190). The lack of discriminatory power with respect to

different possible dimensions of inter-firm heterogeneity has given

rise to explorations that are somewhat ad-hoc in their choice of

the underlying mechanism. For example, in a paper that is part

of a larger study on the diffusion of technology, Lacci, Davis and

Smith investigate a model in which firms have priors about the profi-

tability of the innovation that are normally distributed (among firms),

and the autonomous, constant-rate rightward movement of this

prior distribution with time results in a cumulative normal

diffusion curve! ((7), chapter 6). The model performs extremely

well, somewhat better than Mansfield's logictic curve formulation.

Besides just the element of ad-hoc-ness, it is hard to reconcile

Mansfield's empirical evidence on the lack of correlation between

a given firm's "diffusion leadership" vis-a-vis different innovations

with the attribution of a predominant role for initial heterogeneity

in expectations in the diffusion process -- presumably optimists

should be optimists at least somewhat consistently. Other plausible

dimensions of heterogeneity, including primarily differences in
access to initial information, and differential financial "health"
and access to financial markets, have also been explored in the
literature. While these reasons do, no doubt, play some role,
Mansfield's data casts doubts on the latter, and studies published
in Nabseth and Ray ((6), pp. 69-84) seem to point against a signi-
ficant role for the former, in that average year of first information
seems to be more or less the same for all categories (by time of
adoption) of adopters, i.e., firms which ultimately end up using
the innovation. The basic premise of the "bandwagon" hypothesis
viz. that imitative behavior plays a predominant role, has also been
questioned. Given the diversity of technologies and product mixes
that one finds in any industry to which the innovation is potentially
applicable -- a diversity that is important enough to cause signi-
cant differences in profitability of adoption for all but the landmark
innovations -- there is some reason to question the degree to which
imitative behavior is responsible for the diffusion curve. (On
this point see Gebhardt and Hatzold, "Numerically Controlled Machine
Tools," in (7), chapter 3). Prominent researchers in the field
have raised doubts concerning the usefulness of the analogy
that the "bandwagon" model makes with, for example, models of
epidemic spreading. For instance, Ray (7) expresses the opinion
that, given the diversity among firms, the role that may be played
by imitative behavior, if any, is that of drawing some attention

to the innovation (if first information had not already reached
the firm) and sums up by stating that the "equations probably
demonstrate an effect rather than a cause, and have more descriptive
than analytical value" ((7), page 226). It is probably a fair
assessment that considerable doubts remain about the mechanism
behind an interesting observed empirical regularity.

In what follows, we try to approach the problem from a
different point of view -- one that may be classified as that
of rational sequential capital budgeting decisions. The essence
of our approach is the following. At the risk of taking an
extreme position, we choose to emphasize one dimension of firm
heterogeneity as the primary determinant of the diffusion
phenomenon. Firms within any relevant industry differ in the precise
nature of their activities and technologies. A fundamental charac-
teristic of a very significant innovation, as opposed to more routine
technical improvements, is that its applicability (in the sense
of the profitability of its application), though it is not appreciated
ex ante, turns out ex-post to be more or less the same for a broad
range of technologies and product mixes. (For example, Hakansson
((7), chapter 4) provides evidence that, at least for one major
innovation, the ex-post assessed profitability for various categories
of adopters was quite strikingly homogeneous relative to the
differences between adopters and non-adopters.) Ex-ante, however,
firms of one class find very little in the experience for firms of

another class that convinces them to adopt the innovation, at least

beyond a certain stage of shared learning of information about the

innovation. Instead each firm starts out, quite rationally, on

some optimal sequential procedure of commercial trials to estimate

the unknown mean profitability of the innovation, and the estimation

takes time because of the irreducible technological uncertainty

associated with any technology. In any such sequential procedure,

if estimation and especially waiting are costly, there will be an

optimal stopping set -- a high or low enough realization   history

from trials such that further experimentation would be abandoned,

and a decision made to adopt or reject the innovation. The

phenomenon of diffusion is hypothesized to arrive essentially from

the distribution of first passage times to the optimal stopping

set of this sequential procedure, given the true parameters of the

innovation. It is conjectured that the distribution of first

passage times, which arises without any strong inter-firm interaction

or bandwagon effects, would display some of the major characteristics

observed by Mansfield viz. the logistic form, and the strong depen-

dence on the profitability parameter.[3]   We also explore likely

reasons for the relationship with other parameters that Mansfield

observes, and discuss the effect of additional variables like

"resolution of uncertainty" and their likely effect on the diffusion

curve. These are the issues that we explore in the sections below.

Clearly, the above framework is a vast simplification. For one thing, even if perceived profitability of introduction (relative to installing the newest pre-innovation equipment) was the same for all firms, there would be some differences in adoption behavior because of the differing age structure of equipment; since investments in old equipment are sunk costs, it is only older vintage equipment, with their greater deterioration and lower productivity, that are likely to be replaced. We shall have more to say on this point later. Secondly, it is unreasonable to assert that imitative behavior plays no role at all. Indeed, if there exists an important degree of similarity among firms, a type of "bandwagon" effect can arise that would have nothing to do with initial heterogeneity among firms along any dimension. It would arise solely from the fact that, in the course of its sequential estimation, a given firm rationally takes adoption decisions by other firms as additional observations. For analytical richness, it would have been nice to have been able to take this factor into account -- but there are great difficulties in doing so, and we discuss these later. We believe, however, that the degree of doubts expressed about the relevance of the imitative behavior framework in the literature does provide some justification for pursuing our approach, at least as an exploratory analysis. In section II we discuss the model of sequential experimentation that we employ, and analyze detailed results from this model in section III.

II

We are concerned with the first three of the four fundamental

phases of an innovation -- commercial trials, rapid adoption, slower

gains, and decline -- that have been noted in the pioneering study

of Jerome (4). In particular, we model the commercial trials

process as one of sampling to estimate the unknown mean return of

the new technology, and we assume that the observations are continuous

enough, and the sample path of the new technology's return is a

constant variance diffusion process, in the technical sense, so

that the variance of its profitability is perfectly known. We make

use of results derived by Chernoff (3) on the optimal sequential

procedure. The specific assumptions made by Chernoff (3) about the

relevant cost structure are that (a) there is a cost C per sampling,

and (b) if the sampling stops and the wrong decision is made then

the cost is $K*|\mu|$, where μ is the true mean of the unknown Guassian

process, which we may interpret as a measure of the expected profi-

tability of the innovation, since Chernoff's procedure applies to

testing the mutually exclusive and exhaustive hypotheses that

HO: $\mu > 0$ and Hl: $\mu \leq 0$. The criterion function is expected loss --

but that is not a major headache since we know the variance of the

new technology, and thus a Capital Asset Pricing Model type rela-

tionship may be used to extract a suitable hurdle rate, and μ may

be considered as the mean return from the new technology less this

hurdle rate -- with K being a measure of, say, Net Present Value
per unit of expected excess return.

The part of the cost-structure that has to be treated with
great care in our problem is the "cost of sampling," or the cost of
waiting one more period, while continuing to experiment with the new
technology on a small scale. In our situation, there are likely
to be three types of waiting costs. One is direct -- that of
actual commercial trials and associated minor development work.
The second would be that associated with the postponement of expansion
and replacement plans using pre-innovation equipment of new vintage.
The third element, in a discounted framework, is that associated
with postponing the relative advantage of the new technology over
the old one into the future. The critical feature that distinguishes
the third from the first two is that it is dependent on the estimate
of the relative advantage of the new technology. To accommodate
such a feature would require a basic extension of the Chernoff
framework, which we shy away from for two reasons. First, this
part of the waiting cost would almost certainly give rise to an
asymmetric stopping set -- since when the estimate of the
new technology's profitability is low, it also costs less to
experiment some more, for example. This would make the evaluation
of first passage times extremely diffisult. (The work of Petkav
(9) lends support to this conjecture in a somewhat different model.)
Secondly, though the principles of deriving optimal stopping sets

in Gaussian estimation problems are simple and well understood, the mechanics are sheer tedium -- involving essentially the manipulation of series solutions to a free-boundary problem. Our efforts show that it is almost impossible to (consistently) justify a constant waiting cost of this (third) type, when the profitability of the project as a whole is imperfectly known, and its estimate is changing. We therefore assume that, as regards waiting costs, the first two elements dominate, which is not necessarily very unrealistic.

(A question that naturally arises is that why should firms not go ahead with their plans and switch later to post-innovation equipment if they turn out to be profitable enough? The answer is that for a given level of positive relative profitability of post-innovation equipment, the net present value loss of doing so -- either from opportunity costs arising from the lack of a secondary market, or because of the value adjustment on it once the relative advantage of post-innovation equipment is known -- may easily outweigh the costs of waiting, if there is a significant assessed probability of accepting the new technology. At the same time the same waiting costs (of postponing these plans) may be large compared to the costs of postponing the relative advantage of the new technology. In the real world, there would be significant deadweight costs of adjustment to further strengthen this argument. However, the strict separation of the two types of decisions assumed here is heuristic, and is not likely to be rigorously true.)

We should note that when we talk about postponing the advantages of replacement plans, we are implicitly assuming a world in which either or both of deterioration and embodied technical progress take place, which is natural enough. An element that may serve to reduce the costs of waiting is anticipated decreases in the costs of the new equipment. Further implications of our assumed model for waiting costs are discussed in section III.

In the Chernoff procedure, the experimenter starts out with a normal prior with parameters (μ_0, σ_0^2) about the mean μ of a normally distributed variable X with variance σ^2. After n realizations $\{X_i\}$, the posterior estimate of μ is distributed normally with mean Y_n and variance S_n given by

$$Y_n = \frac{\mu_0 \sigma_0^{-2} + \sigma^{-2} \sum_{i=1}^{n} X_i}{\sigma_0^{-2} + n\sigma^{-2}} \tag{4a}$$

$$S_n = (\sigma_0^{-2} + n\sigma^{-2})^{-1} \tag{4b}$$

More surprisingly, $(Y_n - Y_m)$ for $m > n$, is distributed normally with mean zero and variance $(S_n - S_m)$. An approximation to the discrete sampling procedure is therefore obtained by considering $X = \sum_i X_i$ and Y as continuous-time diffusion processes with respect to the true probability distribution for $\{X_i\}$ and the Bayesian posterior distribution of Y respectively -- and the stopping set for the sequential sampling procedure is obtained by solving a free-boundary

problem with the heat equation, which arises from the underlying
Guassian diffusion process. From the obvious symmetry of the
problem, the optimal stopping sets for accepting H0 and H1 are mirror
images of each other, about the zero-axis of Y, the posterior mean.
As the number of trials tends to infinity, or the posterior variance
of the estimate of μ viz. S_n tends to zero, the optimal stopping sets
are given by the following equations, which are simple enough to at
least numerically solve for the distribution of first passage times
to them. Let

$$X' = \mu_0 \sigma_0^{-2} + X\sigma^{-2} \qquad \text{and} \qquad T = \sigma_0^{-2} + t\sigma^{-2} \qquad (5)$$

Then the optimal stopping set is given by

$$X' \simeq a/4T \qquad \text{where} \qquad a = (\frac{K}{C})\sigma^{-2} \qquad (6)$$

To illustrate one reasonable and intuitively expected feature of
the stopping set, we note that, ceteris paribus, the effect of an
increase in (K/C) is to increase the "performance required" for the
acceptance of the new innovation, which is sensible.

Another sensible feature of the procedure may be seen
transparently by considering the case when the precision of the prior
σ_0^{-2} goes to zero. The stopping set may then be expressed in terms
of X as $X = \pm(K/4C) * \sigma^2/t$ and hence it is more conservative as

the uncertainty connected with the innovation σ^2, which impedes

estimation of μ, goes up.

Since X is a Wiener process with

$$E[dx] = \mu dt \qquad \text{Variance } [dx] = \sigma^2 dt \tag{7a}$$

the first passage time density function $g(t|x_0,t_0)$ to the barrier

B(t), given starting X_0,t_0 satisfies the Kolmogorov backward equation

$$1/2 \ \sigma^2 \ \frac{\partial^2 g}{\partial x^2} + \frac{\mu \partial g}{\partial x} = \frac{\partial g}{\partial t} \tag{7b}$$

Consider the more convenient distribution function $G(t|x_0,t_0) = \int_{t_0}^{t} g(\tau|X_0,t_0)d\tau$. Since equation (7b) has constant coefficients, G

also satisfies it. Let us transform to the variables $X' = \mu_0\sigma_0^{-2} + \sigma^{-2}X$

and $T = \sigma_0^{-2} + \sigma^{-2}t$, to get

$$1/2 \ \frac{\partial^2 G}{\partial x'^2} + \mu \ \frac{\partial G}{\partial X'} = \partial G/\partial T$$

The boundary, from equation (6), is given by $\pm D/T$ where $D = (\frac{K}{4C})\sigma^{-2}$.

For convenience, let us transform from the variable X' to Y, where

$$Y = X'*T + D \quad \text{and} \quad \text{let } E = 2D \tag{7c}$$

Then

$$\left(\frac{\partial G}{\partial X'}\right)_T = T\left(\frac{\partial G}{\partial Y}\right)_T \qquad \text{and} \qquad \left(\frac{\partial^2 G}{\partial X'^2}\right)_T = T^2\left(\frac{\partial^2 G}{\partial Y^2}\right)_T$$

and,

$$\left(\frac{\partial G}{\partial T}\right)_{X'} = \left(\frac{\partial G}{\partial T}\right)_Y + \left(\frac{\partial G}{\partial Y}\right)_T \left(\frac{\partial Y}{\partial T}\right)_{X'}$$

Thus we get to the equation, $G(T|Y_0,T_0)$

$$1/2 \; T^2 \frac{\partial^2 G}{\partial Y^2} + \left(\mu T - \frac{(Y-D)}{T}\right) \frac{\partial G}{\partial Y} = \frac{\partial G}{\partial T} \tag{8a}$$

with the boundary conditions

$$G(0,T) = 0 \qquad\qquad G(E,T) = 1 \tag{8b}$$

and the initial condition

$$G(Y,\sigma_0^{-2}) = 0 \quad, \qquad 0 \leq Y < E \tag{8c}$$
$$= 1 \quad, \qquad Y = E$$

since we are only interested in the probability of absorption at the upper (acceptance) barrier. Technically, $\underset{T\to\infty}{\text{Lim}} \; G(T|Y_0,T_0)$ gives the equilibrium proportion of firms which adopt the new technology, but in reality truly imitative behavior may set in after the profitability of the innovation has been convincingly demonstrated.

In purely mechanical terms, the effects of various parameters of the innovation on the diffusion process may be investigated as follows. The effect of μ, the true profitability per unit of investment, or more precisely the excess of the true mean profitability over the appropriate hurdle rate, is only on the differential equation, and not at all on the boundaries. In contrast, the effect of μ_0 and σ_0^2, which reflect the initial optimism/pessimism and the initial uncertainty about the (mean) profitability of the innovation, is only on the initial conditions. The effect of investment requirement, via (K/C), is solely on the height of the barrier. Lastly, the rate of resolution of uncertainty about μ, which is measured by something like $\dfrac{\partial S_n}{\partial n}^{-1}$, is clearly a decreasing function of σ. The effect is both on the height of the barrier D, and on the relationship between "real time" t and T, and through the latter effect, it is dependent on the degree of uncertainty σ_0^2. That sums up the essential characteristics.

It is an unfortunate fact of life that an analytical solution to this problem has not been found! Therefore, we resort to numerical methods to obtain the first passage time distributions. The details and the results are reported in the next section.

III

In our numerical simulations of the first passage time behavior, we have assumed the following values of the variance, the relative cost ratio (K/C) and the time range. These are

$$\sigma_0 = 1 \cdot 0 \quad \text{or} \quad \sigma_0^{-4} = 1 \qquad \sigma_0^{-2} = 1$$

$$\sigma = 1 \cdot 0 \qquad \sigma^{-2} = 1 \qquad 0 \leq t \leq 1, \quad 1 \leq T \leq 2$$

$$7 \leq \frac{K}{C} \leq 14 \qquad \text{or} \qquad 3 \cdot 5 \leq E \leq 7 \cdot 0$$

μ, the true mean profitability of the innovation, is varied between 1.15 and 3.25 over a range that depends on the value of E. That is because, at least at the initial stage, we only deal with first passage time distributions for unbiased priors i.e.,

$$\text{(starting)} \quad Y = \mu_0 \sigma_0^{-4} + \frac{E}{2} , \quad \mu_0 = \mu,$$

and therefore have to make sure that the initial value of Y, as defined in equation (7c) i.e., $\mu + E/2$, does not exceed E, the boundary of the stopping set for Y . We note that the standard deviation of the initial prior σ_0, is "small" compared to μ (RMU in the tables) for values of μ in the upper part of the range.

These obtained first passage time distributions are then fitted to a logistic curve of the form

$$\text{Log } (G/(1-G)) = a + L*t \tag{9}$$

Table 1 presents our estimates of the important slope parameter L. (The logistic formulation is fitted only over the range $.01 < G < .99$, and over a sub-range of this if the G(t) function does not reach the value 0.99 by t = 1.) The fit obtained is extremely good, the adjusted $\bar{R}^2$ being always in excess of 0.98 and the student T-statistics generally above 30 for both the intercept and the slope coefficients.

Table 1

Slope Parameter of Logistic Curve

$$\text{Log} \frac{G}{1 - G} \quad = \quad a + L\tau$$

Fitted to the First Passage Time Distribution Function $G(\tau)$

(Also, "first and last time points" at which $G(\tau)$ takes
on the values 0.01 and 0.99 respectively -- then are given in
terms of the variable

$$\tau = 100(T - 1)$$

(Regression Estimates for L are per unit of τ)

RMU

E	3.25	3.15	3.05	2.95	2.85	2.75	2.65	2.55	2.45
7.000	(28,49) .43887	(37,58) .4411	(44,65) .4402	(52,74) .4508	(61,82) .4262	(69,91) .4287			
6.500			.4174 (24,46)	.4331 (34,56)	.4286 (43,65)	.4260 (51,74)	.4207 (60,83)	.4102 (68,92)	.3942 (76,101)

E	2.85	2.65	2.55	2.45	2.35	2.25	2.15	2.05	1.95
6.500					(84,110) .3751				
6.250	(25,44) .38794	(40,59) .3714		(54,74) .3576					
6.000	(30,53) .4000	(39,63) .4050	(50,74) .3998	(59,83) .3863	(67,93) .3823	(77,104) .3679	(86,110) .3519		
5.750			(30,50) .3584		(46,66) .3499		(60,83) .3432		
5.500		(9,30) .3425	(20,41) .3611	(30,50) .3531	(36,58) .3653	(45,67) .3563	(53,76) .3446	(62,85) .3282	(68,93) .3132
5.250					(27,48) .3439		(44,67) .3397		(61,86) .3178

RMU	2.35	2.25	2.15	2.05	1.95	1.85	1.75	1.65	1.55
E									*
5.000	(16,38) .3383	(26,48) .3375	(35,58) .3387	(44,67) .3239	(52,77) .3174	(61,87) .3039	(70,97) .2857	(79,100) .2577	
4.750				(34,58) .3260		(52,78) .3053		(70,99) .2726	
4.500			(17,35) .3179	(22,46) .3174	(34,58) .3116	(42,68) .3019	(51,79) .2921	(62,91) .2752	(71,100)** .2521

* Final value of G(t)   0.789

** Final Value of G(t)   0.969

RMU	1.85	1.75	1.65	1.55	1.45	1.35	1.25	1.15	1.05
E									
4.250		(42,69) .2895		(61,92) .2591		(81,100)* .2395			
4.000	(16,44) .2759	(29,57) .2682	(39,68) .2621	(51,81) .24997	(60,93) .2385	(72,100)** .2127	(82,100)*** .2155		
3.750			(29,58) .2753		(51,84) .2457		(73,100)+ .2102		
3.500				(23,55)	(37,69)	(49,84)	(61,99)	(73,100)++	

* Final G(t) =.487   ** Final G(t)=..861   *** Final G(t)=.378

+ Final G(t)=.813   ++ Final G(t)=.700

(Footnote to Table 1) :

The time units for the first three rows were obtained by dividing the same time interval into 125 units rather than 100 units. The exhibited L-s are adjusted for this, but the first and last time points are not.

There is some evidence of the presence of "auto-correlation"
but, given the exceedingly good fits obtained, no correction has
been made. It is useful to point out that, in our framework, there
is a very good reason to expect the first passage time distribution
to have the "general characteristics" of a logistic curve -- though,
in the absence of analytical solutions, exact comparisons (e.g.,
the linearity in the logarithmic form of equation (3)), are not
possible. The reason we expect the "logistic-type" behavior is
that given a stationary or narrowing (optimal) stopping barrier
(in X',T' space) and a positive drift towards that barrier (because
RMU is greater than zero) we expect that with increasing time (or T'),
a larger fraction of the remaining "probability mass" will hit the
barrier in the next marginal instant since, "on average", it is closer
to the barrier. Now, the probability mass that has hit the barrier is
itself an increasing function of time past, and thus the former
serves as a "proxy" for the latter. Thus the proportion of the
remaining mass that is expected to hit the barrier in the next
marginal instant is an increasing function of the proportion that
has already hit it -- which gives rise to the logistic nature of
the curve.

A glance at table 1 is sufficient to notice a significant
positive relationship between RMU and L. Heuristically, this is
also to be expected from a crude mental image of the (hypothesized)
diffusion process viz. that a measure of the "speed-up rate" of

arrivals, or a measure that is inversely related to the dispersion of the first passage time distribution, should go up with an increase of the strength of drift towards the barrier. Similar "intuitive" support can be provided for the apparent negative relationship between RMU and the time taken (from the beginning) for a given percentage (of firms) to hit the barrier. The effect of the barrier height E on this time taken for arrival of a given percentage also seems clear-cut -- the two are positively related, as would be intuitively expected. On the other hand, the effect of E on L is not easily discernible from a casual inspection of table 1, and neither is there a strong intuitive rationale to expect L, which is essentially a measure related to inversely to the dispersion of the first passage time distribution, to react one way versus another to changes in E. We therefore move to more formal investigations of the dependence of L on the parameters, through regression studies.

As part of an attempt to replicate a Mansfield-type test of the dependence of L on the parameters, we regressed L on the (true) mean profitability RMU and the height of the barrier E. The regressions were carried out in both linear and logarithmic form. We carried them out both for the whole sample of seventy-four parameter combinations as well as for a subset that excluded the highest three values of E (7.0, 6.5, 6.0) which, at a casual glance, seemed not to show as strong a dependence between L and RMU. The regression results are summarized below.

$$L = .0316 + .08812*RMU + .0212*E + \tilde{\varepsilon}_1 \qquad (10a)$$
$$\quad (.0114) \quad (.00684) \qquad (.00368)$$

$\overline{R}^2 = 0.9716 \quad$ D.W. $= 2.063 \quad$ F(2.70) $= 1197.5 \quad$ No. of observations $= 74 \quad$ Mean of L $= .3283$

$$L = .05188 + .09857*RMU + .01237*E + \tilde{\varepsilon}_2 \qquad (10b)$$
$$\quad (.01406) \quad (.0078) \qquad (.0045)$$

$\overline{R}^2 = 0.9515 \quad$ D.W. $= 2.078 \quad$ F(2.49) $= 480.6 \quad$ No. of observations $= 52 \quad$ Mean of L $= .2959$

$$\log L = C + 0.670*(\log RMU) + 0.1426(\log E) + \tilde{\varepsilon}_3 \qquad (10c)$$
$$\quad (.0459) \qquad (.0622)$$

$\overline{R}^2 = 0.9596 \quad$ D.W. $= 0.9122 \quad$ F(2.50) $= 593.24 \quad$ No. of observations $= 53$

One quite note-worthy conclusion emerges from the regressions. Whereas increasng RMU increases L and decreases the time taken for a given percentage to "arrive" at the barrier, increasing E has the opposite effect on the time taken for a given percentage to "arrive" but the same effect (in terms of sign) on L. Thus the connotation of increasing "speed of diffusion" that is often attached to L in

the literature may not be a valid one for parameter changes (across innovations) that affect barrier height E and not RMU. The two measures associated with the adoption rate -- one relating (inversely) to the dispersion of the first passage time distribution, and the other related to the time needed for diffusion measured from some beginning point, respectively -- can move in "opposite" directions.

As regards the effects of individual variables, the comparison with Mansfield's results goes as follows. The effect of (mean) profitability on L in our exercise is much the same as the effect of ex-post relative profitability in Mansfield's work -- strong and very significant. It is thus possible that an explanation different from "a psychological law relating intensity of stimulus and reaction" does exist. The other two important variables studied by Mansfield, the size of investment required and the durability of the equipment that the innovation replaces, are somewhat less straightforward to relate to our framework. We shall first take up the durability variable, then deal with the effects of σ, the "irreducible" standard deviation of the new technology's rate of return, and then proceed to a discussion of the likely effect of the size of investment variable.

Regarding durability, Mansfield hypothesized that higher durability would tend to lower the value of L since "... although rational economic calculations might indicate that replacement would be profitable, firms may be reluctant to scrap equipment that

is not fully written off and that will continue to serve for many

years." ((5), page 146) -- an argument that is based entirely

on "irrationality." Empirically, this was (weakly) confirmed.

There exists, however, a plausible alternative explanation based

on rational behavior. In the cross-section (of innovations)

higher durability is likely to be associated with a greater

preponderance of older vintage pre-innovation equipment. For

firms owning such equipment, the waiting costs of postponing

replacement with pre-innovation equipment may be higher, and hence

(K/C) would be lower, and thus the height of the stopping set E

would be lower. This has two effects, which are superficially

"contradictory." It should lower the time between innovation and

acceptance (by say 80%) but, at least according to the results

from our sample, it should lower L, i.e., lead to a flatter first

passage time distribution between the values .01 and .99. The first

point has been noted, though not in this framework, in the

existing literature.[4]

The effect of the σ variable is simple to analyze, and it

relates also to the effect of "resolution of uncertainty," since

the increase in the precision of the posterior estimate of   per

unit of trials/time is directly proportional to σ^{-2}. Ceteris paribus,

as σ^{-2} goes up, (i) barrier height E goes up, and (ii) the increment

of real time Δt associated with a given increment of ΔT goes down.

Both these effects serve to increase L measured per unit of real

time. However, the effect of increasing σ^{-2} on the earliness of
diffusion is ambiguous. The increase in E increases this time
measured in terms of T but the second effect acts to reduce this
time measured in terms of t, real time. Approximately speaking,
increased σ^{-2} would on balance increase/decrease a measure of
the time required for diffusion as E affects the increment in T
required for the first passage time distribution to reach an agreed-
upon criterion value more than/less than proportionately. Based
on Table 1, a more than proportionate effect on the increment in
terms of ΔT, and thus an increase in the increment in terms of Δt,
real time required, would be predicted, which again is somewhat
counterintuitive. However, given that our results are dependent on
the arbitrary magnitude of the prior precision assumed in the
numerical simulations, this last prediction is probably the least
reliable of all.

The third and last important variable, the size of investment
required for the innovation, is probably the most difficult to
relate to our framework. Mansfield hypothesized and empirically
confirmed a negative effect of higher investment on L, because
of the supposed tendency of firms to be more cautious with larger
investments, and possibly having difficulty in raising the money.
Neither factor should loom large for a firm in a competitive,
well-diversified capital market -- and nothing in Mansfield's other
evidence leads us to question the premise of such an environment.
The first passage time approach suggests one plausible alternative
explanation -- if higher investment technologies also tend to have

higher irreducible uncertainty σ^2, then the effect would be a fall

in L as investment goes up. In fact this would tend to be augmented

by a higher hurdle rate, and thus a higher waiting cost holding

net present value constant, which would lead to a further drop

in the barrier height E, and thus in L.[5]

(In passing, it may be pointed out that in the literature

some degree of confusion exists with respect to the likely reason

for the effect of the investment variable on L observed by Mansfield.

Writers who have discussed Mansfield's work (e.g., Norris & Vaizey

(8), and Nelson, Peck and Kalchek (7)) have "explained" the negative

effect of the investment requirement on L as arising from the fact

that, given a certain level of operating cost advantage for the

innovation, the overall profitability of investment in it varies

inversely with the size of the investment required. That reasoning

does not, of course, apply when the profitability variable that is

held constant is the overall rate of return on the investment.)

We believe, however, that the observed effects of both durability

and investment probably arise, in a framework in which no recourse is

taken to imperfect market or imperfect rationality assumptions, for

reasons that are different from, and complementary to, those

discussed above. So far in our discussion we have, aside from some

parenthetical mention in section I, left out any consideration of the

age distribution of equipment in the industry. Higher durability is,

of course, likely to be associated with a broader range of ages for

existing machines. In all likelihood, it will only be profitable

to replace the older machines. Thus the age distribution should, of

course affect the rate of intra-firm diffusion -- as measured, for

example, by the fraction of a "representative" firm's capacity

that is operated by post-innovation equipment. But if the age

distribution of equipment for different firms in the industry is

much the same, and the equipment is fairly divisible, then

this factor may not impinge significantly on the rate of inter-firm

diffusion -- which is what is of concern in this paper -- since

all firms will have some machines or equipment that is worth

(considering) replacing. Now it is very likely that the investment

required (for any significant commercial use) and durability of

equipment are really measures of indivisibility. As the size of

these variables increases, the twin assumptions that individual firms

have equipment age distributions that are identical with the

industry's and that for all firms there are some machines that are

worth replacing become less and less valid. Thus indivisibility

begins to affect the rate of inter-firm diffusion. The effect is

likely to be an increase in the dispersion of the diffusion curve

(and thus a decrease in L) and also the time required for adoption

by a given fraction of firms, because, initially, for some firms with

very new equipment, it is simply not worthwhile to consider replacing

with post-innovation equipment. Obviously, a higher rate of .capacity

expansion for firms is likely to work against this effect.

At this state, it is appropriate to briefly discuss the

difficulties involved in enriching this framework to incorporate the

possibility of firms assigning some observation value to the action

of other firms, a possibility which we had alluded to in section I.

Consider first the (empirically unrealistic) case in which all firms

are identical, i.e., they know that the true mean profitability

of the innovation, whatever it is, is the same for all of them.

If now one makes the (very unrealistic) assumption that each firm

has access to every other firm's history of observations, then it

is like decentralizing a sampling problem -- it speeds up the

sampling process if each firm faces constraints on the number of

samples per unit time, but it will lead all of them to accept or

reject simultaneously, in the absence of other complications. On

the other hand, if firms can only observe adoption decisions by

other firms, which is more realistic, then we have the problem

of what observation value, in a metric that is the same as that

used for own observations, to assign to the observation of an

adoption decision. This assignment, at any given stage of the

sampling process (so that the precision is known) clearly depends

on the optimal stopping set in this set-up. But the optimal

stopping set itself depends on what alternatives for observation

are available, and of what precision and cost. Thus a very

difficult two-way simultaneity is introduced that is non-existent

in the structure of ordinary sequential experimentation formulations --

and this difficulty makes obtaining solutions for even the most simple

cases quite difficult. When, in addition, we introduce differences

between firms -- and our first passage times approach is predicated

on a significant role for this element -- the problem becomes even more intractable. "Intuitively," it seems a reasonable conjecture that this effect should (a) make stopping sets, conditional on a given level of precision, more conservative, given the greater opportunity for observations, but (b) speed up the rate of increase of (posterior) precision because experiences of other firms are also being incorporated to some extent. Intuition then suggests that the proportion of very early adopter should be lowered due to the first effect and that the proportion of very late adopters should be lowered by the second effect. This should lower the dispersion of the diffusion curve, but have an ambiguous effect on the time taken for a given fraction to adopt. However, a rigorous support for these conjectures is well beyond the scope of this paper.

In conclusion, our (vastly simplified) approach does lead to (a) consistency with the major observed features of technology diffusion, and (b) some extensions to the effects of variables hitherto excluded. However, it is probably something of an over-simplification, and it does not permit any clear-cut test to distinguish this approach empirically from one that is based heavily on some kind of imitative behavior by firms. Given the opinions expressed by empirical researchers in this field, though, this framework with its (a) simplicity, and (b) specificity about the process involved, has something to add as a clarificatory and exploratory analysis.

Footnotes

1.   I] was measured as the ratio of the (maximum) pay-back

period demanded for projects in that industry to the ex-post

pay-back period of the innovation -- the implicit rationale being

that, for long-lived projects, the inverse of the pay-back period

is a good approximation of the (internal) rate of return.   I

was measured as the ratio of investment required in the innovation

to average assets per firm, investment required being that for

any significant commercial use.

Footnotes continued

2.   (In reference to equation (3)):

The solution to the differential equation (1) ((5), page 139) is:

$$M(t) = \frac{N[e^{\ell+(Q+\phi)t} - \frac{Q}{\phi}]}{1 + e^{\ell+(Q+\phi)t}}$$

where ℓ is a constant of integration. Using the condition that

$\lim\limits_{t \to -\infty} M(t) = 0$, one gets $Q = 0$, and

$$\frac{M(t)}{N} = \frac{1}{1 + e^{-(\ell+\phi t)}}$$

or

$$\log \left[\frac{M(t)}{N - M(t)}\right] = \ell + \phi t \tag{3}$$

Equation (3) is really the key part of Mansfield's procedure. It

is clear, of course, that with $Q = 0$, the "bandwagon" process of

equation (1) can not start (or get off the ground) with $M = 0$. But

Footnotes continued

since Mansfield only uses ϕ, which is a measure inversely related

to the dispersion of the diffusion curve, it does not serve any

great purpose to split hairs about the intercept term of equation (3).

3.   Obviously, inter-firm equality of profitability can not

be true of every innovation, for then it would be irrational of

firms to ignore others' experience. The equality, for a given

"landmark" innovation, only serves to simplify the story -- and

makes it possible to define concepts like "the" profitability of

the innovation. The crucial part of our approach is that of

independent decision-making by individual firms, at least beyond

a certain stage of shared information. One may think of the

(true mean) profitability of a technical change for different firms

as being independent drawings from a distribution that depends on

the mean rate of technical progress, industry market structure,

etc. For major innovations, with which we are concerned here, the

(values of) the drawings happen to be nearly the same for a broad

range of firms -- but they do not know that ex-ante!

Footnotes continued

4.   Norris and Vaizey (8) suggest that, due to similar

"vintage effects," innovations may spread more quickly in an

industry with a lot of old equipment, given that the profitability

of the innovation is known. Evidence showing that very significant

vintage effects exist, in the sense that average productivity may

differ greatly from the productivity of best practice plants is

provided by Salter (12).

It must be pointed out that the hypothesis that higher

durability equipment is (on average, in the cross-section) associa-

ted with greater profitability of replacement is a plausible, but

not very forceful, conjecture. The higher profitability referred

to has to be over and above that attached to the greater Net

Present Value advantage of a given excess rate of return, which

would arise from higher project life alone. The reason is that

this effect on C, by itself, will be counteracted by a similar

effect on K.

Footnotes continued

5.   This, again, is a plausible, but not necessarily
correct, conjecture. There is some reason to believe, a priori,
that innovations which require larger investments (as a fraction
of firm assets) are more "complex" to adjust to, and thus the
uncertainty of realized performance is greater with them. But
how strong this effect is is something that is difficult to comment
on.

References

1. Breakwell, J. V. and H. Chernoff, "Sequential Tests for the
    Mean of a Normal Distribution II (large t)" Annals of
    Mathematical Statistics, 1964, pp. 162-173.

2. Chernoff, H., "Sequential Analysis and Optimal Design,"
    Society for Industrial and Applied Mathematics (1972).

3. Jerome, H., "Mechanization in Industry," N.B.E.R. (1934).

4. Mansfield, E., "Industrial Research and Technological Innovation;
    an Econometric Analysis," Cowles Foundation (1968).

5. Mansfield, E., "The Economics of Technical Change," (Norton, 1968).

6. Nabseth and Ray, "The Diffusion of New Technological Processes,"
    N.I.E.S.R. (Britain) 1973.

7. Nelson, C., Peck, M. J. and Kalachek, E. D., "Technology,
    Economic Growth and Public Policy," (Brookings) 1967.

8. Norris, K. and Vaizey, J., "The Economics of Research and
    Technology," (Allen & Unwin) 1973.

9. Petkau, A. J., "Sequential Medical Trials for Comparing an
    Experimental with a Standard Treatment," Technical Report No. 1,
    Department of Mathematics, M.I.T. (1975).

10. Rosenberg, Nathan, "Perspectives on Technology."

11. Salter, W., "Productivity and Technical Change,"(Cambridge
    University Press) 1969.

Biographical Note

The author, Sudipto Bhattacharya, was born in Calcutta, India on December 3, 1951. He graduated high school from Wellington College, Wellington, New Zealand in 1967 and received a B.Sc. (Honors) degree from the University of Delhi, India in 1971 and the Post-Graduate Diploma in Business Administration from the Indian Institute of Management, Ahmedabad, India in 1973. From September, 1977 he will be employed on the faculty of the Graduate School of Business of the University of Chicago.