THE BUSINESS CYCLE AND THE STOCK MARKET

by

Andrei Shleifer
A.B., Harvard University
(1982)

SUBMITTED TO THE DEPARTMENT OF
ECONOMICS
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1986

Signature of author____
Department of Economics
May 12, 1986

Certified by____
Peter A. Diamond
Thesis Supervisor

Certified by____
Franklin M. Fisher
Thesis Supervisor

Accepted by____
Richard S. Eckaus
Chairman, Departmental Graduate Committee

ABSTRACT

The three essays of this thesis concern the role of expectations in determining the allocation of resources, particularly in the macroeconomic context. Specifically, all three papers are motivated by the proposition that private agents' beliefs are aggregated into stock market prices, which can therefore influence the allocation of investment.

The first essay does not deal with financial markets explicitly, although it explores the role of animal spirits in determining investment. The essay describes an artificial economy, in which firms in different sectors make inventions at different times, but innovate simultaneously to take advantage of high aggregate demand. In turn, high demand results from simultaneous innovation in many sectors. The economy exhibits multiple cyclical equilibria, with entrepreneurs' expectations determining which equilibrium obtains. These equilibria are Pareto ranked, and the most profitable equilibrium need not be the most efficient. While an informed stabilization policy can sometimes raise welfare, if large booms are necessary to cover fixed costs of innovation, stabilization policy can stop all technological progress.

The second essay explores theoretically an imperfectly competitive economy in which firms gauge from the distribution of share prices information about the productive opportunities of other sectors. They use this information to forecast aggregate demand, which they need to do in order to make investment decisions. When the stock market perfectly reveals technological uncertainty, profit-maximizing decisions of firms yield a unique efficient equilibrium. When the market is not perfectly revealing, there is room for multiple sunspot equilibria with different levels of income. In some cases, the information conveyed by share prices so influences investment decions as to reduce aggregate welfare.

The third essay departs from macroeconomics and deals empirically with the question: Do demand curves for stocks slope down? To this end, it looks at episodes of shifts of the demand for individual securities and considers accompanying price changes. Since September, 1976, stocks newly included into the S&P 500 Index have earned a significant positive abnormal return at the announcement of the inclusion. This return does not disappear for at least ten days after the inclusion. The returns are positively related to measures of buying by index funds, consistent with the hypothesis that demand curves for stocks slope down. The returns are not related to S&P's bond ratings, which is inconsistent with a plausible version of the hypothesis that inclusion is a certification of the quality of the stock.

Thesis Supervisors: Peter A. Diamond
Franklin M. Fisher
Titles: Professors of Economics

ACKNOWLEDGMENTS

Writing a thesis gets one deeply into debt. My debts are many, and I can only acknowledge the largest. My advisors, Professors Peter Diamond and Frank Fisher, guided this research away from many confusions. The final product would have been much inferior without their help. Professor Lawrence H. Summers trained me as an economist and had a huge impact on all the work I have done. Many people at MIT, including Professors Olivier Blanchard, James Poterba, Richard Ruback, and my third reader, Stanley Fischer, offered advice that I had the good fortune to follow. My fellow students and friends — Bob Barsky, John Bound, Greg Mankiw, and Robert Waldmann — improved my thinking and my papers considerably. To all these people I owe large debts.

But debts alone could not get this venture going. In addition, Robert Vishny took a large equity stake in these three papers, one of which he has in fact co-authored. (His liability for other essays is limited.) Working with him has been the great fun of MIT years.

Finally, I would like to dedicate these essays to my parents.

ESSAY ONE

IMPLEMENTATION CYCLES

1. Introduction.

At least since Keynes (1936), economists have suspected that an autonomous determinant of agents' expectations can lead them to do business in a way that makes these expectations come true. Several recent studies confirmed this suspicion by exhibiting economies with multiple self-fulfilling rational expectations equilibria. Most notably, Azariadis (1981) and several subsequent studies[1] investigate such equilibria in overlapping generations models, while Diamond (1982) and Diamond and Fudenberg (1982) study them in a model with search-mediated trade. None of these studies, however, focus on Keynes' specific concern about the influence of the state of long term expectation, or business confidence, on businessmen's plans to undertake or postpone investment projects. A model addressing this concern is presented in this paper[2]. In the model, entrepreneurs hold partly arbitrary, but commonly shared expectations about the future path of the economy, and independently choose a pattern of investment that fulfills these expectations. Expectations influence the cyclical behavior of macroeconomic variables, the efficiency properties of the economy, and, in some cases, long-run development as well.

Specifically, the theory describes the possibilities for both cyclical and noncyclical implementation of innovations occurring despite the steady arrival of inventions[3]. The model is one of a multisector economy, in which each sector receives ideas about cheaper means for producing its output. Such inventions arrive to each sector at a constant rate. When a firm in a sector invents a low-cost technology, it can start using it at any time after the invention. Although the inventing firm can profit from becoming the lowest-cost producer in its sector, such profits are temporary. Soon after the firm implements its invention ("innovates") imitators enter and eliminate all profits. Because of this, the firm would like to get its profits when they are the highest, which is during a general boom. Expectations about the date of arrival of this boom determine whether the firm is willing to postpone innovation until the boom comes.

If all firms owning inventions share the expectations about the timing and size of the general boom, they can time their innovations to make this boom a reality. When firms in different sectors all anticipate an imminent boom, they put in place the inventions they have saved. By innovating simultaneously, firms give a boost to output and fulfill the expectation of a boom. When, on the other hand, firms expect a boom only in the distant future, they may choose to delay implementation of inventions. When firms in different sectors postpone innovation, the economy stays in a slump. A firm in a given sector affects the fortunes of firms in other sectors by distributing its profits, which are then spent on output of firms in all sectors. In turn, they benefit when profits from other sectors are spent on their own products. By waiting to innovate, all firms contribute to the general prosperity of a boom; while the general prosperity of a boom affords them profits that are worth waiting for.

When expectations drive investment, the economy can fluctuate without fluctuations in invention. Any of a number of cycles of different durations can be an equilibrium, depending on agents' anticipations about the length of the slump. The longer is the slump, the bigger is the boom that follows it. One possible equilibrium outcome is the immediate implementation of inventions, in which case output grows without a cycle. When the economy does fluctuate, it falls behind its productive potential as firms postpone innovation, but catches up in a boom. Productivity grows in spurts. An economy with these features is described in Sections 2-3.

The possibility of a cyclical equilibirum sheds doubt on a frequently articulated view that a market economy smooths exogenous shocks. Inventions here can be interpreted as shocks hitting the economy, which are essentially identical each period. But these shocks can be "saved." If they are, the stock of technological knowledge grows steadily, but is embodied into technology periodically. The economy follows a cyclical path when a much smoother path is available.

When expectations are autonomous, the economy can end up in any of its several perfect foresight stationary cyclical equilibria. These equilibria are Pareto ranked, so the economy can settle in a very bad equilibrium. But expectations need not be truly autonomous; they may reflect agents' preferences over equilibria. For example, some equilibria may generate higher profits for all firms whose actions affect the equilibrium path. If the myriads of firms in the economy could coordinate on production plans supporting this equilibrium, it is arguably the natural outcome to expect. The question then arises whether the most profitable equilibrium is the one preferred by consumers. Section 5 shows that if innovation does not require fixed costs, the acyclical equilibrium is both the most profitable and the most efficient. In contrast, Section 6 supplies an example in which, with contemporaneously incurred fixed costs, the most profitable equilibrium is cyclical, but the most efficient one is acyclical. The example raises the possibility of a disagreement between workers and firm owners over the preferred path of the economy.

In the model discussed in Sections 2 through 6, long term development of the economy is independent of which equilibrium obtains. In the long run, all good ideas are put to use, with or without business cycles. Since cycles are inefficient, a countercyclical fiscal policy that could steer the economy to the steady growth equilibrium would be desirable. An example of such a policy, presented in Section 7, is a progressive tax system (or a tax surcharge during booms). When a government intervention reduces the profitability of innovation during booms, it can eliminate cycles and raise welfare.

In general, however, long-term development might rely on the cycle, and an ignorant countercyclical policy might be harmful. For example, if each innovator must incur a fixed cost in the period prior to innovation (e.g., he must build a plant), large sales during booms may be necessary to enable the entrepreneur to cover his fixed costs. Innovations introduced during slumps may lose money, and even steady growth equilibria may fail to sustain innovation because of an insufficient level of aggregate demand. In this case, only a cyclical equilibrium is compatible with implementation of inventions. In an alternative equilibrium (called stone age), firms never expect a boom and do not innovate at all.

If the government understands that long term growth can only be sustained with fluctuations, it will forego stabilization policy. An attempt to eliminate the cycle with aggregate demand management will at best be wasteful, and at worst will steer the economy into the stone age equilibrium. The success of countercyclical policy should be judged in light of the possibility that an ignorant policy can entail substantial welfare losses if it blocks technological progress.

While focusing on the role of expectations and on coordination, I depart significantly from a down-to-earth theory of investment. Capital in the model is a stock of knowledge embodied into a technology which uses no durable assets. Investment constitutes taking available ideas that are not being used, and adding them to the stock of utilized knowledge. The cycles are implementation cycles rather than cycles in physical investment. Knowledge, however, is a very imperfect proxy for a physical asset, since it does not offer the same opportunities for physical smoothing of consumption. Section 8 discusses the consequences of introducing capital into the model, and considers additional assumptions that must be made to preserve the results. It also discusses additional extenstions and concludes.

2. Elements of the Model.

THE CONSUMER

The household side of the economy consists of one representative consumer, who lives for an infinite number of discrete periods. The consumer's preferences are defined each period over a list of N goods which is constant over time. Lifetime utility function is given by:

$$(1)\qquad \sum_{t=1}^{\infty} \rho^{t-1} \frac{\left(\prod_{j=1}^{N} x_{tj}^{\lambda}\right)^{1-\gamma}}{1-\gamma}$$

where $\lambda=1/N$ and x_{tj} is consumption of good j in period t.

I use Cobb-Douglas preferences within a period to abstract from substitution between different goods[4]. In a model addressing macroeconomic questions, we want equilibrium in each sector to be determined by aggregate demand and not by prices in other sectors, a property guaranteed by Cobb-Douglas preferences. The infinitely lived consumer formulation assures that the results are not driven by the restricted market participation property of overlapping generations models. All the results I present also hold in a finite horizon economy.

Assume that in period t each good j is sold at the price p_{tj} on a separate market and that consumer's income is y_t. If interest rates paid at time t is denoted by r_{t-1}, the consumer's budget constraint is

$$(2)\qquad \sum_{t=1}^{\infty} \frac{y_t - \sum_{j=1}^{N} p_{tj}x_{tj}}{D_{t-1}} = 0$$

where $D_t=(1+r_1)\ldots(1+r_t)$ and $D_0=1$.

The assumption of one lifetime budget constraint for the representative consumer relies on perfect capital markets. In particular, it means that an entrepreneur can borrow against the (known with certainty) profits that will be earned from his yet unmade invention. We can think of inventors selling the claims to profits from future inventions in a competitive stock market, so that all these claims are traded from the start. The model thus allows for consumers with heterogeneous wealth levels (specifically, inventors and non-inventors), and for capital market transactions between them. Once these transactions are completed, we can think in terms of a representative consumer with a lifetime budget constraint.

Maximization of (1) subject to (2) yields consumption expenditures c_t in period t such that

$$(3) \qquad c_t^{\gamma} = [\sum_{i=1}^{N} p_{ti}x_{ti}]^{\gamma} = \rho^{t-1} D_{t-1} \cdot \{\Pi\, p_{jt}^{\lambda}\}^{\gamma-1} \cdot \frac{1}{\alpha},$$

where α is the Lagrange multiplier on (2). In addition, we get constant expenditure shares for various goods:

$$(4) \quad x_{ti}p_{ti}=\lambda c_t$$

Assume that physical storage is impossible. Competitive interest rates adjust to make $y_t=c_t$, so that the consumer wants neither to borrow nor to save. Equilibrium interest rates are given by:

$$(5) \qquad 1+r_t = \frac{1}{\rho}\cdot(\frac{y_{t+1}}{y_t})^{\gamma} \cdot \frac{\{\prod_{j=1}^{N} p_{jt+1}^{\lambda}\}^{1-\gamma}}{\{\prod_{j=1}^{N} p_{jt}^{\lambda}\}^{1-\gamma}} .$$

Finally,

(6) $y_t=\Pi_t+L$

where, Π_t are aggregate profits in period t, L is the inelastic labor supply, and wage at each t is taken to be 1 without loss of generality. Throughout, I will measure prices and interest rates in wage units.

MARKET STRUCTURE AND INNOVATIONS.

Prior to period 1, output of each good j can be produced by firms with constant returns to scale technologies output=labor input. Every period, firms play a Bertrand price game, without capacity limitations. The Bertrand assumption is equivalent to assuming competition whenever no innovation takes place, while letting the innovator to be a dominant firm in its market. In period 1, then, equilibrium prices all equal unity.

Each period, one firm in each of n sectors generates an invention. These inventions are made in a very strict order. In the first period, firms in sectors 1,...,n get ideas; in the second period, firms in sectors n+1,...,2n invent, and so on. In period $T^*=N/n$, firms in the last n sectors invent, and in period T^*+1, the next round of inventions begins with sectors 1,...,n. This order is permanent for all rounds of invention.

An invention in each sector is a technology that produces one unit of output with $1/\mu$ the labor it took to produce this output with the best technology known up to then. μ exceeds 1, and is the same for all goods and for all times. Thus ideas from the first round permit a unit of output to be produced with $1/\mu$ units of labor, those from the second round with $1/(\mu^2)$ units, etc[5].

In any period from the date it gets the invention, the firm can enter the market and implement it. Assume that the firm can postpone innovation without the danger of another firm implementing it first (until, of course, the next idea arrives to the sector). When it innovates, the firm enters a Bertrand market, in which it becomes the lowest cost producer. Equilibrium price equals the marginal cost of inefficient firms, but the innovator captures the whole market. He does not want to lower the price since demand is unit elastic, and he cannot raise it without losing all his sales. In the period after innovation, imitators enter and compete away all profits, with the price falling to the marginal cost of the efficient technology, or $1/\mu$ times the old price.

THE DECISION TO INNOVATE.

Consider what happens to a firm that innovates when aggregate demand is y_t, and the marginal cost of an inefficient producer in its sector is w_{ti}. It gets revenue λy_t for output$=\lambda y_t/w_{ti}$, obtained at a unit cost w_{ti}/μ and a total cost $\lambda y_t/\mu$. Its profits are:

(7) $\pi_t = \lambda y_t-(\lambda y_t/w_{ti})\cdot(w_{ti}/\mu)= \lambda(\mu-1)y_t/\mu$

Independence of profits of the unit cost level of inefficient firms is a special feature of the Cobb-Douglas and constant unit cost assumptions; it does not buy any important results. Each firm innovating in period t will make π_t. I use notation $m\equiv\lambda(\mu-1)/\mu$, so that $\pi_t=my_t$.

Importantly, I assume that a firm owning an invention will choose its date of innovation (and hence the only date it makes profits) to maximize the present value of profits. It may be argued that in a representative consumer economy, the firm should do what that consumer wants, and hence if profit maximization leads to an inefficiency, it is an inappropriate objective for the firm. To deal with this objection, the economy I describe can be replicated as in Hart (1982), so that owners of firms do not consume their firms' output. Suppose we have two representative consumers on two identical islands, each laboring on his own island and consuming the fruit of his own labor, but owning firms and saving on the other island. Suppose the two islands are in the same equilibrium, so that interest rates, incomes and profits are the same on both. Then any firm that the representative consumer owns cannot affect the prices he faces by altering its date of innovation. In that case, the owner's objective is profit maximization, since the firm's choices only enter their problem through the budget constraint, and capital markets are perfect. After replication, all the issues I study remain. Having one representative consumer and a profit-maximizing firm is a simplifying, but perfectly legitimate abstraction.

3. Construction of a periodic equilibrium.

The principal decision of the firm holding an invention is to determine when to innovate. In this section, I will show that firms in different sectors, receiving ideas at different times, may all choose to innovate in the period of high profits and high aggregate demand, i.e. when other firms innovate. This synchronization of innovations gives rise to a multiplicity of perfect foresight equilibria. One of them is always the steady growth acyclical equilibrium, in which inventions are implemented immediately. From the set of equilibria with fluctuating output, I focus on constant period cycles.

In a perfect foresight equilibrium, firms form expectations about the path of interest rates and of aggregate demand, and these expectations are fulfilled by the chosen timing of innovations. Firms are assumed to be small, so that each firm ignores its own impact on the behavior of aggregate variables. Similarly, when a firm makes its decisions, it only cares about aggregate data, and not about what is happening in any sector other than its own[6].

Suppose we look for cycles of period $T \leq T^*$, in which inventions accumulated in periods 1 through T are implemented together in period T, called a T-boom. Innovations are imitated in period T+1, which is also the first period of the next cycle (I shall speak in terms of periods 1, ... T, but this should be interpreted modulo T.)

The conditions for existence of a perfect foresight cyclical equilibrium of period T are twofold. First, it must be the case that if firms inventing in periods 1,...,T-1 expect the boom to take place in period T, they choose to innovate in period T rather than in the period they get their ideas or any period prior to T. Second, if firms with inventions expect a boom in period T, they must prefer not to wait past period T to innovate; in particular, they should not want to wait until the next boom. I shall take up these two conditions in order.

To find conditions for postponement, fix T and consider first the periods of no innovation. Prior to the boom, there are no profits earned in the economy and hence in periods 1, ... T-1 income is L. Since prices do not change either, interest rates are given by $1+r_1=....=1+r_{T-2}=1/\rho$. Next consider a T-boom. Since a firm will never sit on a new idea until the next idea comes to its sector and makes the first one obsolete, T must satisfy:

(8) $T<T^*=N/n$.

Since profits are the same in each innovating sector, we have:

(9) $\Pi_T=nTmy_T$,

which combined with (6) yields:

(10) $\pi_T=mL/(1-nTm)=my_T$

Note that the condition $T<1/(nm)$ is implied by (8).

We also apply (5) go get:

(11) $1+r_{T-1}=(1/\rho)(1-nTm)^{-\gamma}$,

since prices do not change from period T-1 to period T.

Now, if a firm getting its idea in period 1 is willing to wait until period T to implement it, so will firms getting their ideas in periods 2,...T-1, since the interest rate is positive throughout, and income stays constant at L. For the same reason, if a firm that gets the idea in period 1 wants to postpone implementing it beyond period 1, it will not want to implement it until time T.

We can now calculate the condition under which the firm that gets its idea in period 1 is willing to delay innovation until period T. That condition is $\pi_T/D_{T-1}>\pi_1$ or

(12) $\rho^{T-1}(1-nTm)^{\gamma-1}>1$

Expression (12) can be interpreted as follows. Profits in this model are proportional to output, while the discount factor is proportional to output raised to the power γ. The more concave is the consumer's utility, the higher must be the interest rate to keep him from wanting to borrow in the period prior to the boom. Discounted profits are thus proportional to output raised to the power $1-\gamma$. Also, $(1-nTm)$ is the share of wages in income, and hence discounted profits are proportional to $(1-nTm)^{\gamma-1}$, since the wage bill is constant. The more firms innovate, the higher is the share of profits and hence the higher is income, and the more profitable it is to innovate at that time. This effect, however, is mitigated by declining marginal utility of income and by discounting. In particular, with logarithmic utility, there can be no delay, since the discount rate is proportional to profits. I shall therefore assume throughout that $0\leqslant\gamma<1$[7]. But as long as $\gamma<1$ and (12) is satisfied, interest rates do not rise by enough prior to the boom to offset firms' preference for getting their profits during tlat boom.

When (12) holds, if all firms but one inventing in periods 1 through T are willing to wait until period T to innovate, that last firm is also willing to wait until period T. We now need to find under what conditions when all but one of the firms innovate in period T, the last one does not want to wait beyond that. For even when a firm waits until the boom, it may want to wait one more period because of price declines in period T+1 and the resulting possibility of negative interest rates (in wage units) at time T.

Two influences can keep the firm from postponing innovation beyond a boom. First, even despite price declines, discounting may be sufficient to render postponement unprofitable. Second, by the time a firm might want to innovate in the future, the next invention may arrive in its sector, thus preventing it from profiting from its own idea. I shall first provide the condition under which a firm does not want to wait beyond a boom even without a danger of its invention being surpassed, and then deal with that danger.

Observe first that if the firm does not want to wait until 2T to innovate, it will not want to wait until any period before 2T either. For if it did, it would also want to postpone innovation from then until 2T, by (12). Observe also that the firm unwilling to wait until 2T would not choose to wait beyond 2T, since delaying innovation from 2T to $2T+t \leqslant 3T$ is just like delaying innovation from T to $T+t \leqslant 2T$ (or even worse if the next idea may be coming into the sector). All we need to find, then, is the condition under which the firm prefers entering at time T to that at 2T.

To get this, apply (5) to obtain:

$$(13) \qquad 1+r_T = \frac{1}{\rho} \cdot \frac{(1-nTm)^{\gamma}}{\mu^{nT\lambda(1-\gamma)}} .$$

The power of μ appears in (13) since imitators drive the price of the goods whose production was innovated last period to $1/\mu$ of their period

T levels (and profits to 0). After imitation, the pattern of interest rates in periods T+1,..., 2T-1 repeats that in periods 1,...,T-1, and profits in period 2T are again given by (7). To prevent delay until 2T, then, we need:

(14) $\rho\mu^{n\lambda(1-\gamma)}<1$.

Condition (14) excludes the possibility that firms want to postpone their innovation indefinitely; it is also equivalent to the transversality condition for the consumer's problem, guaranteeing that lifetime utility (1) is finite in equilibrium.

The inverse of the left hand side of (14) raised to the power T is the discount factor between periods T and 2T. That inverse also equals to the interest rate that would prevail in a steady growth equilibrium. Inequality (14) says that, looking from period T, period 2T profits (which in wage units are the same) should be discounted. The problem is that prices fall after period T, and hence period T interest rate may be negative. Nevertheless, by assuming (14) we insist that, on average, the future be discounted at a positive rate, which will only happen if technological progress is not too fast[8].

When (14) holds, no firm wants to postpone innovation beyond a boom even when the next innovation in its sector will not arrive until after the next boom. When (14) fails, a firm wishes to postpone innovation until the boom just prior to the arrival of the next idea into its sector. In this case, only the cycle of length $T^*=N/n$ can be sustained as a periodic perfect foresight equilibrium, and this cycle always exists when

(14) is false. (Proof: (12) reduces to $\rho^{N/n-1}\mu^{1-\gamma}>1$, which is true whenever $\rho\mu^{n/N(1-\gamma)}>1$). Because failure of (14) leads to the conclusion of infinite lifetime utility, the usefulness of this case is unclear, and I shall ignore it from now on.

The main arguments of this section can be summarized in:

Proposition 1: Suppose the pace of innovation is slow enough that (14) holds. Then for every T satisfying (8) and (12), there exists a perfect foresight cyclical equilibrium, in which all accumulated inventions are implemented simultaneously every T periods.

This result has a simple economic interpretation. If firms can only receive profits in one period, they would like to do so at the time of high aggregate demand. The latter obtains when profits are high, and profits are high when many firms innovate. The rise in interest rates in the period prior to the boom is not sufficient to offset this preference for synchronization.

The paths of utility and of the nominal interest rate over a T-cycle are shown in Figure 1. Over time, the magnitude of the cycle stays constant thanks to Cobb-Douglas preferences, which imply that profits in each boom are the same and each round of cost reductions has the same effect on interest rates. If we detrend the GNP series, we obtain a cyclical pattern with both booms and recessions.

4. Multiplicity of Equilibria.

Proposition 1 suggests that for a given set of parameter values, there may be several periodicities T for which there exists a cycle. In particular, it is obvious that T=1 always works, so the 1-cycle -- a steady growth equilibrium in which ideas are put in place as soon as they are had -- always exists.

To study multiplicity of perfect foresight equilibria of a constant period, define left hand side of (12) as the function

$$f(T) \equiv \rho^{T-1}(1-nmT)^{\gamma-1}.$$

Remember that $mLf(T)$ is the present value that a firm inventing in period 1 attaches to its invention in a T-cycle. We are interested in T's between 1 and N/n, for which $f(T)>1$ when parameters satisfy (14). To describe this set, it is useful to start with two calculations. The proofs of these and subsequent claims are collected in the Appendix.

Lemma 1: $f(T)$ attains a minimum at a positive T_M under (14). Furthermore, $f(T)$ is decreasing for $T<T_M$, and increasing for $T>T_M$.

Lemma 2: Under (14), $f(N/n)<f(1)$.

The lemmas imply that $f(T)$ attains its minimum somewhere to the right of T=1. It may decrease all the way to N/n, or reach a minimum before N/n and rise afterwards, but not all the way to f(1). In fact, from the restrictions imposed so far we cannot ascertain either the sign of $f'(N/n)$, or whether $f(N/n)$ is greater or less than unity. The possibilities for the set of T's that keep $f(T)$ above 1 are then as follows. It can include all T's between 1 and N/n, only low T's, or both low T's and high T's, with a break in the middle. The three possibilities are shown on figure 2. They demonstrate the general multiplicity of equilibria in this model.

Multiple equilibria arise naturally when expectations govern the timing of investment. In a 1-cycle, agents always expect a constant but mild boom, and promptly innovate to make it come true. In a longer period cycle, agents expect a low level of aggregate demand for a time, and correctly anticipate the moment of a big boom. Compared to short cycles, long cycles have longer (and deeper, after detrending) slumps, but also wider spread booms[9]. In addition, there can be equilibria with variable period of the cycle. So long as firms do not want to wait until the next cycle's boom, the period of the cycle today does not affect the period of the cycle in the future. Expectations can support any one of these equilibria, as long as beliefs are common to all the market participants.

It is worth noting that, as long as (12) and (14) hold with a strict inequality, none of these equilibria is sensitive to small perturbations of the aggregate demand process. For example, suppose that a firm in some sector makes a mistake and innovates in a slump. If the impact of this mistake on aggregate demand in the boom is negligible, (12) will continue to hold, and other firms will stay with their planned timing of innovation. Equilibria are thus invariant to small exogenous fluctuations of demand.

5. Coordination, Profitability and Efficiency.

When several T-cycles qualify as equilibria, it becomes an issue which one of them should occur. An almost persuasive view argues that expectations are completely autonomous in this model, and therefore any discussion of equilibrium selection is unwarranted. Alternatively, one might ask which equilibrium firms will prefer, and then maintain that the most profitable equilibrium is the most plausible one.

To do this, consider a firm in period 1, contemplating its own profits in various equilibria. If its profits are the highest in the T^*-cycle, all firms receiving ideas up until T^* would also prefer a boom in T^* to any earlier boom. Moreover, they know they cannot have a bigger boom than the T^*-boom even after T^*, since a new round of inventions precludes delay by some firms. In this case, the T^*-cycle is the most profitable for all firms receiving ideas up until T^*, and in this sense it is focal. (Note that not all firms prefer a boom at T^*; if $T^*=3$, firms getting inventions in period 4 might well prefer the 2-cylce). Alternatively, if firms getting inventions in period 1 prefer a boom right then to a boom in any future period, it is plausible to expect them to be able to coordinate on immediate innovation. In this case, steady growth is a focal outcome[10].

In my example, discounted profits in the T-cycle for a firm getting an invention in period 1 are given by mLf(T). By lemmas 1 and 2, this quantity is the highest in the 1-cycle, which I therefore consider to be a plausible outcome. Though this weakens the case for my model as a predictor of cycles, the next section will present a generalization in which the T^*-cycle may well be the most profitable.

The next obvious question is one of the consumer's preference between equilibria. Though it is intuitively appealing that the consumer should prefer the 1-cycle to all others, this proposition is not trivial to show. True, in the 1-cycle, the consumer gets price reductions the soonest, and the production set of the economy expands at the fastest technologically feasible rate. As a result, if we compare a T-cycle with the 1-cycle, in all periods other than T-booms, the consumer is clearly better off in the 1-cycle. In T-booms, however, high profits may compensate for higher prices. Take, for example, period T^* in the T^*-cycle. In that period, we have innovation and profits in all sectors, and hence (by a standard result in tax theory) no static distortion since the markup is the same on all goods. In period T^* in the 1-cycle, we have the same production set and a distortion, since only n out of N prices exceed the marginal cost. Hence period T^* welfare is higher in the T^*-cycle. In fact, for ridiculously large rates of innovation (i.e. rates of innovation that require unreasonably small discount rates ρ for (14) to hold), the consumer prefers the T^*-cycle. For example, let $T^*=2$, $\gamma=0$, so that (12) and (14) reduce to $\rho\mu>1$ and $\rho\mu^{1/2} < 1$. We calculate that $U(1)-U(2) = (1+\rho\mu^{1/2})(\frac{2\mu}{\mu+1}) - (1+\rho\mu)$. By setting $\rho=1/4$ and $\mu=9$, we satisfy (12) and (14), while $U(1)-U(2)=-.05 < 0$. Nonetheless, the following holds:

<u>Proposition 2:</u> Assume (14) and that $\mu \leqslant T$. Then the consumer's lifetime utility is higher in the 1-cycle equilibrium than it is in the T-cycle equilibrium.

The restriction that $\mu \leqslant T$ is economically meaningless; however, it is weaker than the restriction that $\mu < 2$. If we think of a period as a year, a "reasonable" value for the size of innovation in an average sector cannot be nearly that high.

The consumer thus prefers immediate innovation, and will end up with it if firms can coordinate the timing of innovation to settle on the most profitable equilibrium. On the other hand, if expectations are truly autonomous and lead to a cyclical outcome, the consumer's welfare falls short even of its second best potential attained through immediate innovation. Of course, no equilibrium in this model is efficient.

To pinpoint the sources of inefficiency in the model, consider its deviations from the Walrasian paradigm. These are twofold. First, firms do not act as price takers and, in particular, firms recognize the effect of their innovation today on tomorrow's price. Secondly, the innovators' output improves the productive opportunities of imitators, since the latter cannot imitate until innovation took place. This externality can be described with missing markets, and turns out not to matter as long as imitators earn zero profits[11]. Absence of price taking is thus the culprit of inefficiency. In fact, it can be shown that if firms act as price-takers in a similar model generalized to allow for decreasing returns (so that there are profits in equilibrium), we will not obtain a delay of innovation even with production externalities introduced by imitation.

6. The Case of Fixed Costs.

Suppose that innovation requires a one time expenditure of F units of labor at the time the innovation is implemented. Imitation, as before, comes free in the period after innovation. This change raises the relative desirability of innovation during a large boom, and allows for the most profitable equilibrium to be cyclical.

Adaptation of the earlier analysis yields:

$$(10') \quad \pi_T = \frac{mL-F}{1-nTm} = my_T - F$$

for which we need to assume that

(17) $mL-F \geq 0$.

The condition for preference for delay from period 1 to period T then is:

$$(12') \qquad \rho^{T-1} (1-nTm)^{\gamma-1} \left(\frac{L}{L-nTF}\right)^{\gamma} > 1.$$

Fixed costs improve the possibilities for existence of T-cycles because, with fixed costs, aggregate demand is lower in a T-boom, and therefore r_{T-1} is lower. In fact, the ratio of profits in period T to what they would have been if a firm innovated in period 1 is the same with or without fixed costs. The essential difference fixed costs make is that they lower the interest rate in period T-1, and thus raise <u>discounted</u> T-boom profits. (Note that as long as $mL-F \geq 0$, $F<\lambda L$, and $nTF<nT\lambda L<L$.) The condition that a firm not wish to wait until the next boom remains (14), so we sum up with:

<u>Proposition 3:</u> Whenever there exists a T-cycle in a model without fixed costs, there also exists a T-cycle if fixed costs are low

enough ((17) holds). Furthermore, with fixed costs, a T-cycle exists for some parameter values which do not admit a T-cycle without fixed costs (e.g. logarithmic preferences).

Proposition 3 shows that multiplicity of equilibria is at least as big a problem now as it was without fixed costs. Moreover, since the relative profitability of long period equilibria rises with F, a firm obtaining an invention in period 1 may now prefer the T^*-cycle, even when (14) holds.

Lemma 3: If F is just below mL, the T^*-cycle is the most profitable, provided

(18) $\rho^{(N/n)-1} \cdot (\mu-(n/N)(\mu-1)) > 1.$

It should be noted that condition (18) implies the existence of a T^*-cycle. For parameter values satisfying (14) and (18) simultaneously, firms may very well end up in a T^*-cycle. At the same time, if F is just below mL, the consumer prefers the 1-cycle. The reason is that profits are virtually equal to 0, and hence "high" profits at T^* cannot offset the lower path of prices of the 1-cycle. In this special case, the consumer's preference for immediate innovation is clearcut. Fixed costs thus introduce the possibility of firms "choosing" an equilibrium that the consumer dislikes. If expectations accomodate this selection, we end up with a perfect foresight equilibrium whose efficiency may be substantially lower than that of a less profitable equilibrium.

7. Public Stabilization Policy and Long-run Growth.

In an economy that develops in a T-cycle, fiscal policy can eliminate fluctuations. Such a policy can raise welfare when steady growth is the socially preferred and feasible equilibrium. Consider first the economy without fixed costs, discussed in Sections 2-3. Let the government step in in the first period after a boom, and introduce a progressive income tax $\tau(y_t)$, to be imposed on all income. The proceeds of the tax are thrown into the sea.

In this economy, income and profits in period t are reduced by a factor of $(1-\tau(y_t))$, and interest rates are again given by (5), where the income is after tax. The firm should now maximize its owner's discounted after tax profits. Under these modifications, it does not pay a firm to postpone innovation until period T if τ satisfies:

$$\text{(19)} \qquad \rho^{T-1}\left\{\frac{1-\tau\left(\frac{L}{1-nTm}\right)}{(1-nTm)(1-\tau(L))}\right\}^{1-\gamma} \leqslant 1.$$

Set $\tau(L)=\tau(L/(1-nm))=0$, and for each $T>1$, let $\tau(L/(1-nTm))$ be the tax rate satisfying (19) with equality. In this case, profits are lower in the boom, so even though the interest rate prior to the boom is also lower, since $\gamma<1$, discounted profits from investing in the boom fall. The economy is thus stabilized on the 1-cycle, the government collects no taxes, but stands ready to implement its policy. A progressive tax (or tax surcharge) here resembles an automatic stabilizer (Bailey, 1978). Furthermore, at the time of the announcement, the policy has an infinite multiplier, as income jumps from L to $L/(1-nm)$ without any government expenditure. Finally, this policy raises both welfare and profits, and hence should receive universal support[12].

Even if the government sets its tax variables suboptimally, but still stabilizes the economy on steady growth, the amount of harm such a policy can do is limited. If the government sets $\tau(L/1-nm) > 0$, it collects revenues, buys goods and throws them into the sea. Still, such policy can only waste what is collected; it cannot arrest technological progress.

This, however, is not the case in a more general model. What makes the cases studied in Sections 2-6 special is that long-run development of the economy is independent of the particular cyclical path that it follows. Eventually, all good ideas are put to use, with or without business cycles. While fiscal policy stabilizes growth, it has no consequences for development.

An alternative possibility is that the cycle is essential for development. This would be the case if, for example, firms could not cover their fixed costs when they expect other firms to innovate as soon as they invent. Thus aggregate demand during steady growth is too low to sustain it as an equilibrium. Only in a boom of a cyclical equilibrium might aggregate demand be high enough to enable firms to cover fixed costs. In addition, there will exist a stone age equilibrium, in which because firms do not expect other firms to innovate, innovation is unprofitable and never takes place. Cyclical synch-ronization of innovations is thus essential for implementation of inventions. Appendix B presents an example of such an economy.

If cyclical growth is essential for innovation, stabilization policy can do more harm than good. The reason is that if taxes render innovation unprofitable even during booms, there may be no times when the firm can earn a profit from implementing its invention. As a result, the economy will settle in the stone age equilibrium. Too aggressive a fiscal policy, while getting rid of the cycle, can actually endanger technological progress.

8. Conclusion.

The examples discussed in this paper have attempted to illustrate the impact of entrepreneurs' expectations about the future path of macroeconomic variables on their decisions to undertake or postpone investment projects. An economy in which aggregate demand spillovers favor simultaneous implementation of projects in different sectors was shown to exhibit cyclical equilibria, with duration of slumps governed largely by expectations. In some examples, business cycles were either the most profitable or even unique outcomes. Furthermore, although countercyclical policy stabilizes the economy in some cases, an aggressive intervention can also interfere with long-run development[13].

The model I discussed can be amplified to study the nature of cyclical equilibria in a somewhat more realistic context. For example, if inventions come in different sizes (e.g. different μ's relative to the same F), firms with big ideas need not wait for the boom, even when firms with small ideas do. Alternatively, a very large unanticipated innovation (or some other shock) can have a large enough impact on aggregate demand so as to trigger a boom prior to its otherwise anticipated time[14]. Finally, if inventions arrive into some sectors more often than they do into others, economy wide equilibrium can consist of overlapping cycles of different periodicities, as emphasized by Schumpeter (1939).

Some extensions of the model are suggestive of ways of getting to a unique equilibrium. For example, suppose some periods (such as the Christmas season) are characterized by an especially high marginal utility of consumption, and therefore by low interest rates preceding them. The present value of profits earned from innovating in such periods might be especially high, resulting in their selection as booms. Although in the model without fixed costs cyclical equilibria with seasonal booms might be focal, they are not unique. With fixed costs, however, such equilibria can be made unique.

In evaluating the usefulness of this model, it might be fruitful to recall four conditions that seem to be responsible for cyclical equilibria. First, there must be a constantly replenished supply of pure profit opportunities. Second, these opportunities cannot be exploited forever, without pure profits being eliminated by entry. Third, profits in different sectors of the economy must spill over into higher demand in other sectors. Fourth, this spillover must be significant at the moment profits are received: intertemporal smoothing of consumption should not make the moment of receipt of income irrelevant for demand.

As this discussion suggests, I do not regard innovation to be the critical part of the story: it is simply an extremely convenient way to model temporary pure profit opportunities. Furthermore, I consider the first three conditions to be quite appropriate for a market economy.

Absence of capital, however, is a critical assumption that cannot be eliminated without substituting an alternative. For suppose we add capital to the model. Then in a period of a slump, when the consumers realize that they will be better off in the future, they will attempt to dissave and to consume now, thereby reducing the future capital stock and smoothing out consumption between periods. In this case, there will be no general boom, and implementation cycles will be impossible in equilibrium. As I specified the model, physical dissaving is not feasible because there is no capital[15]. When all the adjustment to fluctuations in income occurs through interest rates, the incentives for firms not to wait for the boom are insufficient to eliminate cycles.

With capital, we need additional assumptions to accomodate implementation cycles. First, borrowing constraints can restrict opportunities for consumption smoothing. The results of this paper can be developed in an overlapping generations model of capital with the conclusion that, if entrepreneurs cannot borrow against future profits, cyclical equilibria are feasible. An alternative formulation, which is perhaps a fruitful subject for future research, is to consider a model with durable irreversible investment as in Arrow (1968). The effect of durable capital should be to limit the amount of physical dissaving that the economy can do. As a result, durable capital may accomodate implementation cycles, though I have not verified this possibility. Finally, the economy may be subject to uninsurable and unanticipted shocks, in which case the mechanism discussed in this paper will work to accentuate cycles, though it will not cause them.

APPENDIX A.

Proof of Lemma 1:

Setting $f'(T)=0$, we obtain that $T_M = \frac{1}{nm} + \frac{1-\gamma}{\ln\rho}$.

Taking logs of both sides of (14), we get that $\ln\rho + n\lambda(1-\gamma)\ln\mu < 0$, so

$$\frac{1}{\ln\rho} > - \frac{1}{n\lambda(1-\gamma)\ln\mu} .$$

Therefore, $T_M = \frac{1}{nm} + \frac{1-\gamma}{\ln\rho} > \frac{1}{nm} - \frac{1-\gamma}{n\lambda(1-\gamma)\ln\mu} =$

$$= \frac{1}{n\lambda} [\frac{\mu}{\mu-1} - \frac{1}{\ln\mu}] > 0,$$ where

the last inequality follows since $\ln\mu > (\mu-1)/\mu$ for $\mu>1$.

The sign of the derivative of $f(T)$ is the sign of $(1-nTm)\ln\rho + nm(1-\gamma)$,

which is negative for $T < T_M$ and positive for $T > T_M$.

Proof of Lemma 2:

First, it can be verified that

(A) $\mu(1-n/N)+n/N < \mu^{1-n/N}$ when $\mu > 1$ and $0 < n/N < 1$.

Now, the claim of the lemma amount to asserting that

$$X \equiv \{\mu-(n/N)\mu\}^{1-\gamma} \cdot \rho^{N/n-1} < 1.$$

But (14) implies that $\rho < \mu^{-n/N(1-\gamma)}$, so

$$X < \{\mu-(n/N)(\mu-1)\}^{1-\gamma} \cdot \{\mu^{-(1-\gamma)n/N}\}^{(N-n)/n}$$

$$= [\{\mu-(n/N)(\mu-1)\} \cdot \mu^{-(N-n)/n}]^{1-\gamma},$$

The last expression is smaller than 1 provided $\mu-(n/N)(\mu-1) < \mu^{1-n/N}$, which is just (A).

Proof of Lemma 3:

Applying (10'), we can compute that

$$\pi_{T^*} = \rho^{N/n-1} \cdot \mu^{1-\gamma} \cdot \left[\frac{L}{L-NF}\right]^{\gamma} \cdot (mL-F), \text{ and}$$

$$\pi_1 = \frac{mL-F}{1-nm} .$$

Therefore, $$\frac{\pi_{T^*}}{\pi_1} = \mu^{1-\gamma} \cdot \left[\frac{L}{L-NF}\right]^{\gamma} \cdot (1-nm) \cdot \rho^{N/n-1}$$

$$\approx \mu(1-nm)\rho^{N/n-1} \quad \text{when } F \approx mL.$$

But $\mu(1-nm)=(\mu-(n/N)(\mu-1))$, so when (18) holds, $\frac{\pi_{T^*}}{\pi_1} > 1$, as claimed.

Proof of Proposition 2:

Notation: Denote $\alpha=n/N$ so that $T^*=1/\alpha$. Let L(T) be the equilibrium lifetime utility of the agent in a T-cycle equilibrium for $T=1,\dots, T^*$. I will prove later that to compare L(1) with L(T) it is enough to look at the first T periods. Accordingly, denote by U(T) the total utility attained over the first T periods in a T-cycle equilibrium, and U(1) the total utility attained over the first T periods in a 1-cycle equilibrium. Denote by v(T) utility in a T-boom. Also, for each of the first T periods of the 1-cycle equilibrium, define:

$f_t = \prod_{j=1}^{N} x_{tj}^{\lambda}$, where x_{tj} is equilibrium consumption of good j in period t in the 1-cycle, and define g_t as the corresponding quantity for T-cycle.

Thus, $U(1) = \sum_{i=1}^{T-1} \rho^{i-1} (f_t^{1-\gamma})/(1-\gamma)$ and $U(T) = \sum_{i=1}^{T-1} \rho^{i-1} (g_t^{1-\gamma})/(1-\gamma)$.

The proof is long and tedious, so I outline the steps first. Step 1 proves that, when comparing welfare in a 1-cycle equilibrium to that in the T-cycle equilibrium, it is enough to look at the first T periods of the individual's life. Steps 2 and 3 restrict the parameter space to the cases that are most favorable to T-cycle welfare: Step 2 shows that it is sufficient to look at $\gamma=0$, and Step 3 shows that, for $\gamma=0$, it suffices to look at the highest permissible ρ, which by (14) is $\rho=\mu^{-n/N}$. Step 4 shows that v(T) is never greater than μ^{α}, the value it attains when $T=T^*$. Step 5 shows that $U(1) > U(T)$ as long as $\mu<T$ and completes the proof. Step 6 supplies a counterexample for $\mu>T$.

<u>Step 1:</u> $L(1)>L(T)$ iff $U(1)>U(T)$.

For both the 1-cycle equilibrium and the T-cycle equilibrium, the history of periods $T+1,\ldots,2T$ is the same as the history of periods $1,\ldots,T$, except nT prices are $1/\mu$ of their old levels. This makes the utility in period $T+x$ (where $x=1,\ldots,T$) equal to $\rho^T\cdot\mu^{nT\lambda(1-\gamma)}$ times the utility in period x. Extending this argument to future periods, we get:

$$L(1)=\frac{U(1)}{1-\rho^T\mu^{nT\lambda(1-\gamma)}} \qquad \text{and} \qquad L(T)=\frac{U(T)}{1-\rho^T\mu^{nT\lambda(1-\gamma)}}.$$

Inequality (14) ensures that these expressions are finite, and thus the assertion is proved.

<u>Step 2:</u> Other things equal, if $U(1)>U(T)$ for $\gamma=0$, then $U(1)>U(T)$ for $\gamma>0$.

Observe that $\partial U(1)/\partial f_t = (f_t)^{-\gamma}\cdot\rho^{t-1}$ and $\partial U(T)/\partial g_t = (g_t)^{-\gamma}\cdot\rho^{t-1}$. In the 1-cycle, income (in wage units) stays constant over time, but prices are falling. Thus,

(A) $f_i>f_j$ for $i>j$.

Also, before period T, the individual enjoys both higher income and lower prices in the 1-cycle than he does in the T-cycle. Thus,

(B) $g_i<f_i$ for $i<T$.

If $g_T < f_T$ also, as may be the case if $T<T^*$, we are done with proving that $U(1) > U(T)$. The question is what happens when

(C) $g_T > f_T$,

so we assume from now on that this is the case. Our maintained hypothesis is that

$$\sum_{i=1}^{T}\rho^{i-1}f_i > \sum_{i=1}^{T}\rho^{i-1}g_i.$$

Now suppose that $\gamma>0$, and apply the mean value theorem to find that:

$U(T)-U(1) = \sum_{i=1}^{T}\rho^{i-1}h_i^{-\gamma}(g_i-f_i)$, where

(D) $f_T < h_T < g_T$, and

(E) $g_i < h_i < f_i$ for $i<T$.

Now combine (D), (A), and (E) to show that $h_T > f_T > f_i > h_i$ for $i<T$.

The last inequality implies that, for any $i<T$, we have

(F) $h_i^{-\gamma} > h_T^{-\gamma}$.

Using (F) and (B), we then obtain that

$$\sum_{i=1}^{T}\rho^{i-1}\, h_i^{-\gamma}(g_i-f_i) < \sum_{i=1}^{T}\rho^{i-1}\, h_T^{-\gamma}(g_i-f_i)$$

$$= h_T^{-\gamma}\left(\sum_{i=1}^{T}\rho^{i-1}g_i - \sum_{i=1}^{T}\rho^{i-1}f_i\right) < 0 \text{ by}$$

assumption. This proves the claim.

<u>Step 3:</u> If $\gamma=0$ and $U(1)>U(T)$ for some ρ_1, then $U(1)>U(T)$ for all $\rho<\rho_1$.

We are taking step 2 into account, and also assuming that

$$\sum_{t=1}^{T-1}\rho_1^{t-1}(f_t - g_t) > \rho_1^{T-1}(g_T - f_T).$$

Multiply both sides by $\dfrac{\rho^{T-1}}{\rho_1^{T-1}}$ to obtain:

$$\sum_{t=1}^{T-1}\frac{\rho^{T-1}}{\rho_1^{T-t}}(f_t-g_t) > \rho^{T-1}(g_T - f_T).$$

But $\frac{\rho^{T-1}}{\rho_1^{T-t}} < \rho^{t-1}$ when $\rho<\rho_1$. Hence

$$\sum_{t=1}^{T-1} \rho^{t-1}(f_t - g_t) > \sum_{t=1}^{T-1} \frac{\rho^{T-1}}{\rho_1^{T-t}} (f_t - g_t) > \rho^{T-1} (f_T - g_T). \quad \text{QED.}$$

Since (14) imposes an upper bound on ρ in terms of μ, and since we taking $\gamma=0$, we assume from now on that $\rho = \mu^{-\alpha}$.

Step 4: Utility in a T-boom, v(T), is bounded above by μ^{α}.

A calculation reveals that

$$v(T)= \frac{1}{1-T\alpha\,\frac{\mu-1}{\mu}}\,\rho^{T-1} = \rho^{T-1}\,\frac{\mu}{(T^{*}-T)\alpha(\mu-1) + 1} = \frac{\mu^{(T^{*}-T)\alpha}}{(T^{*}-T)\alpha(\mu-1)+1}\,\mu^{\alpha}.$$

The last equality follows since $\rho=\mu^{\alpha}$. By the mean value theorem,

$\mu^{(T^{*}-T)\alpha} = 1 + (\mu-1)(T^{*}-T)\alpha\cdot y^{-T\alpha}$, for some $1<y<\mu$. Then $y^{-T\alpha} < 1$, and so

$\mu^{(T^{*}-T)\alpha} < 1 + (\mu-1)(T^{*}-T)\alpha$. This implies, using the last expression for v(T) that $v(T) < \mu^{\alpha}$.

Step 5: $U(1) > U(T)$ for $\mu < T$.

• When $\rho=\mu^{-\alpha}$ and $\gamma=0$, we can compute that

$$U(1) = \frac{1}{1-\alpha\,\frac{\mu-1}{\mu}} \cdot T$$

Applying the mean value theorem to $f(x)\equiv 1/x$ for x between 1 and $1-\alpha\frac{\mu-1}{\mu}$,

we obtain that $U(1)= T\{1-\alpha \frac{\mu-1}{\mu}(\frac{-1}{(1-Y)^2})\}$ for some $0 < Y < \alpha \frac{\mu-1}{\mu}$.

But then $U(1) > T\ (1 + \alpha \frac{\mu-1}{\mu})$.

• When $\rho=\mu^{-\alpha}$, $\gamma=0$ and $v(T)<\mu^{\alpha}$, we have

$U(T) \leqslant 1+ \mu^{-\alpha}+\ldots+ \mu^{(-T+1)\alpha}+ \ \mu^{\alpha} < (T-1) + \mu^{\alpha}$. The last inequality follows since each of the T-2 middle terms is less than 1.

Applying the mean value theorem for $f(x)=x^{\alpha}$ for x between 1 and μ, we obtain that $(T-1) + \mu^{\alpha} = (T-1) + 1 + (\mu-1)\alpha Z^{\alpha-1}$ for some $1 < Z < \mu$.

But since $Z^{\alpha-1} < 1$, we have $U(T) < T + \alpha(\mu-1)$.

• Putting the bounds on U(1) and U(T) together, we get

$U(1) - U(T) > T + T\alpha \frac{\mu-1}{\mu} - T - \alpha(\mu-1) = \alpha(\mu-1)(\frac{T}{\mu} - 1) > 0$ by assumption.

This completes the proof.

Appendix B.

This appendix describes an economy which grows either in cycles or with no innovation at all. Suppose that an innovating firm must incur a fixed cost in the period prior to innovation. If the innovation reduces the unit cost from that of the currently used technology by a factor of μ, then this fixed cost is F. If, however, a round of innovation has been skipped, and the innovation improves the currently used technology by a factor of μ^2, then the fixed cost is 2F (similarly 3F for μ^3, etc.). This technology captures the notion that more dramatic innovations are costlier to implement. In other respects, the economy is the same as that of Section 2.

Calculation of equilibria generally follows Section 3, except now y_t exceeds c_t by the amount of fixed cost investment needed for period t+1 innovation. If an innovation requires a fixed cost aF, the cost to the firm is $aF(1+r_t)$, since aF must be saved in period t and savings earn interest. Thus fixed costs are a limited form of capital. With savings in the model, interest rate expressions must be modified to allow for divergence of income from consumption.

Consider an economy in which $T^*=2$ and $\gamma=0$. Thus half of all sectors receive an invention each period. I will present an example in which (I) equilibria where each sector implements every n'th round of invention as soon as it arrives, while skipping the first n-1 rounds, do not exist (in particular, for n=1, steady growth equilibrium does not exist), (II) the 2-cycle exists, (III) the stone age equilibrium, in which entrepreneurs expect no innovation to take place — and none does — also exists.

Requirement (I) amounts to the condition that an innovating firm's profits be negative in an equilibrium where every k periods N/2 sectors reduce costs by a factor of μ^k, followed next period by the same reduction in the other N/2 sectors. For k=1, this is the steady growth equilibrium. The condition that ensures that (I) is satisfied is:

$$\text{(B1)}\quad \frac{\mu^k-1}{\mu^k} \cdot \frac{1}{N}\left(\frac{L-\frac{N}{2}Fk}{1-\frac{\mu^k-1}{2\mu^k}}\right) - \frac{Fk}{\rho\mu^{k/2}} < 0, \qquad \text{for } k=1,2.....$$

To satisfy requirement (II), we calculate the 2-cycle. The interest rate before the boom is $1+r_1=1/\rho$, and the interest rate before the slump is $1+r_2=1/\rho\mu$, since all prices fall after the boom. Profits are given by:

$$\text{(B2)}\quad \pi_1 = m(L-NF) - F(1+r_2) = \frac{1}{N}\,\frac{\mu-1}{\mu}\,(L-NF) - F(1+r_2)$$

$$\text{(B3)}\quad \pi_2 = m\mu L-F(1+r_1) = \frac{1}{N}\,(\mu-1)L - F(1+r_1),$$

in the slump and boom, respectively. We are seeking parameters for which

$$\text{(B4)}\quad \frac{\pi_2}{1+r_1} > 0 > \pi_1 \quad , \text{ or}$$

$$\text{(B5)}\quad \rho\{\frac{1}{N}\,(\mu-1)L\} - F > 0 > \frac{1}{N}\left(\frac{\mu-1}{\mu}\right)(L-NF) - \frac{F}{\rho\mu}$$

The last condition for the 2-cycle is (14) from Section 3, which here reduces to

$$\text{(B6)}\quad \rho\mu^{1/2} < 1.$$

Finally, for the stone age equilibrium to exist (requirement III), it must be unprofitable for a firm to innovate alone in a period, or

(B7) $\frac{\mu-1}{\mu} \; \frac{1}{N} L \; - \; \frac{F}{\rho} < 0.$

Let L=100, μ=2, ρ=.7 and NF=60. For these parameter values, conditions (B1) and (B5)-(B7) can be shown to hold. In this case, equilibria with innovation but without synchronization do not exist, and bunching is necessary for technological progress.

Macroeconomic stabilization policy of the type described in Section 7 will not work in this economy. To establish this, let τ_1 and τ_2 be the income tax rates for busts and booms, respectively, and observe that, as before, taxation reduces consumption and income. Because the consumer's utility is linear in income, after tax interest rates will not be affected by the imposition of taxes. However, a firm pays for its plant at pretax interest rates, which increase as a result of imposition of the tax. These interest rates are given by:

(B8) $(1+R_1) = \frac{1}{\rho(1-\tau_2)}$

(B9) $(1+R_2) = \frac{1}{\rho\mu(1-\tau_1)}$

In this example, taxation strictly raises interest rates that firms have to pay for their capital. As a result, if innovation was not profitable in a slump before taxes were introduced, it will not be profitable with taxes. Nor will taxes make innovation in a steady growth equlibrium profitable, thereby permitting such equilibrium to reappear. Furthermore, a firm may no longer be able to break even if it innovates in a boom. Profits in a 2-boom are now given by

(B10) $\pi_2 = (\frac{1}{N}(\mu-1)L - F(1+R_1)) \cdot (1-\tau_2)$

For the parameter values from the example and $\tau_2 > .15$, π_2 is negative. This means that a tax rate of 15% or higher on a 2-boom's income sends the economy into the stone age equilibrium, which exists regardless of the level of the tax.

FIGURE 1
THE PATHS OF UTILITY AND INTEREST RATES OVER THE CYCLE

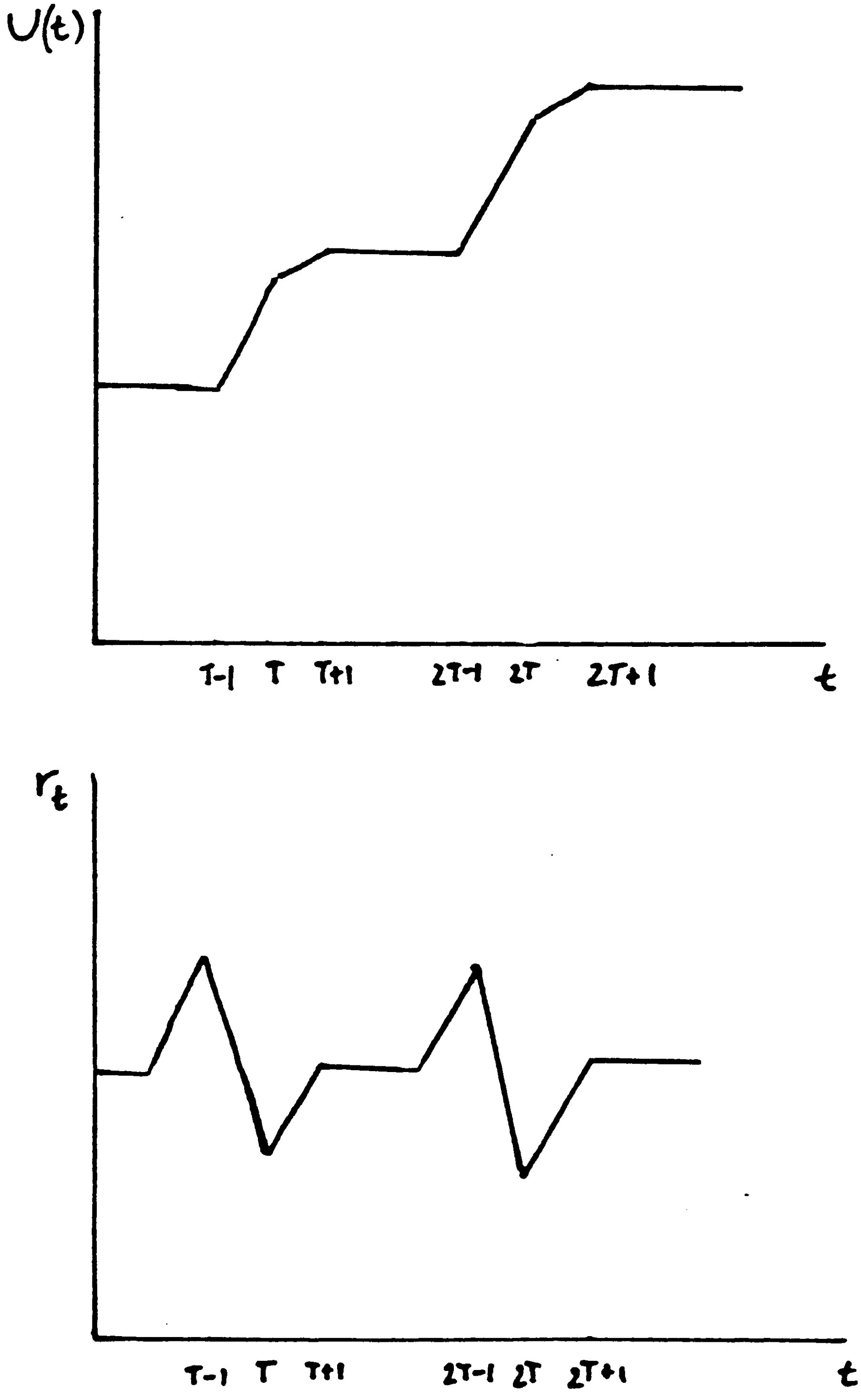
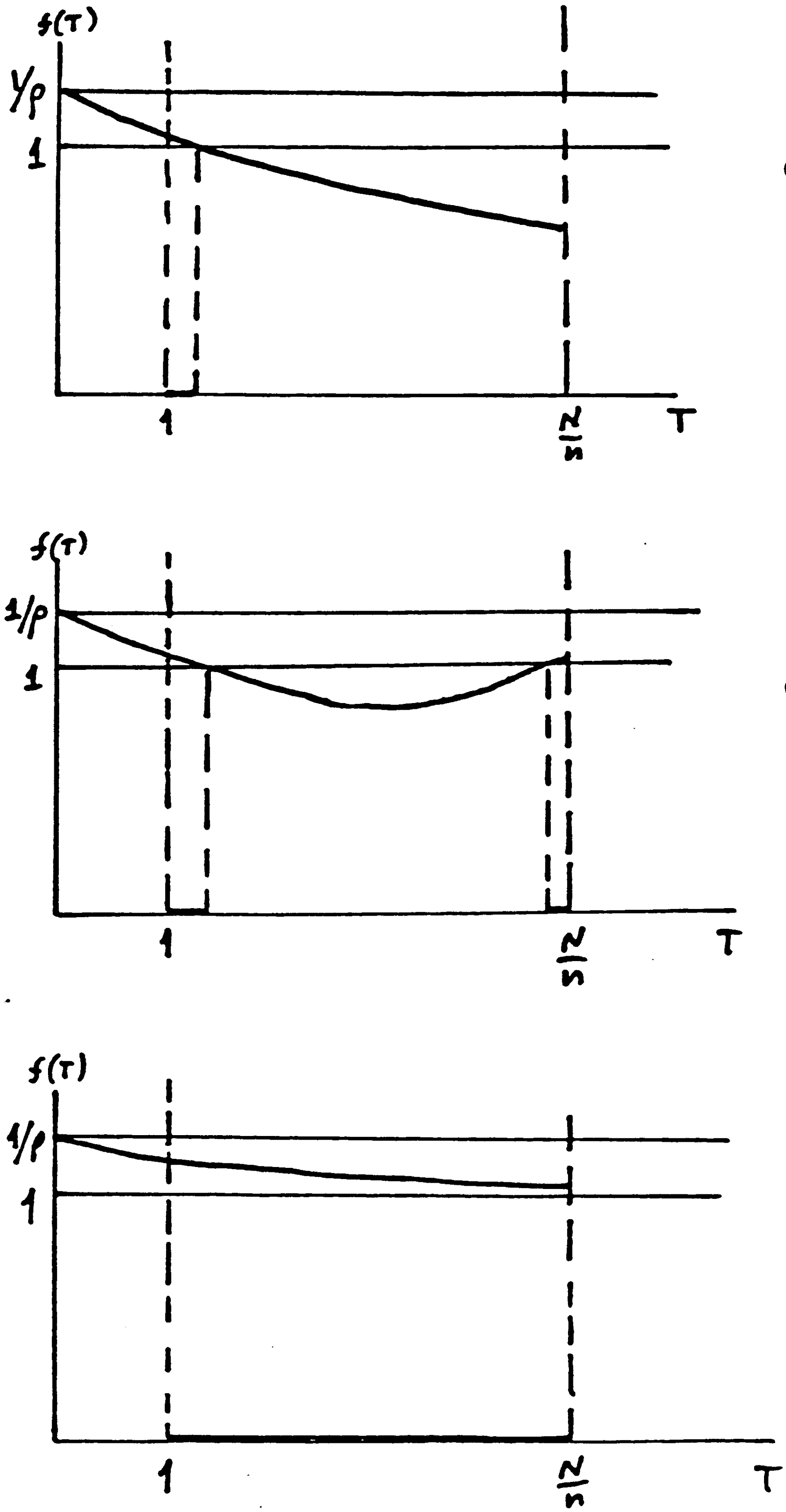


FIGURE 2

MULTIPLICITY OF CYCLICAL EQUILIBRIA



(A) only short cycles are feasible

(B) short and long cycles are feasible

(C) all cycles are feasible

FOOTNOTES.

1. Cass and Shell (1983), Grandmont (1983), Farmer and Woodford (1984).

2. Several recent theoretical studies are related to the current discussion. Weitzman (1982) and Solow (1984) discussed multiple Pareto ranked equilibria in economies with increasing returns; my work sheds doubt on the importance they attribute to the increasing returns assumption. Rotemberg and Saloner (1984, 1985) study oligopolies that intensify exogenously started booms through price wars or inventory buildups. Their work shares with mine the "unsmoothing" character of private agents' response to macroeconomic fluctuations, though they do not model endogenous cycles. Finally, Judd (1985), discusses a completely different mechanism that leads to innovation cycles. In his model, there is an infinite supply of possible innovations, but when firms introduce too many new products within a short time, they compete for the same consumer resources and reduce profits from each particular product. Furthermore, when imitation leads to price reduction for recently introduced products, consumers substitute towards these products, making entry by yet newer products even less profitable. As a result, after a period of innovation, entrepreneurs wait until the secular growth of the economy renders further innovation profitable. Judd's mechanism is almost the opposite of mine: innovations in his model repel rather than attract other innovations.

3. Because of the essential role played by innovation in this paper, the theory of business cycles it presents might (incorrectly) be thought

to be Schumpeterian (1939). Schumpeter thought the innovation process to be essentially autonomous and completely independent of market demand. His inventors create markets rather than adapt to enter good markets. In contrast, Schmookler (1962, 1966) believed that innovation occurs in markets where demand is substantial and profits from innovation can be great. My theory, then, is more Schmooklerian than Schumpeterian. Schmookler, however, insisted that expectations are adaptive, and that innovation takes place in those markets where demand has proven to be high. My work, in contrast, emphasizes foresight as a determinant of the timing of innovation.

[4]. Cycles of the type I discuss will be the easier to sustain, the less substitutibility there is between different goods.

[5]. Because prices are rapidly falling, it is important to keep in mind that restrictions on the speed of innovation are necessary to keep life-time utility finite. These will be discussed later in the paper.

[6]. I could alternatively assume a continuum of infinitesimal sectors, except for the difficulty of writing a utility function for a continuum of goods. The essential assumption is that firms take the behavior of aggregate variables as given, and ignore their own impact on those variables. It is misleading, therefore, to interpret this model as a game between sectors.

[7]. In contrast, Grandmont (1983) requires high γ's to generate a cycle in an overlapping generations model.

[8]. An alternative interpretation of (14) can be made if we take as

numeraire the price of a good whose production is innovated in period T. Then by period 2T, the real price of each good whose production has not been innovated at time T rises by a factor of μ, as do the real wage, aggregate income and profits. Condition (14) says that the real discount rate between periods T and 2T actually exceeds μ. As before, it amounts to saying that the force of time preference dominates the rate of increase of real income and profits.

9. In contrast, Diamond and Fudenberg (1982) exhibit a business cycle which at every stage provides agents with a lower utility flow than does the good stationary equilibrium.

10. In this model, as well as in the extension with fixed costs from Section 6, cycles of period T, with $1<T<T^*$, cannot be the most profitable for all firms receiving inventions between times 1 and T.

11. Because of constant returns, imitators always earn zero profits (since they always play the Bertrand game against at least the innovator). Hence they cannot afford to pay anything for the right to imitate. Even if we open personalized markets for imitation rights (following Arrow (1970)), but let the innovator set the price, he could sell nothing at any positive price; nor will he give away his idea at zero price. As a result, personalized markets will clear at a small positive price, at which both supply and demand equal zero in all periods. Put another way, in making his decisions the innovator can ignore these markets, and therefore their absense is irrelevant.

12. One potential disadvantage of this policy is that it is not subgame

perfect. When it comes to a boom, the government would prefer not to tax and waste the income.

13. This paper has not considered micro interventions, such as patent policy. For example, a policy granting each innovator a permanent patent ensures immediate implementation of inventions. Such policy, however, is socially very costly.

14. Farrel and Saloner (1984) studied a different mechanism for such a domino effect; some results similar to theirs hold in my model.

15. With fixed costs standing in for capital, as in Section 7, physical dissaving again is not feasible.

REFERENCES.

Azariadis, Costas (1981), Self-fulfilling Prophecies, Journal of Economic Theory, vol. 25, pp. 380-396.

Arrow, Kenneth J. (1968), Optimal Capital Policy with Irreversible Investment, in Wolfe, ed., "Essays in Honor of J. R. Hicks."

Arrow, Kenneth J. (1970), Political and Economic Evaluation of Social Effects of Externalities, in Margolis, ed., "The Analysis of Public Output," NBER.

Bailey, Martin N. (1978), Stabilization Policy and Private Economic Behavior, Brookings Papers on Economic Activity 1978:1, pp. 11-60.

Cass, David and Karl Shell (1983), Do Sunspots Matter?, Journal of Political Economy, vol. 91, pp. 193-227.

Diamond, Peter A. (1982), Aggregate Demand Management in Search Equilibrium, Journal of Political Economy, vol. 90, pp. 881-894.

Diamond, Peter A. and Drew Fudenberg (1982), An Example of "Animal Spirits" and Endogenous Business Cycles in Seach Equilibrium, MIT Mimeo.

Farmer, Roger and Michael Woodford (1984), Self-Fulfilling Prophecies and the Business Cycle, CARESS Working Paper #84-12.

Farrel, Joseph and Garth Saloner (1984), Standardization, Compatibility, and Innovation, MIT Working Paper #345, April.

Grandmont, Jean-Michel (1983), Endogenous Competitive Business Cycles, Discussion Paper No. 8316, CEPREMAP.

Judd, Kenneth (1985), On the Performance of Patents, Econometrica, vol. 53, no. 3, pp. 567-586.

Hart, Oliver D. (1982), A Model of Imperfect Competition with Keynesian Features, Quarterly Journal of Economics, vol. 97, pp. 109-138.

Keynes, John M. (1936), "The General Theory of Employment, Interest, and Money," London: Macmillan.

Rotemberg, Julio and Garth Saloner (1984), A Supergame-Theoretic Model of Business Cylces and Price Wars During Booms, M.I.T. Working Paper #349, July.

Rotemberg, Julio and Garth Saloner (1985), Strategic Inventories and the Excess Volatility of Production, M.I.T. Mimeo, April.

Schmookler, Jacob (1962), Invention, Innovation, and Business Cycles, in "Variability of Private Investment in Plant and Equipment, Part II, Some Elements Shaping Investment Decisions," materials submitted to Joint Economic Committee of Congress of the United States, 87'th Congress, Washington, D.C.: Government Printing Office, pp. 45-55.

Schmookler, Jacob (1966), "Invention and Economic Growth", Cambridge: Harvard University Press.

Schumpeter, Joseph E. (1939), "Business Cycles", New York: Harper & Row.

Solow, Robert M. (1985), Monopolictic Competition and the Multiplier, M.I.T. Mimeo.

Weitzman, Martin L. (1982), Increasing Returns and the Foundations of Unemployment Theory, Economic Journal, vol. 92, pp. 787-804.

ESSAY TWO

THE STOCK MARKET AS A CAUSE OF INVESTMENT

(WITH ROBERT W. VISHNY)

1. Introduction.

Numerous recent studies have documented the stock market's distinctive performance as a predictor of investment and output. Thus Moore (1983), Fischer and Merton (1984), and Fama (1981) showed that the market anticipates investment and output, while Doan, Litterman and Sims (1984) found that the stock market Granger causes investment, even controlling for other macroeconomic variables. Since the stock market is a forward looking variable, these results are not surprising. The question that is of some interest, however, is whether the market is just a passive predictor of the future or whether it actually causes investment. In this paper, we formally develop a theory of an investment causing stock market. In doing so, we hope both to evaluate the ability of such a theory to conform to the empirical findings mentioned above, and to conduct a welfare analysis of the informationally efficient stock market in a macroeconomic model.

In our theory, agents learn from share prices information that is helpful in forecasting aggregate demand. Unlike in the standard signal extraction macro models (Lucas (1972), Grossman and Weiss (1982)), where agents use macroeconomic variables such as money and interest rates as a source of information, agents in our model use the whole distribution of share prices to gauge the productive potential of the economy. Stock prices are a good source of such information since self-interest prompts informed agents to incorporate their private information into stock prices (Grossman (1976)). By observing the whole array of stock prices, firms forecast aggregate demand more accurately, and choose their investment policies based on these forecasts. Since movements in stock prices alter agents' forecasts and hence their investment decisions, the stock market causes investment.

To set up a benchmark for evaluating economies with imperfectly informed firms, Section 2 presents a full information economy. In the model, firms have increasing returns technologies, with fixed costs varying across firms. All agents know the actual distribution of fixed costs (which will be uncertain in subsequent sections). Each firm must decide whether to utilize its increasing returns technology or to leave the market to less efficient fringe firms. The profit-maximizing choice depends on expected demand, since only in a large enough market can the efficient technology break even. Demand, in turn, depends on profits of other firms, since profits are distributed to the consumer and spent by him. Aggregate demand spillovers through the distribution of profits make firms interested in the productive potentials of other firms in the economy. Since profitability of each sector depends on the level of fixed cost that the efficient firm in that sector faces, each firm uses its knowledge of the economy to forecast aggregate profits and demand, and thus to make an accurate investment decision. In the benchmark case of perfect information, the economy has a unique perfect foresight equilibrium, in which investment decisions are efficient. In other words, a perfectly informed planner would have each firm make the same decision as it does in the free market equilibrium.

In contrast, Section 3 presents the same economy, except now agents have imperfect knowledge about technological opportunities of other sectors. They make forecasts of aggregate demand based on their priors as well as observation of their own fixed cost. In this case, rational expectations equilibria exist, but are not, in general, unique or efficient. The sources of inefficiency are twofold. The first is the inability of firms to accurately condition their investment choices on circumstances of other sectors, since decisions must be made on the basis of imperfect information. An eqully well informed central planner would face the same difficulty. The second source of inefficiency stems from divergence of private and social interests in the presence of externalities. A firm's losses (gains) have adverse (beneficial) impact on profits of other firms, and the firm ignores this impact in making investment decisions. While this externality is immaterial in our complete information example, it can distort investment decisions under incomplete information.

Section 4 shows how an introduction of an informationally efficient stock market into the economy of Section 3 can solve its problems. If claims to each firm's profits are traded, then each firm's stock price will reflect insiders' knowledge of its fixed cost. Agents can then condition their investment decisions on the distribution of share prices of other firms. In our economy, share prices are completely revealing, and the unique stock market equilibrium is the perfect foresight equilibrium of the full information economy. In equilibrium, the stock market homogenizes possibly diverse expectations and focuses them on the accurate prediction of income. Since it contains information that is not available elsewhere, the stock market performs as a leading indicator and a cause of investment.

The results of Section 4 rely on there being only one dimension of payoff-relevant uncertainty. One stock price is then capable of revealing all the needed information about a firm. More generally, uncertainty might concern, for example, both present and future opportunities of a firm. Accordingly, Section 5 presents a model with uncertainty both about each firm's fixed cost and about the timing of its investment opportunity. In that model, share prices are not fully revealing, and firms make decisions in the face of residual uncertainty.

The positive and normative properties of the noisy and fully revealing stock markets differ substantially. Although the noisy stock market remains a leading indicator and a cause of investment, it can now make mistakes. Our results thus suggest how the stock market can be a good leading indicator on average, while still predicting "nine of the last five recessions." In Section 6 we attempt the beginnings of the analysis of normative properties of a noisy stock market by means of an example. Since residual uncertainty remains, equilibrium investment decisions are not always efficient, in contrast to the revealing stock market case. More importantly, the example shows that, in the presence of demand externalities, an introduction of an information-producing stock market can entail a welfare loss relative to the uninformed model.

In Section 7, we briefly summarize our results and compare them to a model where the market is only a passive predictor. Our reasoning suggests why available econometric evidence cannot distinguish the two models. Our analysis is therefore intended to clarify the properties of a causal stock market, rather than to suggest how to distinguish it from a passive leading indicator.

2. The Full Information Economy.

The benchmark economy described in this section sets the stage for the subsequent analysis[1]. It shares with the models to follow the assumptions about preferences, technology and markets, but uses a particularly simple information structure.

Consider a one period economy with a representative consumer, who has Cobb-Douglass preferences defined over a unit interval of goods. All goods have the same expenditure shares. Thus, when his income is y, the consumer can be thought of as spending y on every commoditiy. The consumer is endowed with L units of labor, which he supplies inelastically, and he owns all the profits of this economy. Taking his wage as numeraire, his budget constraint is given by:

(1) $y = \Pi + L$,

where Π is aggregate profits.

Each good is produced in its own sector, and each sector consists of two types of firms. First, each sector has a competitive fringe of firms which convert one unit of labor input into one unit of output with a constant returns to scale technology. In addition, each sector has a unique firm ("monopolist") that has access to an increasing returns to scale technology. This technology requires the input of F units of labor to cover the fixed cost, where F can be either low, F_1, or high, $F_2>F_1$, but also yields $\alpha>1$ units of output for each unit of labor input after that. The fraction of sectors where the monopolist's fixed cost is F_1 is denoted by n; the remaining 1-n sectors have monopolists with a fixed cost F_2. In this section, the fixed cost of each sector is publicly known (displayed on a billboard), and therefore n is publicly known. Much of this paper examines the consequences of uncertainty about n.

The monopolist in each sector decides whether to use his technology or to abstain from production altogether. He uses his technology ("invests") only if he can earn a profit. The price he charges if he produces equals unity, since he loses all his sales to the fringe if he charges more, and he would not want to charge less when facing a unit elastic demand curve. When income is y, the profit of a monopolist with a fixed cost F (where F is either F_1 or F_2) is:

$$(2)\quad \pi = \frac{\alpha-1}{\alpha}\, y \;-\; F \equiv ay - F.$$

The monopolist invests as long as $y \geqslant F/a$. We assume that $a\cdot L - F_1 > 0$, so that low fixed cost producers always invest in equilibrium, and that $a\cdot L - F_2 < 0$. Thus the choice of the F_2-monopolist will always be the problem to solve.

The reason for combining monopoly and a competitive fringe in each sector is as follows. Since we are interested in firms' investment decisions as a function of expected demand, we need firms that take demand curves, rather than prices, as given. To this end, we introduce monopoly. Unfortunately, a monopolist facing a unit elastic demand curve charges an infinite price. To avoid this we bring in the fringe. With more substitution between goods, the fringe would not be necessary, although it would be harder to keep track of relative prices.

The assumption of only two types of fixed costs suggests two candidate pure strategy equilibria. In the first, called optimistic, F_2-firms invest; in the second, called pessimistic, they do not.

Assuming that F_2-firms invest, aggregate profits in the optimistic equilibrium, for a fixed n, will be given by:

(3) $\Pi(n)=n(ay^o(n) - F_1) + (1-n)(ay^o(n) - F_2)$.

Combining (3) with (1), we obtain the expression for income, namely,

(4) $y^o(n)=\alpha(L - nF_1 - (1-n)F_2)$.

The term in parentheses is labor used in actual production (after covering fixed costs); at unit prices, income is equal to the output produced with that labor. For the optimistic equilibrium to exist for this n, F_2-firms must choose to invest when they expect demand $y^o(n)$, or

(5) $ay^o(n) - F_2 \geqslant 0$.

Using (4), we can restate (5) as a condition on n; specifically,

(6) $n \geqslant \dfrac{F_2 - L\cdot a}{(F_2 - F_1)a} \equiv \bar{n}$.

Only for n's above the cutoff level $\bar{n}$ will the economy have a large enough productive potential to sustain profitable investment by inefficient monopolists. When (6) is satisfied, the economy has an optimistic equilibrium with income given by (5). When (6) fails, such an equilibrium does not exist[2].

Proceeding analogously, we calculate profits and income in the pessimistic equilibrium as:

(7) $\Pi(n) = n(ay^{p}(n) - F_1),$

(8) $y^{p}(n) = \dfrac{L - nF_1}{1 - an}.$

For the pessimistic equilibrium to exist, it must be the case that

(9) $ay^{p}(n) - F_2 \leqslant 0,$

or, using (8),

(10) $n \leqslant \dfrac{F_2 - L\cdot a}{(F_2 - F_1)a} = \bar{n}.$

For n's below $\bar{n}$, the pessimistic equilibrium exists; otherwise, profits earned by F_1-firms are high enough to warrant investment by F_2-firms. Conditions (6) and (10) together show that (except when $n=\bar{n}$), the pessimistic and optimistic equilibria cannot coexist. Thus, the equilibrium is unique[3].

Figure 1 illustrates how income is determined in this economy by showing both $y^{o}(n)$ and $y^{p}(n)$ as a function of n. Note that $y^{o}(0) < y^{p}(0)$, since the investing F_2- firms are losing money when n=0, and also that $y^o(1)=y^p(1)$ since there are no F_2- firms in this case. Since $y^o(\bar{n}) = y^p(\bar{n}) = F_2/a$, $n=\bar{n}$ is the case with multiple equilibria in which F_2- firms can play any mixture of investing and abstaining while earning zero profits. For $n>\bar{n}$, the profits of all firms are actually higher when F_2- firms invest; while for $n<\bar{n}$, everyone benefits when F_2- firms stay out of the market and do not lose money. Income in this economy is given by the maximum of the two curves, or

(11) $y(n) = \text{Max }(y^o(n), y^p(n))$.

In this model, y(n) is the highest attainable income for the prices given by the technology of the fringe. Since all these prices are equal to unity, second best (restricted by monopoly pricing) welfare maximization is tantamount to income maximization. Investment decisions in this economy are therefore efficient.

Perfect information about technological opportunities in other sectors enables each firm to make the correct forecast of income and thus to invest efficiently in each state of the world. In addition, firms have a positive external effect of profits of other firms (and hence on welfare) if and only if their own profits are positive. Thus even though firms ignore their own contribution to aggregate demand, they will make the same investment decisions as they would had they taken it into account. Except for the pricing decision, profit maximization promotes social welfare and investment is efficient.

3. The Incomplete Information Model.

If n is not observed, each monopolist has to make his investment decision on the basis of two pieces of information. The first is the commonly shared prior $g(n)$, which by the assumption of rational expectations is the same as the actual density of n's. The second is his observation of his own fixed cost, which he can use to update his beliefs about n. Firms do not have any other information about fixed costs of other firms, and, for now, have no means of obtaining such information. Because fixed costs differ across firms, the posterior distributions $g(n|F)$ will also differ. For example, a firm that draws F_2 will take into account the fact that its draw is more likely when n is low.

Since marginal investment decisions are made by F_2-firms, we next solve for one such firm's problem. The monopolist whose own fixed cost is F_2 revises his beliefs according to Bayes rule, to form a posterior:

$$(12) \qquad g(n|F_2) = \frac{(1-n)g(n)}{1 - \int_0^1 ng(n)dn}.$$

When this firm conjectures incomes $y_2(n)$ in state n, it invests provided

$$(13) \qquad a\int_0^1 y_2(n)g(n|F_2)dn - F_2 > 0.$$

Since profits are linear in income, an F_2-firm only cares whether expected income (according to its posterior) exceeds F_2/a — the income at which it breaks even.

Unlike in the previous section, we can no longer be sure that all equilibria are in pure strategies. Supposing that each F_2-firm invests with probability q, a fraction q of F_2-firms invest in a mixed strategy equilibrium. In this case, income in state n is given by:

$$(14)\quad y(n) = \frac{L - nF_1 - q(1-n)F_2}{1 - an - aq(1-n)}.$$

For q=1, this expression reduces to (4), while for q=0, it reduces to (8). Putting (13) and (14) together, we have an optimistic equilibrium if (13) holds when q=1, a pessimistic equilibrium if (13) fails when q=0, and a mixed strategy equilibrium when expected profits equal F_2/a for some $0<q<1$. An application of the intermediate value theorem shows that if neither an optimistic nor a pessimistic equilibrium exists, then a mixed strategy equilibrium exists. Thus there always exists an equilibrium, although for some constellations of parameter values there are multiple equilibria. In this model it can be shown, however, that if an optimistic equilibrium exists, then it is unique.

Figure 1 can again be used for a diagrammatic exposition. If the optimistic equilibrium is played, then for a given realization of n, the ex post income is equal to $y^o(n)$. Such an equilibrium is sustainable if the expected value of $y^o(n)$, taken with respect to an F_2- firm's beliefs, is above F_2 /a (which is equivalent to the expected value of n being above $\bar{n}$). Similarly, pessimistic equilibria for various realizations of n are points on the $y^p(n)$ curve, and are sustainable if the expectation of $y^p(n)$ is below F_2/a. The locus of incomes in a mixed strategy equilibrium where F_2 - firms invest with probability q lies between $y^o(n)$ and

$y^p(n)$ curves. The mixed strategy equilibrium for that q exists if the expected income is F_2/a. The nonlinearity of the $y^p(n)$ locus is the mathematical reason for the possibility of multiple equilibria.

Although we discuss efficiency at some length is Section 6, a few summary observations are in order here. First, in every equilibrium under uncertainty, there is a potential efficiency loss since firms are unable to condition their investment decisions on the particular realization of n. For example, when n is realized below $\bar{n}$, investment by F_2-firms is inefficient. Note that a social planner operating in the face of the same uncertainty could not avoid such mistakes. In addition, firms' investment choices might even differ from those of a planner with the same information. The reason is that type 2 firms do not fully internalize their effect on aggregate profits in the model of incomplete information. For example, the planner might prefer them not to invest even when they can on average make money. Section 6 discusses welfare economics of incomplete information equilibria in some detail.

The main message of the analysis of equilibrium under uncertainty is that investment decisions in this case are imprecise, and an institution supplying firms with additional information about n is needed. As the next section shows, an informationally efficient stock market can meet precisely this need.

Section 4: The Revealing Stock Market.

In contrast to the previous section, where prices of shares were not observed, suppose now that each monopolist has publicly traded shares. Suppose also that trading by the well-informed agents (insiders) is allowed, so that the agents who know the fixed cost of a firm can trade to incorporate their information into share prices. In this economy, agents can use both the information about their own fixed costs and the observed distribution of share prices to make inferences about n.

Equilibrium share prices will be determined by the no-trade-by-insiders condition. That is, we require that each firm's share price be equal to the conditional expected value of its profits given the aggregate distribution of share prices and the insiders' knowledge of the fixed cost. Specifically, the value of firm i, v_i, is given by:

(15) $v_i = ay_i^e - F^i$,

where F^i is firm i's fixed cost and y_i^e is aggregate income expected by its managers. When the stock market is fully revealing, as it will turn out to be in this section, in equilibrium all agents hold the same expectations about future income. More generally, agents other than firm i's insiders have a different forecast of expected aggregate demand. For example, the manager of firm j who observes a fixed cost that is different from that of firm i will make a different forecast than firm i's insiders. In this case, the no-trade-by-insiders condition is not sufficient to eliminate trade. To justify absence of trade at these prices in a world with risk-neutral agents we appeal to the argument that, in a more general context, rational agents would not trade against insiders except for liquidity reasons. Alternatively, equilibrium in securities markets with diverse expectations can be reestablished by making traders risk-averse. To avoid these difficulties, we always construct examples where no agent wants to trade at equilibrium security prices.

Under these assumptions, there will be two values of share prices in equilibrium: v_1 and v_2. The reason is that all monopolists that draw the same fixed cost will have the same priors, the same private information, and observe the same stock market. (In addition, if there are multiple equilibria, they observe the same sunspot.) Hence they all expect the same income. Generically, $v_1 \neq v_2$, and hence agents will be able to deduce n by counting the number of firms with value v_1. With n known, the economy has a unique perfect foresight equilibrium of the full information case[4]. The only equilibrium stock market prices v_1 and v_2 are equal to the actual profits of the underlying firms.

The stock market in this economy is completely revealing. In equilibrium, private information is not used to forecast income, since the stock market conveys all that agents need to know. (Private information is, however, used to make efficient investment decisions.) The market thus homogenizes expectations about income, that would otherwise differ because of diverse information about fixed costs. Since the equilibrium in the perfect information economy is unique and efficient, the stock market eliminates both multiplicity and welfare loss due to uncertainty. The average income is strictly higher in the presence of a completely revealing stock market. These results, however, do rely on complete revelation.

Many of the empirical regularities concerning the stock market can be rationalized in this model. The stock market is a flawless leading indicator, in the sense that its value stands in a one-to-one relationship with future income. The market contains information that is dispersed throughout the economy and is impossible to obtain from other sources[5]. As a result, the market Granger causes investment, even controlling for other macro variables. Finally, movements in share prices will be positively correlated across firms, since positive profits spillovers change firms' values in the same direction.

Section 5: the Non-revealing Stock Market.

In reality, the price of a share reflects many characteristics of the firm that are known to the insiders but not to the general public. In particular, if share prices reflect both current and future opportunities, the market might have no way of distinguishing present and future profits even when all the insider information is incorporated into prices. The example of this section builds on the present/future confusion in order to generate some predictions about an imperfectly revealing stock market.

Consider a two period economy with the same demand and market structure each period as those in Sections 2-4. Each sector has a monopolist (with an increasing returns to scale technology) each period, but it is a different firm in period 1 than in period 2. No firm earns profits in both periods. The consumer lacks time preference and hence the interest rate is 0. Finally, shares of firms that earn profits in period 1 and of firms that earn profits in period 2 are traded together, without it being known when a particular firm earns its profits. Thus observers have no way of telling whether a given firm is a period 1 or a period 2 monopolist (while insiders have this information), even when they can identify its fixed cost. Information conveyed by share prices is therefore ambiguous.

The price of a firm's shares is again determined by the expectations and private information of its managers[6]. We only consider symmetric equilibria[7], in which managers of firms with the same fixed cost in different periods have the same information (i.e., the same sunspot, among other things) and hence the same expectations about average income in their respective periods. Equilibrium share prices of their firms are therefore the same. For example, the price of shares of a firm with the fixed cost F_1 is v_1 regardless of the period in which profits are obtained.

Again, we focus on the decision of an F_2-firm. Let n_1 and n_2 be the proportions of F_1-firms in periods 1 and 2, respectively, and let $s=(n_1+n_2)/2$ be the fraction of firms with stock price v_1. Conditional on observing s, the posterior density of n_1 by a manager whose F_2 is in period 1, and the posterior density of n_2 by the manager whose F_2 is in period 2 are the same; denote them by $g(n_1|s, F_2)$. Then

$$(16) \qquad g(n_1|F_2,s) = \frac{(1-n_1)g(n_1)g(2s-n_1)}{\int_0^{2s} g(x)g(2s-x)(1-x)dx}.$$

In the equilibrium where fraction q of F_2-firms invest each period, income expected by an F_2 - monopolist in his period is:

$$(17)\ y^e(q,F_2,s) = \int_0^{2s} \frac{L - nF_1 - q(1-n)F_2}{1 - an - a(1-n)q}\, g(n_1|F_2,s)dn_1.$$

The stock market in this model does not resolve all of the payoff-relevant uncertainty. It reveals the fixed cost of each firm, but not the period in which that fixed cost applies. Agents therefore learn $s=n_1+n_2$, but not how many F_1-firms there are in each period. As a result, residual uncertainty remains and the conclusions regarding multiplicity and efficiency of equilibria follow those of Section 3. Section 6 deals with the welfare properties of equilibria with an imperfectly revealing stock market.

The stock market in this case remains a leading indicator and a cause of investment. In addition, since there are multiple sunspot equilibria, the market can move substantially without any change in n. If sunspots change, income shifts dramatically even without significant (if any) technological shocks. Moreover, the market can "make mistakes." For example, if the mood shifts from optimism to pessimism, the market might fall even when s rises. In a sense, the market is moving against the new information, although it accurately reflects future changes in income. Note that although the stock market here propagates changes in the mood (sunspots), it does not cause them.

Section 6: An Illustration of Welfare Issues.

In the previous sections, we have only briefly discussed the welfare economics of equilibria under uncertainty. The model with two types turns out to be too restrictive to easily generate interesting examples, although we have constructed some[8]. Instead, we look at an example with three levels of fixed costs as a way to show how multiple sunspot equilibria arise under uncertainty and how the stock market can lower welfare when firms maximize profits in the presence of uncertainty and aggregate demand spillovers.

In the example, we assume that L=100, α=4, and that there are three levels of fixed costs: F_1=0, F_2=60 and F_3=115. The marginal investors will always be the F_3-firms. The first part of the example addresses the issue of multiplicity. Subsequently we extend the example to introduce the stock market. The two states of the world are summarized in Table 1.

Consider the case of uncertainty. The posterior probability that a firm drawing F_3 attaches to state 1 is (.2•.5)/(.2•.5+.5•.5)=2/7. The respective posterior probability for state 2 is 5/7. Calculations show that this example has three equilibria: optimistic, pessimistic, and mixed strategy with fraction .86 of F_3-firms investing. As an illustration of such calculations, consider the optimistic equilibrium. If all firms invest, income in state 1 equals

y_1 = 4(100-.8•60-.2•115)=116,

while income in state 2 equals

y_2= 4(100 - .5•115)=170.

From the viewpoint of an F_3-firm, then, expected income is

$y^e = (2/7)y_1 + (5/7)y_2 = 154.6$,

and expected profits are $(3/4)\cdot154.6 - 115 = .93$. Since expected profits of an F_3-firm are positive if it expects an optimistic equilibrium, it invests. The optimistic equilibrium therefore exists. Table 2 summarizes the outcomes in each type of equilibrium in both states of the world.

If we look at the average income in this economy, it is equal to 143 in the optimistic and in the mixed strategy equilibrium, and 145 in the pessimistic equilibrium. Thus the consumer would prefer F_3-firms to abstain from investment. At the same time, for obvious reasons, the expected profits of F_3-firms are the highest in the optimistic equilibrium. If profit-maximizing F_3-firms could pick an equilibrium, they would pick the inefficient one. Since they are the only marginal firms, there is nothing to keep them all from investing.

The reasons for the above inefficiency can be understood from Table 2 by considering a move from the pessimistic to the optimistic equilibrium in each state. In state 1, F_3-firms together lose 5.6 as a result of this shift[9]. In addition, they impose a negative externality on F_2-firms, which makes the latter lose 8.4. Overall, the income decline in state 1 from a shift to optimism is 14. In state 2, such a shift results in a profit gain of 6.25 for F_3-firms and of 3.75 for F_1-firms, for a total of 10. While a shift to optimism results in an average gain of $(6.25-5.6)/2=.325$ for F_3-firms, it also results in an average gain of $(3.75-8.4)/2=-2.325$ for other firms, with a total income loss of two units. This is precisely the difference in expected incomes of pessimistic and optimistic equilibrium. Because F_3-firms ignore the externality that their investment imposes on other firms, they invest to the detriment of social welfare.

This example is suggestive of the possibility of reduction in aggregate welfare with an introduction of the stock market. Table 3 presents parameters of an extended example that shows how this can work.

Because state 1 in this example is sufficiently bad, the unique equilibrium without the stock market entails no investment by F_3-firms and average income of 124.27. Once the stock market is introduced, however, agents can tell whether, on the one hand, they are in state 1, or, on the other hand, state 2 or 3. States 2 and 3 yield identical distributions of stock market prices. In state 1, then, agents know the state and F_3-firms abstain from investing, just as they did without the stock market. Income is 103.53. In states 2 and 3, the situation is identical to the first half of this example, where three equilibria existed. If F_3-firms play optimistically and invest, the average income is 143, and the overall average (counting state 1) is 123.27, which is below 124.27. The stock market reduces welfare in one equilibrium (while not raising it in any other), since it prompts inefficient behavior in some states. Furthermore, there is a good reason to believe that this inefficient equilibrium will be played by profit-maximizing F_3-firms. When firms maximize profits in the presence of uncertainty and aggregate demand spillovers, access to additional information gives them the latitude to make profitable, but socially inefficient choices. In this case, the information that the economy is in state 2 or 3 leads F_3-firms to invest in those states contrary to the wishes of the planner.

Although this example is suggestive of possibilities of interesting behavior in aggregate demand spillover models under uncertainty, it does not tell us much about what typically occurs. Are sunspots common in such models? How likely are inefficiencies due to underinvestment as opposed to overinvestment? Does an introduction of an informationally efficient stock market usually raise welfare? Work aimed to answer these questions will be reported in the next draft of this paper.

Section 7: Summary and Conclusions.

The model of this paper replicates several empirical observations about the behavior of the stock market over the cycle, but also suggests an important allocative role for the stock market. The reason for this role is that stock market prices convey information that is relevant to predicting the productive potential of the economy and therefore aggregate demand. Because firms use this information in making their investment decisions, the stock market causes output. Because these investment decisions need not be socially optimal, information transmission through share prices has complex welfare consequences, including the possibility of welfare reduction accompanying the introduction of the stock market.

In summarizing this work, we should note that it does not invalidate the view that the stock market is only a passive predictor of the future. Suppose, for example, that fixed costs in our model were non-discretionary, so that all firms always had to invest. In this case, stock prices have a purely passive role, but because of uniqueness of information they convey falsely appear to Granger cause output. Differentiation between the two views is therefore difficult on the basis of currently available evidence.

We should also stress the limits of our findings on welfare economics of an imperfectly competitive stock market economy. We have only illustrated the possibility that the information conveyed by the stock market can be welfare-reducing when it is used by firms whose investment decisions have spillover effects which they do not internalize. How typical or important such examples are remains to be established.

TABLE 1.

PARAMETERS OF THE EXAMPLE WITH MULTIPLE EQUILIBRIA.

	Fraction of firms with fixed cost $F_1=0$	Fraction of firms with fixed cost $F_2=60$	Fraction of firms with fixed cost $F_3=115$	Probability of the state
State 1	0	.8	.2	.5
State 2	.5	0	.5	.5

TABLE 2.
PROFITS IN THE DIFFERENT STATES IN THE EXAMPLE.

State 1:

TYPES	Profits if F_3-firms invest		Profits if F_3-firms do not invest		Profits if F_3-firms invest with prob. .86	
	of one firm of the given type	of all firms of the given type	of one firm of the given type	of all firms of the given type	of one firm of the given type	of all firms of the given type
F_1	87	0	97.5	0	89.17	0
F_2	27	21.6	37.5	30	29.17	23.33
F_3	-28	-5.6	0	0	-25.83	-4.44
TOTAL		16		30		18.9

State 2:

TYPES	Profits if F_3-firms invest		Profits if F_3-firms do not invest		Profits if F_3-firms invest with prob. .86	
	of one firm of the given type	of all firms of the given type	of one firm of the given type	of all firms of the given type	of one firm of the given type	of all firms of the given type
F_1	127.5	63.75	120	60	125.33	62.66
F_2	67.5	0	60	0	65.33	0
F_3	12.5	6.25	0	0	10.33	4.44
TOTAL		70		60		67.1

TABLE 3.
PARAMETERS OF THE STOCK MARKET EXAMPLE.

STATE	PERIOD 1			PERIOD 2		
	number of F_1-firms	number of F_2-firms	number of F_3-firms	number of F_1-firms	number of F_2-firms	number of F_3-firms
#1 p=.5	0	.2	.8	0	.2	.8
#2 p=.25	0	.8	.2	.5	0	.5
#3 p=.25	.5	0	.5	0	.8	.2

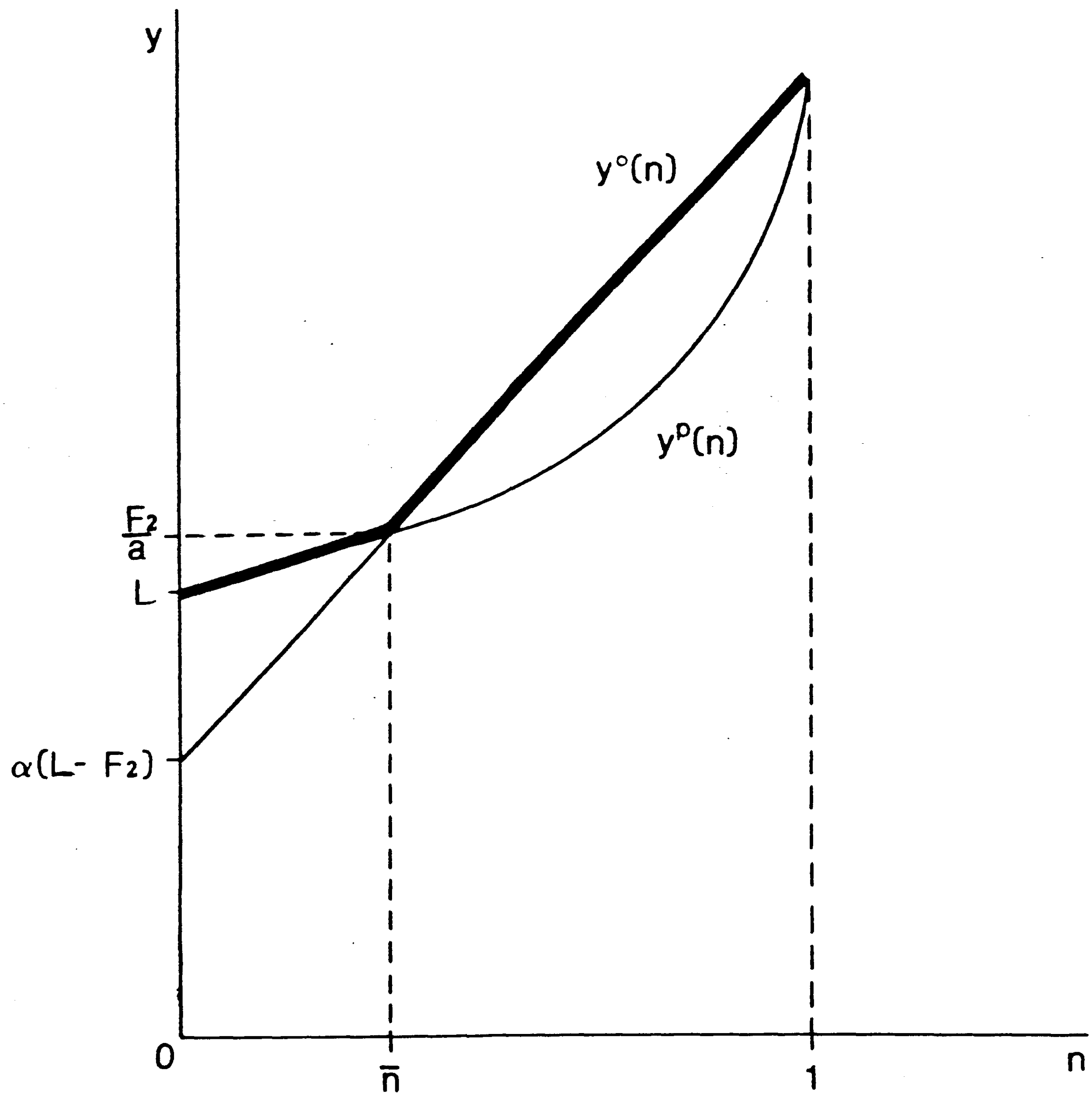


Figure 1

Determination of Equilibrium Income

FOOTNOTES.

1 Models with aggregate demand spillovers that are related to this one have been developed by Weitzman (1982) and Shleifer (1987), among others.

2 When (6) holds with equality, F_2-firms break even when they invest. As will be clear from the subsequent discussion, the economy then has a continuum of equilibria, with different fractions of F_2-firms investing, but with the same income $y=F_2/a$.

3 In the case of perfect information, mixed strategy equilibria do not exist. The uniqueness result generalizes to any finite number of fixed cost levels, but not to a continuum of such. For this precise reason, however, the case with continuum of fixed cost levels must be deemed degenerate.

4 In general, agents use the whole distribution of share prices to infer the distribution of fixed costs and to make an accurate forecast of aggregate income. The revelation result with unidimensional uncertainty was first obtained by Grossman (1976).

5 Surveys are not really a good substitute for the stock market, since agents have no incentive either to participate or to tell the truth. Self-interest of insiders should lead them both to incorporate their information into share prices and to hide it from surveys.

6 In this model, if the no-trade-by insiders condition is satisfied, no other speculator will have a reason to trade either. This feature, however, is special to our completely symmetric example.

7 There may also exist asymmetric equilibria, in which the stock market

is completely revealing since the values of profits of firms with the same fixed cost operating in different periods now differ.

8 In particular, we have constructed an example with a unique pessimistic equilibrium that is Pareto inferior to a non-equilibrium outcome in which F_2 - firms play a mixed strategy. If the state could order firms to play these mixed strategies, it would do so.

9 Considering a change in profits of F_3-firms as a whole and weighing it by prior probabilities is exactly equivalent to considering a change in profits of a representative firm and weighing it by its own posterior beliefs.

REFERENCES.

Doan, Thomas; Robert Litterman; and Christopher A. Sims (1984), Forecasting and Conditional Projection Using Realistic Prior Distributions, Econometric Reviews, pp. 1-100.

Fama, Eugene F. (1981), Stock Returns, Real Activity, Inflation, and Money, American Economic Review, 71, pp. 545-565.

Fischer, Stanley and Robert C. Merton (1984), Macroeconomics and Finance: the Role of the Stock Market, Carnegie-Rochester Conference Series on Public Policy, 21, pp. 57-108.

Grossman, Sanford J. (1976), On the Efficiency of Competitive Markets Where Traders have Diverse Information, Journal of Finance, 31, pp. 573-585.

Grossman, Sanford J. and Laurence Weiss (1982), Heterogeneous Information and the Theory of the Business Cycle, Journal of Political Economy, 90, pp. 699-727.

Lucas, Robert E. Jr. (1972), Expectations and the Neutrality of Money, Journal of Economic Theory, 4, pp. 103-124.

Moore, Geoffrey H. (1983), Business Cycles, Inflation, and Forecasting, Cambridge: Ballinger.

Shleifer, Andrei (1987), Implementation Cycles, Journal of Political Economy, forthcoming, February.

Weitzman, Martin L. (1982), Increasing Returns and the Foundations of Unemployment Theory, Economic Journal, 92, pp. 787-804.

ESSAY THREE

DO DEMAND CURVES FOR STOCKS SLOPE DOWN?

1. Introduction.

Several important propositions in finance rely on the ability of investors to buy and sell any amount of the firm's equity without significantly affecting the price. For example, the home leverage idea behind the Modigliani-Miller theorem [13] and simple cost of capital rules obtain under the maintained assumption of horizontal demand curves for the firm's equity. In addition, most of the common elaborations of the efficient markets hypothesis (such as CAPM or APT) predict horizontal on nearly horizontal demand curves for stocks. In these models, the stock price is an unbiased predictor of underlying value, maintained through the workings of arbitrage. To the extent that stocks have close substitutes, that underlying value is not significantly dependent on supply. Thus the (excess) demand curve for a security is (nearly) horizontal.

Recognizing the importance of the assumption of horizontal demand curves for stocks, financial economists have long been interested in testing it directly. Traditionally, they have done so by examining stock price reactions to buyer and seller initiated large block trades. Negative price reactions to large block sales (and converse for purchases) have been found by Scholes [19], Holthausen, Leftwich and Mayers [7] and Mikkelson and Partch [11]. This evidence, however, is also consistent with the information hypothesis, stating that an offer to buy a large block may signal good news about the stock, thus entailing a price increase. Large block trade studies are therefore inconclusive on the hypothesis that demand curves for stocks slope down (the DS hypothesis).

This paper examines stock inclusions into the S & P 500 Index (hereafter, the Index) to examine the DS hypothesis in the context where information effects probably play no role. Every year since 1966, between 5 and 35 firms have been removed from the Index, usually as the result of takeovers[1]. When S & P takes a stock out of the Index, it simultaneously includes a new firm. Neubert [15] states the following six criteria for inclusion: size, industry classification, capitalization, trading volume/turnover, emerging companies/industries, responsiveness of the movements of stock price to changes in industry affairs. All of these criteria are public information, and none of them is concerned with future performance of the firm.

Subsequent to the announcement of the inclusion, a substantial portion of the firm's shares is bought by index funds, which are funds attempting to mimic the return on the S & P 500 for institutional clients[2]. Though these funds do not necessarily replicate the S & P 500 exactly, and may spread out their buying of a newly included stock over several days, they usually buy what in recent years could have been up to 3% of the newly included firm's equity. Such buying represents an outward shift of the demand curve for the firm's equity, and more importantly, one resulting from demand by buyers whose interest is not prompted by good information. If the demand curve is horizontal, inclusion of a stock into the S & P 500 should not be accompanied by a share price increase. In contrast, if the demand curve slopes down, we should observe a share price increase at the announcement of the inclusion.

Section 2 of this paper presents the description and the results of an event study of stock inclusions into the S & P 500. Section 3 discusses several explanations for the observed results, and attempts to discriminate between these explanations empirically. Section 4 concludes that the DS hypothesis is probably an important part of the observed share price behavior at the inclusion of a stock into the S & P 500.

2: The Event Study.

To perform the event study of stock inclusions into the S & P 500 Index, it is necessary to identify the dates at which the market learns about the inclusion of each stock, the so-called announcement dates (ADs). Revisions of the Index are made effective on weekdays (lately it has always been on Wednesdays) after the market closes. Since September of 1976, S & P has been running an early notification service, whose subscribers are notified about the changes in the composition of the Index within minutes after these changes are made (but again, after the market closes). Thus the day after the inclusion is the appropriate announcement date to examine price changes in the period after September, 1976.

Prior to September 1976, changes in the Index were recorded in a monthly Cumulative Index to Standard and Poor's Outlook (CISPO), which published a complete listing of stocks in the S & P 500 every month. The announcements of changes did not include the actual dates on which they were made; in fact this information was not available then and is not available now for that period. I assume that the announcement date for this period is the day on which subscribers received the CISPO containing the relevant change in the S & P 500. Since CISPO is mailed out ahead of its official publication date, this announcement date actually coincides with the publication date.

The sample includes firms added to the S & P 500 Index between 1966 and 1983. I started with 331 firms. Of those, 34 firms were removed from the sample because CRSP had no data on them (e.g. OTC stocks), 13 firms were removed because their inclusion was perfectly anticipated (e.g. regional telephone companies in 1983 or companies that were already part of S & P 500 and were reincluded after they changed their name subsequent to a merger). Because of the difficulty of identifying the announcement date, I also excluded 17 firms that were included into the Index on June 30, 1976, when the Index underwent a major revision[3]. Finally, I excluded 21 firms in the earlier period because I could not ascertain the announcement dates, and data provided by the S & P Corporation were faulty[4]. After these exclusions, the sample was reduced to 246 firms, or 74% of the initial sample. Share prices were obtained from CRSP files.

To examine share price behavior surrounding inclusion into the S & P 500, I performed a daily event study following Fama, Fisher, Jensen and Roll [5] as implemented by Ruback [18]. Specifically, the market model is applied to descibe the behavior of asset returns, using the value weighted market portfolio from CRSP files. To account for possible risk changes due to inclusion into the S & P 500, stock return equations were estimated separately on the observations before and after the inclusion[5]. Residuals from these equations, called prediction errors, were then averaged across observations for a given day τ relative to the announcement date (AD) to get the Average Prediction Error (APE_τ). The APE_τ measures the mean abnormal performance for a given day relative to the announcement of the inclusion into the S & P 500. The sum of these APE_τ over event days τ_1 through τ_2 yields the Cumulative Average Prediction Error ($CAPE_{\tau_1,\tau_2}$), which measures abnormal performance over an interval of event time. Ruback [18] describes the procedure for obtaining consistent estimates of variances of CAPEs, accounting for possible first order serial correlation in prediction errors that is common to event studies.

The basic results of the event study are presented in the top panel of Table 1. To account for the change in the definition of the announcement date, the results for the periods before and after September, 1976 are presented separately. The results also reveal that prior to September, 1976, there was, on average, no significant price increase on the announcement date. In contrast, since September, 1976, there has been a 2.79 percent AD abnormal return, which is statistically significant at any reasonable confidence level. For over 95% of individual observations, the AD abnormal return is positive in this period. Both statistically and substantively, the inclusion of stocks into the S & P 500 has been accompanied by large abnormal announcement date returns.

In either period there is no evidence of prices starting to rise prior to the announcement date; if anything, the cumulative abnormal returns are negative in the twenty days prior to the AD. In the earlier period, a price increase ahead of the AD would be evidence of the market's learning about the forthcoming inclusion before publication of CISPO and incorporating this information into the price. Absense of such price increases for that period suggests that the choice of the AD is not responsible for the result. For the period after September, 1976, a price increase starting early would be evidence of the market predicting the inclusion of the stock, and revaluing the shares ahead of time. But even on the day prior to the AD, the one day average excess return is not statistically different from 0, confirming that inclusion into the Index is not anticipated.

The last two rows of Table 1 examine share price behavior subsequent to the AD. Though point estimates show share price declines, these declines are not statistically significant (this result holds for individual days after the AD also). In addition, the magnitude of the point estimates of declines is much smaller than that of the AD abnormal return. Thus in the period since September, 1976, prices do not fall significantly in the twenty days after the AD.

To examine the total return from inclusion into the S & P 500, we must look at the cumulative abnormal return starting with the AD. The test of this is a t-test of Cumulative Abnormal Prediction Errors (CAPEs) defined above, starting with the AD. Because these tests determine not only whether stock prices fall after the AD, but also whether the decline offsets the gains on the AD, they are appropriate tests of permanence of the abnormal return from inclusion. The first column of Table 2 presents these tests for the post September, 1976 period. The tests support the hypothesis that positive AD price effects last for at least a month. The cumulative return for the first six and eleven days is significant at very high confidence level, though, as discussed earlier, point estimates indicate some share price declines. As we look further ahead starting from the AD, the standard error of the cumulative return rises. As a result, the twenty one day cumulative return is not statistically significant, although there is no evidence of prices continuing to fall after eleven days judging from point estimates. Even after twenty one days, the cumulative return is significant at the 85% level, and is only a percent below the AD return.

We conclude from these results that, since September, 1976, inclusion of a stock into the S & P 500 Index earned its shareholders a close to 3% announcement date capital gain, most of which has persisted for at least 10 to 20 trading days. The data cannot tell if the duration of this gain is even longer.

3. Analysis of the Results.

In this section, the results of the event study are discussed in greater detail. I first examine additional evidence that may bear on the DS hypothesis, and then evaluate several other possible explanations of the results.

A. Further Tests of the DS Hypothesis.

An important implication of the DS hypothesis is that the share price increase on the announcement date should be positively related to the shift of the demand curve[6]. Index funds have grown dramatically over time. Though precise time series data are not available, calculations based on surveys of index funds by Pensions and Investment Age [3,16] suggest that index funds owned less than .5% of S & P 500 in 1975, 1.4% in 1979 and 3.1% in 1983. The results of the event study suggest that, indeed, the average abnormal AD return cannot be found before 1976, and rose dramatically after that. Unfortunately, this may be due to inaccurate choice of the announcement date in the earlier period (although there is no evidence of abnormal returns prior to publication of CISPO). Alternatively, I split the later sample into the September, 1976 to 1980 subperiod with 44 observations and the 1981 to 1983 subperiod with 58 observations. The last two columns of Table 2 present the results of the event study for the two subperiods. The average abnormal AD return was 2.27% with a t-statistic of 7.63 in the first subperiod, and 3.19% with a t-statistic of 10.1 in the second subperiod. A one-sided t-test of equality of abnormal returns across subperiods rejected at 98% confidence level (with value of 2.2). Consistent with the growth of index funds over the post September, 1976 period, the abnormal announcement date returns have also grown, providing some evidence for the DS hypothesis.

Another interesting result emerging from Table 2 is that although in the earlier subperiods we cannot reject the hypothesis that the abnormal AD return disappears within six days with 95% confidence (we still can reject it for four days), for the later subperiod, even the twenty one day cumulative abnormal return is statistically significantly different from 0. In the later subperiod, not only does the abnormal return remain for a long time, there is some evidence that it increases as time passes. However we look at it, the abnormal returns from stock inclusion into the S & P 500 seem to have grown over time, paralleling the growth of index funds.

An alternative measure of buying by index funds is the excess announcement date volume. Representatives of Vanguard Index Trust and Wells Fargo Index Fund have suggested to me that these funds buy the necessary shares within a few days of the inclusion. To gauge the amount of each stock bought by index funds and self-indexing investors, we look at abnormal daily volume on the announcement date, defined as the difference between the AD volume and the average daily volume in the previous six months, both expressed as a fraction of the number of shares outstanding. We also look at abnormal announcement week (AW) volume defined similarly. While the latter measure of buying by index funds probably includes more of the relevant transactions, it is also much noisier. Abnormal volumes, just like abnormal returns, show <u>no</u> tendency to be positive prior to September, 1976; therefore I restrict the analysis to post September, 1976 data[7]. Since September, 1976, the average abnormal AD volume has been .340%, while the average abnormal AW volume has been 1.012%[8]. Both of these numbers are substantially smaller than the fraction of S & P 500 held by index funds, which may be explained either by slow buying by index funds, or by withdrawal of other investors from the market as index funds are buying. Volume measures are clearly very noisy indicators of index fund buying.

To examine the association between abnormal volumes and abnormal returns, we run a cross-sectional regression of abnormal AD return (RETURN) on abnormal AD volume (ABADVOL):

$$\underset{}{\text{RETURN}} = \underset{(7.55)}{2.41} + \underset{(2.13)}{1.22} \cdot \text{ABADVOL} \qquad \bar{R}^2 = .04,\ N=84.$$

t-statistics in parentheses

To the extent that ABADVOL measures the buying by index funds, the significant positive slope estimate is consistent with downward sloping demand curves for stocks. The true coefficient is probably even larger, since ABADVOL measures abnormal buying with error, and therefore, by a standard errors-in-variables argument, the coefficient is biased toward zero. An alternative test is to regress abnormal returns on the announcement date volume (ADVOL). In this case, usual volume (USVOL) should be included independently into the regression, since we expect that the price effect should be smaller for wider trading stocks. The results are:

$$(1) \quad \text{RETURN} = \underset{(6.73)}{2.77} + \underset{(1.64)}{.982} \cdot \text{ADVOL} - \underset{(-2.51)}{1.59} \cdot \text{USVOL} \qquad \bar{R}^2 = .05,\ N=84.$$

The coefficient on the ADVOL is positive and significant at the 90% level, while the coefficient on usual volume is negative and significant at the 95% level. Both of these results, as well as earlier results based on assets in index funds, are consistent with downward sloping demand curves. Thus the data on "quantities", though imprecise, broadly support the DS hypothesis.

B. The Information Hypothesis.

Every observation that I produced so far can be explained by some version of the information hypothesis, stating that S & P's inclusion of a stock into the Index certifies the quality of the company, and thus entails a price increase. If this hypothesis is supplemented by the notion that S & P responded to growth of index funds by improving the quality of its process of selecting new companies for the Index, the information hypothesis can explain the growth of abnormal AD returns as index funds grew. It can also explain the positive volume/return relationship, if we believe that larger surprizes are associated with greater turnover of the stock. Similarly, one can say that inclusion means more for thinly trading stocks, thus explaining the negative coefficient on usual volume in regression (1).

Although these rationalizations are quite contrived, the argument that inclusion into the S & P 500 certifies quality has some appeal. To make the S & P 500 a convenient way to hold the market, S & P should avoid excessive turnover in its Index. When S & P replaces a stock in the S & P 500, index funds have to change their portfolios and to incur transaction costs along the way. Since they have to pay a higher price for the stock than the previous day's close, they also lose some of the return on the Index (in 1983, this loss may have been as high as ¼%). If holding S & P 500 is costly and inconvenient, index funds may decide to hold some other index. When this happens, S & P loses profits it earns form selling information about its Index.

For these reasons, S & P might be worried about the longevity of firms it includes into the Index. In fact, Neubert [14] expresses S & P's concern about excessive turnover in the S & P's 500. If S & P knows the likelihood of financial distress for different firms, inclusion should be good news about the firm's prospects. S & P's experience with rating bonds makes it plausible that it has the necessary expertise, even though S & P is unlikely to have any truly inside information.

Several considerations suggest that the informational value of the inclusion of a stock into the S & P 500 should not be too great. First, the purpose of the S & P 500 Index is to be a proxy for the market, not a listing of future "winners". Even if some industries are riskier than others, S & P must include firms from these industries in the Index in order to keep it a veritable proxy for the market. Danger of bankruptcy is a second order concern. More importantly, the information communicated by the inclusion must be above and beyond all the other information the market has, and in particular, it must add to the information the market has about S & P's own bond ratings for newly included firms.

The last observation can be used to test a version of the information hypothesis. This test examines the relationship between the abnormal AD return and S & P's own bond ratings of firms included into the S & P 500 Index. Presumably, if S & P rated the bonds of a particular firm as unsafe, inclusion should result in a greater upward revaluation of the shares than inclusion of a firm with a good bond rating. If longevity really is what S & P cares and knows about, the certification value of including a firm whose bonds S & P has already rated to be safe cannot be too great. This hypothesis points to a negative correlation between the abnormal AD return and the quality of bonds.

To test this prediction, I ran a cross-section regression of abnormal AD returns on a dummy equal to 1 if rating is A or better (DUM1, 26 cases out of 84), and a dummy equal to 1 if S & P did not have a rating for the firm's bonds (DUM2, 48 cases)[9]. In the remaining 10 cases the rating is below A. We get:

$$\underset{(3.07)}{\text{RETURN} = 2.34} + \underset{(.397)}{.356 \cdot \text{DUM1}} + \underset{(.837)}{.660 \cdot \text{DUM2}} \qquad \bar{R}^2 = -.02,\ N=84$$

t-statistics in parentheses

There is no statistically significant evidence that stocks without S & P bond ratings earned a higher return. Nor is there any evidence that stocks with high bond ratings earned lower excess returns from the inclusion than did the stocks with low bond ratings. Though this is by no means a rejection of the information hypothesis as a whole, this result sheds doubt on a plausible theory that S & P has special informatiOn about firms' longevities.

C. Other Explanations.

Several alternative explanations of the results in this paper will be briefly discussed below. The first is transactions costs. Scholes [19] noted that if sellers of large blocks of securities must bear costs of rebalancing their portfolios, the buyers who initiate the transaction must compensate the sellers for these costs. As a result, buyer initiated trades should take place at a premium. Although this consideration is undoubtedly important in other contexts, it is of limited use in explaining the findings of this paper (although it might explain some part of them). In the early 80's, the trading costs of institutions have fallen to around 50 basis points per share, some of which is the price impact of trades (as opposed to the brokerage fees or the bid/ask spread). Assuming that the sellers to index funds are also institutions, they should demand at most 1% to sell their shares if demand curves are horizontal[10]. While the transaction costs of institutions have fallen in the early 80's, the price effects we observe have increased to over 3%, which is significantly higher than what compensation for transactions costs ought to be. This suggests that transaction costs are at most a part of the complete story.

Another possible explanation for the observed effects is market segmentation. It states that certain kinds of investors are only interested in stocks that are part of S & P 500, and that inclusion of a stock into the Index invites buying by these investors. Note that this theory explains price increases only if the demand curves by initial holders of the shares of the newly included stock are downward sloping. Its predictions are similar to those of the general DS hypothesis, except that market segmentation predicts additional volume in the aftermath of the inclusion. Because I cannot ascertain at what time the investors interested only in S & P 500 shares buy them, I treat this theory as a special case of the DS hypothesis.

A final possibility is the liquidity hypothesis. On this view, inclusion may be followed by a closer scrutiny of the company by analysts and investors, by a greater institutional interest in the stock, and therefore by an increase in public information about it. As a result, the stock will trade more widely, become more liquid, and the bid-ask spread on the stock will fall[11]. This lowers the required rate of return on the stock and thus leads to a price increase immediately after the inclusion. Consistent with this view, Arbel, Carvell and Strebel [1] document that "neglected" firms earn higher returns than do firms widely held by institutions even after correcting for size.

The liquidity view implies that inclusion into the Index should lead to higher excess returns for less well known stocks. As a test of this implication I calculated separately abnormal returns on stocks that at the time of their inclusion into the S & P 500 either were or were not already part of Fortune 500. The idea is that Fortune 500 stocks are on average larger, relatively better known and more closely scrutinized than other new entries into the Index. I did not find any difference in excess returns for Fortune 500 and non-Fortune 500 stocks. Since membership in Fortune 500 signifies size, while membership in S & P 500 signifies only representativeness, being part of the former should be more important in attracting institutional investors. Furthermore, new entries into the S & P 500 Index are often very large firms that were not initially needed to produce a representative basket. It is hard to imagine that they receive additional attention as a result of the inclusion[12].

4. Conclusion: Other Evidence and Implications.

A variety of other evidence on transactions involving substantial amounts of stock is consistent with the DS hypothesis. Recent analyses of large block trades (Holthausen, Leftwich and Mayers [7]) found price responses lasting for at least several hours, which the authors interpreted to be "permanent." New share issues are commonly found to be accompanied by share price declines (Hess and Frost [6]), although the usual explanation of this effect is that the market learns from a new share issue that the stock is overpriced. In conflict with the latter explanation, Loderer and Zimmerman [10] find that when a firm has several classes of common stock with different voting rights but identical dividend streams, and issues new stock of only one class, there is a negative volume effect on the price of shares of that class, while there is no such effect on the price of shares of other classes. This finding suggests that there is a downward sloping demand curve for individual securities. Parallel to the new issues effect, share buybacks by firms are accompanied by price increases, again in line with the DS hypothesis. The DS hypothesis is also strongly corroborated by tax-loss selling explanations of the January effect, which were recently given new life in the work of Schultz [20] and Rozeff [17]. Finally, though more ambiguously, the DS hypothesis may be part of the reason for large takeover premia (Jensen and Ruback [8].

A plausible reason for downward sloping demand curves for stocks is disagreement among investors over the value of the securities that is not resolved through the observation of price. There is strong direct evidence on the prevalence of such disagreement, and on its importance in security price determination (Cragg and Malkiel [4]). Specifically, the divergence of opinions among analysts turns out to be a good measure of the riskiness of the security and therefore of the required rate of return on it. Miller [12] and Varian [21] explain how such disagreement yields downward sloping demand curves.

If the DS hypothesis is an important feature of financial markets, the empirical relevance of several propositions in corporate finance requires reexamination. For example, the assumptions underlying the Modigliani-Miller theorem are violated, which means, importantly, that tests of sources of deviation from MoMi that rely on its holding absent these deviations may be inappropriate. More generally, the variant of the efficient markets hypothesis that states that the price of a security equals its fundamental value that is agreed upon by all investors is violated if disagreements affect share prices (Black [2]). The importance of these issues for financial economics suggests a clear need for additional theoretical and empirical work.

TABLE 1

AVERAGE ABNORMAL RETURNS SURROUNDING INCLUSION OF STOCKS INTO THE S&P 500 INDEX.

DAYS RELATIVE TO THE ANNOUNCEMENT DATE (AD)	AVERAGE CUMULATIVE PREDICTION ERROR	
	1966 - 1975 (BEFORE THE EARLY WARNING SERVICE) N = 144	SEPT, 1976 - 1983 (AFTER THE EARLY WARNING SERVICE) N = 102
AD - 20 through AD - 1	-2.86 (-2.85)	-1.49 (-1.25)
AD	-.192 (-.918)	2.79 (12.4)
AD + 1 through AD + 10	-.065 (-.091)	-.859 (-1.03)
AD + 11 through AD + 20	1.12 (1.57)	-.154 (-.184)

Notes:

1. t-statistics are included in parentheses
2. details of calculations are provided in the text

TABLE 2

TESTS OF PERSISTENCE OF ABNORMAL RETURNS SUBSEQUENT TO INCLUSION INTO S&P 500.

DAYS OVER WHICH THE ABNORMAL RETURN IS CUMULATED	AVERAGE CUMULATIVE PREDICTION ERROR		
	FULL SAMPLE SEPT,1976-1983	EARLY SUBSAMPLE SEPT,1976-1980	LATER SUBSAMPLE 1981 - 1983
AD	2.79 (12.4)	2.27 (7.63)	3.19 (10.1)
AD through AD + 5	2.22 (3.46)	1.58 (1.88)	2.70 (3.10)
AD through AD + 10	1.94 (2.20)	1.09 (.946)	2.58 (2.17)
AD through AD + 20	1.78 (1.46)	-.807 (-.505)	3.77 (2.28)
AD through AD + 60	1.71 (.819)	-1.05 (-.383)	3.85 (1.36)

Notes:

1. t-statistics in parentheses
2. details of calculation of abnormal returns are provided in the text

FOOTNOTES.

1 In 1984, which is a high turnover year, of the 30 companies dropped, one was removed due to bankruptcy, one due to liquidation, one due to financial insolvency, one because it became unrepresentative of any S & P industry group, and the rest because of mergers, acquisitions or leveraged buyouts.

2 In addition to index funds, there are several pension funds, such as CREF, that do their own indexing, and whose holdings or time of buying could not be readily documented. Ring [16] has suggested that self-indexing pension funds in 1984 may have had as much as 50 billion dollars linked to various markets indices, primarily S & P 500. Mrs. Ring also told me that the number may be as low as 20 billion. In earlier years, the numbers are not known at all.

3 CISPO published a story about the revision of the Index on July 7, but listed the newly included stocks only on July 14. Since secrecy was not a big issue then, people working at S & P suspect that anyone who wanted to know about these changes could find out before July 14.

4. The data I received contained no inclusion dates for the period prior to January 1976, but only the dates on which changes in the Index were specifically noted. These written announcements give inappropriate announcement dates, since CISPO often printed revised lists of the S & P 500 before the change was actually brought to the readers' attention. Using the written announcements to define announcement dates had no impact on results for the early period.

[5] The results were not materially different when returns were not corrected for market movements. Similarly, combining the before and after estimation pe-riods did not make much difference. These results are, therefore, not presented.

[6] One question arising in this context is what the relevant shift of the demand curve is. The discussion that follows presumes that the relevant shift is current needs of index funds. Alternatively, one might imagine that all the future needs of index funds, arising from future cash inflows, are relevant for this shift. In that case, current assets in index funds might only be a very crude measure of future demand, and relying on current assets or even current volume as proxies for shifts in demand curves is inappropriate.

[7] Including these observations into the regressions below will only strengthen the result, but for spurious reasons.

[8] The volume data were obtained from Data Resources Incorporated and are available for only a subsample of observations.

[9] Surprisingly, only 42% of firms had S&P-rated bonds at the time they were included into S & P 500. This may be because the majority of bonds are privately placed and hence are not rated, because S & P does not rate all publicly traded bonds, or because many newly included firms do not have bonds at all.

[10] As long as there are some institutional holders of the security, the marginal investor in the horizontal demand curves story can always be taken to be an institution. This observation addresses one objection

to my results, namely that sellers to index funds must be compensated for the necessity to realize their capital gains immediately. For institutions, this is not an issue. If the sellers to index funds are Exchange members (who indeed own significant amounts of shares), the transaction costs premium should be even smaller.

[11] Kyle [9] and Vishny [22] present models in which better information about an asset makes the market in it thicker and lowers the bid-ask spread. This increase in liquidity might lower the required return.

[12] Many very large firms are not in the S & P 500, as they are not needed for the representative basket, while many small firms are. Newly included firms are often much better followed by analysts than many firms already in the Index.

REFERENCES.

[1] Arbel, Avner; Steven Carvell; and Paul Strebel. Giraffes, Institutions, and Neglected Firms. Financial Analysts Journal (1983), May-June, 57-63.

[2] Black, Fischer. Noise. Journal of Finance (1986), forthcoming, July.

[3] Capozzoli, Cathy. Index Assets up to $9 billion as of May 1980. Pensions and Investment Age (1980), July 7.

[4] Cragg, John and Burton Malkiel. Expectations and the Structure of Share Prices. Chicago: the University of Chicago Press (1982).

[5] Fama, Eugene; Lawrence Fisher; Michael Jensen; and Richard Roll. The Adjustment of Stock Prices to New Information. International Economic Review 10 (1969), February, 1-21.

[6] Hess, A. and P.A. Frost. Tests for Price Effects of New Issues of Seasoned Securities. Journal of Finance 36 (1982), 11-25.

[7] Holthausen, Robert; Richard Leftwich; and David Mayers. The Effect of Large Block Trades: a Cross-Sectional Analysis, Mimeo (1984).

[8] Jensen, Michael and Richard Ruback. The Market for Corporate Control: the Scientific Evidence. Journal of Financial Economics 11 (1983), 5-50.

[9] Kyle, Albert S. Market Structure, Information, Futures Prices and Price Formation. In Storey, G.G. ed., International Agricultural Trade, Boulder: Westview Press (1983).

[10] Loderer, Claudio and Heinz Zimmerman. Rights Issues in Switzerland:

Some Findings To Consider in the Debate over Financing Decisions. Mimeo (1985).

[11] Mikkelson, Wayne H. and M. Megan Partch. Stock Price Effects on Secondary Distributions. Journal of Financial Economics 14 (1985), 165-194.

[12] Miller, Edward. Risk, Uncertainty and Divergence of Opinion. Journal of Finance 32 (1977), 1151-1168.

[13] Modigliani, Franco and Merton Miller. The Cost of Capital, Corporation Finance, and the Theory of Investment. American Economic Review 48 (1958), 261-297.

[14] Neubert, Albert S. Frequently Asked Questions About S&P 500 Indexes. In Chicago Mercantile Exchange, Market Perspectives 2, July/August (1984).

[15] Neubert, Albert S. The Ins and Outs of the S&P 500. In Chicago Mercantile Exchange, Market Perspectives 3, February (1985).

[16] Ring, Trudy. Index Funds Show Significant Growth. Pensions and Investment Age. April 1 (1985).

[17] Rozeff, Michael. The Tax-Loss Selling Hypothesis: New Evidence from Share Shifts. Working Paper Series No. 85-16, University of Iowa (1985).

[18] Ruback, Richard. The Effect of Discretionary Price Control Decisions on Equity Values. Journal of Financial Economics 10 (1982), 83-105.

[19] Scholes, Myron. The Market for Securities: Substitution vs. Price

Pressure and the Effect of Information on Share Price. Journal of Business 45 (1972), 179-211.

[20] Schultz, Paul. Personal Income Taxes and the January Effect: Small Firm Stock Returns before the War Revenue Act of 1917: A note. Journal of Finance 40 (1985), 333-344.

[21] Varian, Hal. Differences of Opinion and the Volume of Trade. Michigan Discussion Paper C-67 (1985).

[22] Vishny, Robert W. Market Structure in Speculation and Brokerage, Mimeo (1985).