working paper
department
of economics

ASYMMETRIC BUSINESS CYCLES:
THEORY AND TIME-SERIES EVIDENCE

Daron Acemoglu
Andrew Scott

95-24 Aug. 1995

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

ASYMMETRIC BUSINESS CYCLES: THEORY AND TIME-SERIES EVIDENCE

Daron Acemoglu
Andrew Scott

95-24 **Aug. 1995**

95-24

ASYMMETRIC BUSINESS CYCLES:

THEORY AND TIME-SERIES EVIDENCE

Daron Acemoglu
Department of Economics
Massachusetts Institute of Technology

and

Andrew Scott
Institute of Economics and Statistics and All Souls College,
Oxford University

August 1995

JEL Classification: E32
Keywords: Asymmetries, Cyclical Indicator, Intertemporal Increasing Returns Regime Shifts, Temporal Agglomeration, Unobserved Components.

Abstract

We offer a theory of economic fluctuations based on intertemporal increasing returns: agents who have been active in the past face lower costs of action today. This specification explains the observed persistence in individual and aggregate output fluctuations even in the presence of i.i.d shocks, essentially because individuals respond to the same shock differently depending on their recent past experience. A feature of our model is that output growth follows an unobserved components process with special emphasis on an underlying cyclical indicator. The exact process for output, the sharpness of turning points and the degree of asymmetry are determined by the form that heterogeneity takes. We suggest a general formulation for models with latent cyclical variables which, under certain assumptions, reduces to a number of state space models that have been successfully used to model U.S. GNP. Using our general formulation we find that allowing for richer heterogeneity enables us to obtain a better fit to the data and also that U.S. business cycles are asymmetric, with this asymmetry manifesting itself as steep declines into recession. We estimate a strongly persistent cyclical component which is not well approximated by discrete regime shifts nor by linear models. Our estimates of the relative size of intertemporal returns needed to explain U.S. GNP vary but on the whole are not implausibly large.

We owe a special debt to James Hamilton whose insightful comments improved an earlier version of this paper. We also thank Philippe Aghion, Charlie Bean, Ricardo Caballero, Paul David, Steve Durlauf, Vittorio Grilli, Christopher Harris, Rebecca Henderson, Per Krusell, Jim Malcomson, John Moore, Chris Pissarides, Danny Quah, Kevin Roberts, Donald Robertson, various seminar participants and two anonymous referees for useful comments. Remaining errors are the authors' responsibility.

1. Introduction

Aggregate economic fluctuations are characterized by successive periods of high growth followed by consecutive periods of low activity. The transitions between these periods of high and low growth are often marked by sharp turning points and considerable evidence suggests that at these moments the stochastic properties of the economy change and display asymmetries, see *inter alia*, Neftci (1984), Diebold and Rudebusch (1989), Hamilton (1989) and Acemoglu and Scott (1994). The importance of tracking these movements in the business cycle is reflected in the considerable attention paid to a variety of coincident and leading indicators (e.g. Stock and Watson (1989), and the papers in Lahiri and Moore (1991)).

A natural way to model temporal agglomeration[1] and asymmetries in economic fluctuations is to assume non-convexities, such as discrete choice or fixed costs at the individual level, because such non-convexities imply that individuals concentrate their activity in a particular period. It is this implication which has been analyzed with considerable success in the (S,s) literature. However, while the presence of fixed costs can account for the discreetness of economic turning points, it does not naturally lead to persistence - once an individual undertakes an action they are less likely to do so in the near future. The key reason is that although the presence of fixed costs leads to increasing returns, these are *intratemporal*; the full extent of economies of scale arising from fixed costs can be exploited within a period[2]. As a consequence persistence in aggregate fluctuations relies on aggregation across heterogeneous agents: either more agents investing in the past increases the profitability of investment for others (e.g. Durlauf (1991) and (1993)) or aggregate shocks affect agents differently, leading to a smoothed response over time (e.g. Caballero and Engel (1991)).

This paper emphasizes an alternative explanation of persistent individual and aggregate output fluctuations by suggesting that fixed costs introduce *intertemporal increasing returns,* that is, the returns from an activity this period are higher if the activity occurred last period and as a consequence an agent who was active in the recent past is more likely to be active now. In the presence of such intertemporal linkages temporal agglomeration arises from two sources; active agents maintaining a high/low activity level in successive periods because of intertemporal increasing returns and agents acting together, either in response to a common shock or because of strategic complementarities. We show how a model of the former explains a number of empirical features of business cycle fluctuations and also offer a framework which enables an economic interpretation of unobserved component models of U.S. output which emphasize the underlying cyclical indicator.

The relevance of intertemporal increasing returns depends on (i) whether individual behavior is persistent and (ii) whether firm's technology and costs contain important intertemporal aspects. The

---

[1] We use this term, following Hall (1991), for the bunching of high economic activity over time.

[2] To overcome this problem empirical implementations of (S,s) models sometimes include time-to-build considerations or decreasing returns at high levels of investment, e.g. Caballero and Engel (1994).

answer to these questions will vary depending upon the type of activity under consideration. For instance, in the case of radical changes in the capital stock, e.g. Cooper and Haltiwanger's (1993) automobile retooling, intertemporal effects are unlikely to be important, and this is supported by the lack of persistence in these cases. However, Section 2 surveys micro evidence that firm level investment is persistent and findings from technology studies, the management science literature and organizational theory which supports the notion that many important discrete decisions (e.g. investment in new technology, product development, innovation, maintenance) exhibit some degree of intertemporal increasing returns. These empirical findings motivate the theoretical model of Section 3 in which a firm has to decide each period whether to undertake both maintenance and investment. Maintenance is modelled as having two effects: (i) increasing the productivity of existing technologies and (ii) facilitating the adoption of new innovations. The interaction of these two roles leads to intertemporal increasing returns: firms find it profitable to maintain the newly adopted technologies and this in turn reduces the costs of adopting future innovations. As a result, investment costs are lower when the firm has invested last period, and a natural asymmetry is introduced in individual behavior; in response to a range of shocks, the agents will find it profitable to invest only if they have invested in the recent past.

Section 4 examines the aggregate economic fluctuations implied by individual level intertemporal increasing returns, and illustrates how our model accounts for the main features of temporal agglomeration and asymmetric fluctuations: the significance and persistence of a cyclical component in output fluctuations, and the importance of turning points and business cycle asymmetries. An important feature of our model is its tractability. Not only can we characterize how the persistence of business cycles and the sharpness of turning points are determined but we also make direct contact with a number of econometric studies which rely upon an unobserved component model of output growth, e.g. Harvey (1985), Watson (1986), Clark (1987). This generality enables us to nest a variety of different types of business cycle asymmetries. And this theme is pursued in Section 5 where, in the context of our model, we examine the circumstances under which turning points between booms and recessions are most marked, so that output growth approximates a shifting regime process as in the discrete state Markov model popularized by Hamilton (1989). In all these exercises, we find that the form of heterogeneity (the distribution of idiosyncratic shocks) is the main determinant of the time-series properties, and also of the asymmetric nature of fluctuations.

Section 6 takes our model to data. First, we look at the connection between our theoretical model and a number of estimated regime shift models to infer the size of intertemporal increasing returns necessary to explain U.S. business cycles. However, the real advantage of our formulation is its ability to nest different unobserved component models with different specifications regarding the form of heterogeneity. We thus estimate two versions of our general model on U.S. data. This exercise yields a number of interesting results. First, we find that a version of our general model with normally distributed idiosyncratic shocks does considerably better than the version with uniformly distributed shocks, but this latter is precisely what previous empirical work has used. Therefore, our model offers an alternative time-series representation grounded in economic theory that outperforms a number of popular models. Moreover, the difference between the versions can be heuristically thought to be the importance of *asymmetry* terms in the version with normal shocks. Thus, our investigation also establishes that U.S. business cycles are asymmetric and illustrates that these asymmetries manifest themselves by sharp downswings in economic activity which cannot be captured by linear models.

Since our formulation is derived from an underlying economic model, we also use the estimation results to make inferences about the nature of the aggregate economy. For instance, we find that the cyclical dynamics of the U.S. GNP cannot be characterized as discrete shifts between different regimes, which, in our theoretical model implies that the idiosyncratic uncertainty is playing an important role in the propagation of economic shocks and this is also confirmed by the fact that we estimate the variance of idiosyncratic shocks to be larger than that of the aggregate shock. We also again find that the size of non-convexity implied by the estimation result is modest and that overall the model provides a good fit to the data.

2. Individual Persistence and Intertemporal Increasing Returns

Our interest in temporal agglomeration focuses attention on problems in which fixed costs are important and so agents have to make a discrete choice about whether or not to perform a certain activity. A crucial feature of our analysis will be the investigation of how an agent's current *qualitative* choice (i.e. whether or not to perform a certain activity) influences the same decision tomorrow. In the standard case, an activity that requires a fixed cost will be bunched within a period of time, essentially because fixed costs imply the existence of intratemporal increasing returns to scale. This observation lies at the heart of (S,s) models (e.g. Scarf (1959)) and implies that a brief period of activity is followed by periods of inactivity at the individual level. However, while there is considerable evidence supporting (S,s) models (e.g. Bertola and Caballero (1990), Caballero and Engel (1994), Doms and Dunne (1994), Cooper, Haltiwanger and Power (1994)), there is also substantial evidence that firm level investment decisions are characterized by significant persistence. For instance, using UK firm level data, Bond and Meghir (1994) find significant autoregressive effects in investment behavior. They estimate equations for the investment-capital ratio and find $I_t/K_t = \alpha + 0.856\, I_{t-1}/K_{t-1}(1-0.122\, I_{t-1}/K_{t-1})+\beta \mathbf{Z}_t + e_t$ where $\mathbf{Z}_t$ is a vector of firm relevant variables and e_t is a white noise disturbance. Evaluating the quadratic term at its sample mean yields an AR(1) coefficient of around 0.75. Bond et al (1994) estimate AR(1) and AR(2) models using Belgian, French, German and UK firm panel data finding strongly significant investment lags, with the sum of the autoregressive coefficients around 0.3. Even the evidence in Doms and Dunne (1994) reveals that there is significant investment occurring in most years of the sample for each firm and that concentrated investment bursts are spread over several years. These findings suggest that while intratemporal economies of scale (lumpiness) are important there may also be intertemporal linkages which can profitably be exploited. This raises the question, what form of investment produces these intertemporal linkages?

The most obvious form of intertemporal economies of scale is learning-by-doing, whereby past experiences increase productivity or reduce costs. More explicitly, consider the case where incorporating new knowledge is a slow and costly process, limiting the degree to which productive investments can be undertaken within a period. However, the more familiar an individual is with the most recent technology vintages, the cheaper it is to adopt the latest version. In contrast an individual not using the most recent vintage faces compatability problems when dealing with the frontier technology. The result will be that only limited innovative moves are taken within each period, with cost incentives making forward steps more likely to come from active agents. The idea that intertemporal linkages might lead agents to take a sequence of small investment steps is given empirical support in studies of technological innovation. In this literature a distinction is drawn between incremental improvements of an existing technology and radical improvements which destroy the existing technology and start an alternative production or organizational method, e.g. Tushman and Anderson's (1986) *"incremental change"* versus *"technological discontinuities"*. There is widespread agreement that the less spectacular incremental changes account for most of the productivity gains (see Abernathy (1980), Myers and Marquis (1969) and Tushman and Anderson (1986)).

More importantly for our paper, there is also a consensus that incremental innovations are more likely to come from firms who have been active at the earlier stages of product development. In Freeman's (1980, p.168) words *"the advance of scientific research is constantly throwing up new discoveries and opening up new technical possibilities, a firm which is able to monitor this advancing frontier by one means or another may be one of the first to realize a new possibility"*. Arrow (1974) and Nelson and Winter (1982) also emphasize the advantages possessed by incumbent innovators in being able to further cope with incremental changes. For instance, Nelson and Winter stress the importance of engineers understanding new technologies, suggesting that firms who innovate have an advantage in doing so again. Abernathy (1980, p.70), using evidence from industries diverse as automobiles, airlines, semi-conductor and light bulb manufacturing, notes that *"Each of the major companies seems to have made more frequent contributions in a particular area"* suggesting that previous innovations in a field facilitate future innovations. One possible explanation of these findings are fixed effects: some firms may simply be good at innovating in certain areas. However, the industry wide work of Hirsch (1952), Lieberman (1984) and Bahk and Gort (1993) suggests that more than just individual fixed effects is operating. In the remainder of this paper we shall focus on this form of investment and its implications for aggregate fluctuations.

3. Individual Behavior in the Presence of Intertemporal Increasing Returns

(i) *The Environment*

We assume that agents (or firms) are risk-neutral and forward looking, and seek to maximize their return (either profit or utility). The key decision is whether or not to invest in order to benefit from technological progress. More explicitly, each period a new technology becomes available. This technology has a stochastic productivity that is revealed at the beginning of the period. The agent decides whether to adopt this technology or not, for instance whether to install a new software. If the technology is adopted, the productivity of the agent increases permanently, starting from the current period[3]. To obtain the highest return from this innovation, its compatibility with existing technologies needs to be monitored. In particular, at the end of the installation period there is the option to gain additional productivity from the innovation by maintenance[4] (e.g. fine-tuning). Under these assumptions the firm's output is

$$y_t = y_{t-1} - \alpha_0 + (\alpha_1 + u_t)s_t - \alpha_2 s_t(1 - m_t) \tag{1}$$

where α_0, α_1 and α_2 are positive parameters and s_t and m_t are binary decision variables that equal 1 if investment (s_t) and maintenance (m_t) are undertaken this period, and equal zero otherwise. If the new technology is adopted (s_t=1), the productivity of the firm is higher permanently. When there is no maintenance effort at the end of the period (m_t=0), this increase in productivity is not as large as it could be (α_1-α_2+u_t instead of α_1+u_t). Deterministic depreciation is denoted by α_0 (our results are unchanged if depreciation is avoided when m_t=1). We assume that u_t is a serially uncorrelated random shock to the productivity of investment with distribution function denoted by F(.). Maintenance costs are assumed to be equal to a positive constant, γ_0 (i.e. $C_t^m=\gamma_0 m_t$). A distinctive feature of (1) is that as in (S,s) models, our focus is on decisions that have a discrete nature, e.g. whether to adopt the recent technology vintage or whether to maintain existing machines. The discrete nature of the choice explains why the productivity shock in (1) is multiplicative with the investment decision; productivity shocks are associated with one vintage of technology and therefore only affect output when the new technology is adopted.

In this model, maintenance also has an additional role other than ensuring the maximum performance of the latest innovation. More specifically, we assume investment costs are given by:

$$C_t = (\gamma_1 - \gamma_2 m_{t-1})s_t \tag{2}$$

where both γ_1 and γ_2 are positive so that when equipment is maintained the firm's investment costs next period are lower. Here γ_2 is the extra investment cost incurred by a plant which did not maintain

---

[3] Time-to-build considerations can easily be incorporated in the return function and only serve to change the timing of returns.

[4] See Pennings and Buitendam (1987) for the importance of maintenance type activities.

last period. In terms of our computing example, all existing bugs in the system need to be removed to ensure the optimal implementation of the new software, implying that next period the instalment of another new program is less costly. In contrast, if the computer is not maintained, technical assistance is necessary both at the beginning of the period for installation (to remove bugs) and at the end of the period (for fine tuning to increase productivity if so decided).

Each period the firm decides whether to invest (s_t=0 or 1) and also whether to maintain (m_t=0 or 1). Denoting the discount factor of the representative firm by β and the per period return by r(.), we have:

$$r(s_{t+j},m_{t+j};m_{t+j-1},y_{t+j-1},u_{t+j})= \\ y_{t+j-1}-\alpha_0+(\alpha_1+u_{t+j})s_{t+j}-\alpha_2 s_{t+j}(1-m_{t+j})-\gamma_0 m_{t+j}-(\gamma_1-\gamma_2 m_{t+j-1})s_{t+j} \tag{3}$$

and the maximization problem of the firm at time t is:

$$\max_{\{s_{t+j}\},\{m_{t+j}\}} E_t\{\Sigma_{j=0}^{\infty}\beta^j r(s_{t+j},m_{t+j};m_{t+j-1},y_{t+j-1},u_{t+j})\} \tag{4}$$

subject to (1) and taking y_{t-1}, u_t, m_{t-1} as given. In period t+j, the state variables are m_{t+j-1}, u_{t+j}, y_{t+j-1} and the choice variables are s_{t+j} and m_{t+j}.

Whether the firm invests in any period depends on u_t. In contrast, the return to maintenance only depends on whether or not investment occurs. If no current investment is undertaken, the only benefit of maintenance is the cost reduction that it brings in the following period if investment then occurs. In contrast, if there is current investment, future productivity also increases by α_2. As a result there are three possibilities regarding maintenance: (i) always maintain (ii) never maintain, (iii) maintain only when there is investment. Because we wish to make maintenance a decision of the firm rather than a fixed characteristic, we concentrate on (iii). To this end we assume[5] :

Assumption A:

$$\beta\gamma_2<\gamma_0<\frac{\alpha_2}{1-\beta} \tag{5}$$

The first inequality can be understood by noting that $\beta\gamma_2$ is the maximum benefit from maintenance in the absence of current investment: if there is investment next period costs are lower by γ_2, otherwise there is no benefit. Therefore, the first part of the inequality implies maintenance is not

---

[5] Assumption A is stronger than we require but simpler to understand than the necessary condition,

$\beta\gamma_2\int_{\omega_0-\omega_1}^{\infty}dF(u)<\gamma_0<\frac{\alpha_2}{1-\beta}+\beta\gamma_2\int_{\omega_0-\omega_1}^{\infty}dF(u)$, where ω_0 and ω_1 are constants defined below.

worthwhile just to obtain future cost savings. In the presence of current investment, the minimum gain from investment is the present value of the productivity increase due to maintenance, $\alpha_2/(1-\beta)$. Consequently the second part of the inequality states that even without future cost savings, it is profitable to maintain if there is current investment. It follows that when (5) holds, we can limit our attention to the case where the firm only maintains when it invests, thus $m_t=s_t$, and the per period return simplifies to:

$$r(s_{t+j};s_{t+j-1},y_{t+j},u_{t+j})=y_{t+j}-\alpha_0+(\alpha_1+u_{t+j})s_{t+j}-(\delta_0-\delta_1 s_{t+j-1})s_{t+j} \tag{6}$$

where $\delta_0=\gamma_0+\gamma_1$ and $\delta_1=\gamma_2$. Assumption A therefore enables us to write our problem in a way which focuses on the intertemporal increasing returns arising from the interactive term $\delta_1 s_{t+j}s_{t+j-1}$. More generally, (6) can be interpreted as the period return for a firm which benefits from lower current investment costs, if it invested last period. Under this interpretation our model captures the idea that firms who are used to adapting to a changing environment face lower costs of further changes.

(ii) Optimal Decision Rules

Using (6) the maximization problem at time t can be written in a dynamic programming form with the value function V(.) satisfying:

$$V(y_{t-1},s_{t-1},u_t)=\sup_{s_t}\ \{r(s_t;y_{t-1},s_{t-1},u_t)+\beta E_t V(y_{t-1}-\alpha_0+(\alpha_1+v_t)s_t,s_t,u_{t+1})\} \tag{7}$$

Solving the agent's optimization problem we obtain (proof in the appendix):

Proposition 1: The value function

$$\begin{cases} s_t=1 \quad and \quad V(y_{t-1},s_{t-1},u_t)=\phi_0+\phi_y y_{t-1}+\phi_s s_{t-1}+\phi_u u_t \\ \qquad\qquad if\ u_t\geq\omega_0-\omega_1 s_{t-1} \\ s_t=0 \quad and \quad V(y_{t-1},s_{t-1},u_t)=\phi_0'+\phi_y y_{t-1} \\ \qquad\qquad if\ \ u_t<\omega_0-\omega_1 s_{t-1} \end{cases} \tag{8}$$

is the unique function that satisfies (7) (ϕ's denote constant coefficients).

To understand the dichotomous nature of the value function, consider the case where the agent does not invest ($s_t=0$). The disturbance u_t is irrelevant to future profits and with no investment costs, it does not matter whether the firm invested/maintained last period and so the value function depends only upon y_{t-1}. However, if the firm invests the value function depends linearly upon y_{t-1}, s_{t-1} and u_t. The optimal choice of s_t depends on whether the investment shock, u_t, is above a certain critical value. Shocks below this value lead to insufficiently high productivity to be embodied in the firm's technology. The critical value depends on s_{t-1} due to the intertemporal non-separability in the cost function. To understand what determines these critical values note that the firm invests this period iff the difference between the left hand side of (7) evaluated at $s_t=1$ and 0 is positive. Using the Appendix and (8), the agent invests iff

$$[1-\beta\int_{\omega_0-\omega_1}^{\omega_0} dF(u_{t+1})]^{-1}\times[\frac{1}{1-\beta}\alpha_1-\delta_0+\beta\delta_1\int_{\omega_0-\omega_1}^{\infty} dF(u_{t+1})+\frac{\beta}{1-\beta}\int_{\omega_0-\omega_1}^{\omega_0} u_{t+1}dF(u_{t+1})]+\delta_1 s_{t-1}+\frac{1}{1-\beta}u_t\geq 0 \qquad (9)$$

It is this inequality which implicitly defines the coefficients ω_0 and ω_1 in (8) (see (A3) in the Appendix). The intuition behind this expression is a good way of illustrating the main features of our model. The firm is comparing the strategy $s_t=1$ with $s_t=0$[6]. If $s_t=1$ production increases by α_1+u_t for all periods compared to $s_t=0$, which has a net present value (including costs) of

$$(1-\beta)^{-1}(\alpha_1+u_t)-(\delta_0-\delta_1 s_{t-1}) \qquad (10)$$

Any further benefits from choosing $s_t=1$ depend upon future values of u_t. There are three possible cases:

(i) If $u_{t+1}\in(-\infty,\omega_0-\omega_1)$ the agent will not invest regardless of s_t and there are no consequences beyond (10).

(ii) If $u_{t+1}\in[\omega_0-\omega_1,\omega_0)$ the household's t+1 investment decision depends upon s_t. The shock is only favorable enough for investment if the household benefits from cost savings arising from past investment i.e. $s_{t+1}=1$ if $s_t=1$ and $s_{t+1}=0$ otherwise. In this case, investing *today* means a difference in expected discounted value next period of

$$\beta\int_{\omega^0-\omega^1}^{\omega^0} dF(u_{t+1})\times\{(1-\beta)^{-1}(\alpha_0+\int_{\omega_0-\omega_1}^{\omega_0} u_{t+1}dF(u_{t+1}))-(\delta_0-\delta_1)\} \qquad (11)$$

where the first integral represents the probability that $u_{t+1}\in[\omega_0-\omega_1,\omega_0)$ and the second is the expected value of u_{t+1} conditional on $u_{t+1}\in[\omega_0-\omega_1,\omega_0)$. If both u_{t+1} and u_{t+2} fall in the region $[\omega_0-\omega_1,\omega_0)$ this same additional benefit accrues in t+2, suitably discounted. In other words, s_{t+2} only equals 1 when $s_{t+1}=1$, which in turn will only be the case when $s_t=1$. As $\{u_t\}$ is assumed an i.i.d sequence this additional benefit at t+2 is (11) multiplied by $\beta\text{Prob}(\omega_0-\omega_1<u_{t+1}\leq\omega_0)$. A similar logic holds for all future periods and summing over time yields

$$\{1-\beta\int_{\omega_0-\omega_1}^{\omega_0} dF(u_{t+1})\}^{-1}\times[\{\beta\int_{\omega_0-\omega_1}^{\omega_0} dF(u_{t+1})\}\times\{\frac{\alpha_1}{1-\beta}+\delta_1-\delta_0\}+\frac{\beta}{1-\beta}\int_{\omega_0-\omega_1}^{\omega_0} u_{t+1}dF(u_{t+1})] \qquad (12)$$

---

[6] Our results are unchanged if m_t and s_t lie in the continuum [0,1] rather than take discrete values. In this case, if (9) holds as a strict inequality agents choose one corner, $s_t=1$ and $m_t=1$; if (9) is strictly negative, $s_t=m_t=0$. If (9) holds as an equality agents are indifferent between any choice that has $m_t=s_t$. If in this case we impose that the agent chooses $s_t=m_t=1$ our results hold exactly.

Equation (12) is the net expected value of all future investment projects the firm benefits from only if it invests today. If the firm does not invest at time t, and future shocks fall in the region $[\omega_0-\omega_1,\omega_0)$, the firm will not invest in the future. However, if the firm invests today the same sequence of shocks leads the firm to invest and reap the benefits given by (12). Thus there is an important *asymmetry in the way firms respond to investment shocks* in high and low activity states, and as a consequence the marginal propensity to invest varies between these states. This state dependence relies entirely on $\omega_1>0$, which from (A3) in the appendix is equivalent to $\delta_1>0$, or in other words, the presence of intertemporal increasing returns.

(iii) Finally, if $u_{t+1}\in[\omega_0,\infty)$, the shock is so favorable agents invest regardless of whether they benefit from lower costs. However, while investment decisions are the same irrespective of s_{t-1}, costs are not. If $s_t = 1$ the cost of choosing $s_{t+1}=1$ is lower by δ_1, which has expected present value of $\beta\delta_1(1-F(\omega_0))$. The same benefit accrues at t+2, if both $u_{t+1}\in[\omega_0-\omega_1,\omega_0)$ and $u_{t+2}\in[\omega_0,\infty)$, with expected value of $\beta^2\delta_1[F(\omega_0)-F(\omega_0-\omega_1)][1-F(\omega_0)]$, with similar expressions holding for t+3, etc. Summing over time gives

$$\{1-\beta\int_{\omega_0-\omega_1}^{\omega_0} dF(u_{t+1})\}^{-1}\beta\delta_1\int_{\omega_0}^{\infty} dF(u_{t+1}) \tag{13}$$

This expression represents the reduction in future costs arising from current investment and again reflects the persistence in s_t, captured by the integral between $\omega_0-\omega_1$ and ω_0.

The sum of (10), (12), and (13) is equal to (9) and characterizes the optimal decision rule of households. The most important feature of the decision rule, summarized by (9), is the dependence of current actions on past decisions. Due to intertemporal increasing returns, shocks in the range $[\omega_0,\omega_0-\omega_1)$ lead to investment if received by an agent who has been active in the past ($s_{t-1}=1$) but not for an agent who has not invested at t-1, making individual behavior persistent.

4. Cyclical Fluctuations in the Aggregate Economy

(i) *Characterizing Output Fluctuations*

We now turn to the implications of individual level intertemporal increasing returns for aggregate economic fluctuations. We assume the economy consists of a continuum of agents, normalized to 1, each of whom faces the technology described above. Because the intertemporal increasing returns are internal to the firm, we allow for correlation between the different investment decisions (denoted s_t^i for agent i) of heterogenous agents by assuming the existence of an aggregate productivity shock v_t. Heterogeneity is first introduced via an idiosyncratic shock, ε_t^i such that each agent receives a shock $u_t^i=v_t+\varepsilon_t^i$. We assume that ε_t^i is drawn from a common distribution function G(.) (with associated density function g(.)) and that v_t is i.i.d with distribution function, H(.), and density function, h(.). Finally, we assume ε_t^i is uncorrelated between individuals and over time. Both shocks are normalized to have zero mean and are assumed to be observed before agents make their investment decisions.

The decision rule of the individual is given in section 3 and takes the form of invest iff $u_t^i \geq \omega_0 - \omega_1 s_{t-1}^i$. Conditioning on the aggregate shock, this can be written as, invest iff:

$$\epsilon_t^i \geq \omega_0 - \omega_1 s_{t-1}^i - v_t \tag{14}$$

where ω_0 and ω_1 are defined by the distribution of u_t^i, F(.). Investment decisions vary among firms due to the idiosyncratic shock, requiring s_t to be indexed by i. Defining S_t as the proportion of agents that invest in period t (equivalently, the aggregate propensity to invest), we have:

Proposition 2:

Aggregate output follows the process

$$\begin{aligned} \Delta Y_t &= \int \Delta y_t^i di = \alpha^0 + \alpha^1 S_t + \int_{i \in [s_t^i = 1]} u_t^i di \\ &= \alpha^0 + (\alpha^1 + v_t) S_t + \int_{\omega_0 - v_t}^{\infty} \epsilon g(\epsilon) d\epsilon + S_{t-1} \int_{\omega_0 - \omega_1 - v_t}^{\omega_0 - v_t} \epsilon g(\epsilon) d\epsilon \end{aligned} \tag{15}$$

where

$$\begin{aligned} S_t &= \{1 - G(\omega_0 - v_t)\}(1 - S_{t-1}) + \{1 - G(\omega_0 - \omega_1 - v_t)\} S_{t-1} \\ &= \{1 - G(\omega_0 - v_t)\} + \{G(\omega_0 - v_t) - G(\omega_0 - \omega_1 - v_t)\} S_{t-1} \end{aligned} \tag{16}$$

To understand (16), note that of the S_{t-1} firms who invested last period only those with an idiosyncratic shock greater than $\omega_0 - \omega_1 - v_t$ will invest now. Therefore, the measure of firms who invest in two successive periods is $(1-G(\omega_0-\omega_1-v_t))S_{t-1}$. Of the $(1-S_{t-1})$ firms who did not invest only those with an idiosyncratic shock greater than $\omega_0 - v_t$ will do so now and their measure is $(1-G(\omega_0-v_t))(1-S_{t-1})$. Therefore, as shown in Fig.1, S_t is a weighted average of points on the distribution function of idiosyncratic shocks, where the location of these points depends upon the aggregate shock and the weights depend on S_{t-1}.

(ii) The Nature of Business Cycle Fluctuations

Proposition 2 outlines an unobserved components model for GNP, as the law of motion for output consists of both a measurement equation, (15), and a state equation, (16). The importance of the state equation is that due its non-linear nature, it enables us to explain the asymmetries and the temporal agglomeration discussed in the introduction. S_t represents the average number of firms who are investing and is most naturally interpreted as the cyclical component of output, which is crucial in generating output fluctuations (e.g. Watson (1986), Hamilton (1989)). Variations in S_t not only alter the growth rate of output via (15), but also provide persistence because S_t follows a time varying

AR(1) process with autoregressive coefficient equal to $G(\omega_0-v_t)-G(\omega_0-\omega_1-v_t)$. Persistence is caused by shocks in the region $[\omega_0-\omega_1,\omega_0)$: in the case where $s_t^i=0$ a value of u_{t+1} in this region means it is optimal for $s_{t+1}^i=0$, whereas with $s_t^i=1$, $s_{t+1}^i=1$ would be optimal. Therefore, agents who invested last period have a higher propensity to invest this period than those who did not. In the aggregate this implies that i.i.d shocks are converted into persistent cyclical fluctuations. The autoregressive nature of the cyclical component S_t relies upon δ_1 being non-zero and so depends entirely on intertemporal increasing returns, if $\delta_1=0$ S_t is white noise. The need for an unobserved components approach arises from the fact that monitoring output growth requires tracking the number of agents who are/are not investing. This is needed because each type of agent responds to shocks differently, each group displaying persistence in their behavior. As a result, it is not sufficient to monitor only the shocks which impact on the economy, but also who will be affected by them.

The AR parameter in (16) is in general time varying and this feature enables us to account for the empirical findings of business cycle asymmetries; essentially, the impact effect of aggregate shocks on output varies over the business cycle. Referring back to Fig.1, changes in v_t shift the position of the two points along the horizontal axis: a high v_t shifts the chord AB down and S_t will be higher. However, the exact impact that v_t has depends upon both S_{t-1} and the cross sectional distribution of shocks, i.e. upon both the slope of the chord AB (which is determined by G(.) and v_t) and the weights on the two points (determined by S_{t-1}). As a result, the non-linear autoregressive form of (16) is a source of path dependence in our model as well as persistence; a shock which changes S_{t-1} both affects S_t through the AR coefficient but *also alters the way that the economy responds to future shocks* due to the interaction between v_t and S_{t-1}.

However, under the assumption of uniformly distributed idiosyncratic shocks these asymmetric interactions between S_{t-1} and v_t are absent. In this case the AR coefficient in (16) is constant and (15) and (16) approximate the standard "return to normality" state space model, variants of which have been used to successfully model U.S. GNP by Harvey (1985), Watson (1986) and Clark (1987)[7]. More generally, (15) and (16) allow for a number of alternative component based models which account for a wide range of asymmetries and cyclical fluctuations. Each of these different models of the cycle arise from alternative assumptions on the distribution of idiosyncratic shocks. However, although our model allows an important role for heterogeneity, we do not need to monitor complicated changes in the cross-sectional distributions to keep track of the state of the economy: what matters for business cycles is not the exact relative position of each agent over time but the distribution function for idiosyncratic shocks around particular ranges. The latter serves as a sufficient statistic for all distributional issues, considerably simplifying the analysis of the next subsection.

---

[7] The uniform distribution case only approximates the return to normality model due to the non-normality of the measurement equation disturbance and because, if v_t is such that either ω_0-v_t or $\omega_0-\omega_1-v_t$ is outside the support of ε_i, the autoregressive coefficient is no longer constant.

Motivated by these observations, we now turn to investigate how the degree of intertemporal increasing returns and heterogeneity interact to determine the time series properties of the aggregate economy, and of specifically the cyclical component S_t.

(iii) Determinants of the Time Series Properties of the Business Cycle

The non-linear nature of our model means there is no unique definition of persistence, but one candidate is the degree of serial correlation in S_t conditional upon v_t:

$$\frac{\partial S_t}{\partial S_{t-1}}|_{v_t}=G(\omega_0-v_t)-G(\omega_0-\omega_1-v_t) \tag{17}$$

Given G(.) is a distribution function we have:

Corollary 1: Business cycle persistence is non-decreasing in δ_1, the degree of intertemporal increasing returns.

Persistence is due to the fact that some agents have shocks in the interval $[\omega_0-\omega_1, \omega_0)$ and invest in this period only because investment costs are lower due to recent high activity. As ω_1 determines the measure of these marginal agents, and is itself a positive function of δ_1, serial correlation is strengthened when intertemporal increasing returns are higher.

To illustrate how the nature of the business cycle varies with different degrees of increasing returns consider the following illustrative simulation. Assume aggregate and idiosyncratic uncertainty to be equally important with both having a variance of 0.25, the former being normally and the latter uniformly distributed, and let the gains from learning-by-doing, α_1, be equal to 1.52 (see Section 6.1 for a justification of this choice). Figs. 2 and 3 show the cyclical component arising from these assumptions for the case $\delta_1/\delta_0=1/3$ and 1/2. Given the strong path dependence in our system Figs.2 and 3 are drawn for the same (suitably scaled) sequence of random shocks. Both figures illustrate how intertemporal increasing returns convert i.i.d shocks into cyclical fluctuations, however, the cyclical indicator is far more persistent in Fig.3 than in Fig.2. The increased learning-by-doing persuades firms to continue to invest even in the presence of mediocre productivity shocks, significantly reducing the noise in the cyclical component.

To understand the impact of distributional affects on the cyclical component we turn to a more general measure of persistence than (17), which was defined conditional on a given value of v_t. Integrating across all possible values of the aggregate shock, we arrive at a global measure of persistence;

$$P=\int\frac{\partial S_t}{\partial S_{t-1}}|_{v_t}\, h(v_t)dv_t \tag{18}$$

Taking a first-order Taylor expansion of P around μ_v, the mean of v_t which is assumed zero, we obtain

$$P \approx G(\omega_0)-G(\omega_0-\omega_1)+[g(\omega_0)-g(\omega_0-\omega_1)]\int(v_t-\mu_v)\ h(v_t)dv_t \tag{19}$$

By definition the second term is zero, and if we take a further first-order Taylor expansion of G(.) around ω_0, P can be approximated by $g(\omega_0)\omega_1$. Thus we can state

Corollary 2: For given ω_0 and ω_1, an increase in $g(\omega_0)$ will increase P.

The intuition behind this corollary is again related to the fact that individual persistence arises from shocks in the region $[\omega_0, \omega_0-\omega_1)$. In the aggregate economy the distribution of idiosyncratic shocks is important because it determines the density of agents who are around this critical region. For our global measure of persistence, $g(\omega_0)\omega_1$ is a measure of the number of such marginal agents, so that the higher is $g(\omega_0)$ the greater is serial correlation. Corollary 2 therefore implies that *spreads* of g(.) around ω_0 (which will often be produced by increases in heterogeneity, represented by increases in the variance of ε^i) will reduce persistence to the extent they lower the number of agents in the region $[\omega_0, \omega_0-\omega_1)$. The intuition of Corollary 2, that the persistence in the dynamics of this economy are linked to the form of the distribution of the idiosyncratic shocks, is important for our results in section 6, where we will investigate the empirical performance of different versions of this general models distinguished by the form of the heterogeneity and thus by the degree of asymmetry in the law of motion for output growth.

Now, to further analyze the impact of aggregate uncertainty on the business cycle we take a second-order Taylor expansion of (17) around μ_v followed by an additional first-order Taylor expansion around ω_0. This gives:

$$\begin{aligned} P &\approx G(\omega_0)-G(\omega_0-\omega_1)+[g'(\omega_0)-g'(\omega_0-\omega_1)]Var(v_t) \\ &\approx G(\omega_0)-G(\omega_0-\omega_1)+g''(\omega_0)\omega_1 Var(v_t) \end{aligned} \tag{20}$$

so we have;

Corollary 3: Increases in the variance of aggregate shocks reduce (increase) the persistence of the cyclical component if g(.) is concave (convex) around ω_0.

Corollary 3 states the surprising result that for a large class of idiosyncratic distributions, the more volatile are aggregate investment shocks, the less important is the business cycle - in the sense that the cyclical component becomes less persistent and aggregate output fluctuations are increasingly driven directly by v_t and not by the state equation. The intuition behind this is that when g(.) is concave around ω_0, increased volatility of the aggregate shock leads to the critical investment threshold being located at points with low density. In contrast, in the case of a convex g(.) function at ω_0, more aggregate variability will take us to values of the density function that are on average higher and thus increase serial correlation due to the increased weight of marginal agents.

(iv) Structural Heterogeneity

The purpose of this subsection is to show that when we extend our model to allow for different types of heterogeneity across agents, the importance of the cyclical component and the degree of non-linearities associated with the business cycle may be enhanced. In particular we show that increasing the dispersion of agents can, under certain conditions, increase the amount of persistence provided by our cyclical indicator. Given that $\{S_t\}$ represents the extent of co-movement between agents, this is a surprising result which re-iterates the limitations of representative agent models.

We have so far only considered what may be termed stochastic heterogeneity using the terminology of Caballero and Engel (1991), that is heterogeneity in the form of idiosyncratic shocks. Whereas in practice, agents also differ significantly in their technological possibilities/preferences as well as the opportunities they face, often captured in microeconometric work by the inclusion of fixed effects. To include these influences we consider *structural heterogeneity* by allowing investment cost functions to be firm specific. We assume agent i has investment costs given by[8]

$$C_t^i=(\delta_0^i-\delta_1 s_{t-1}^i)s_t^i \tag{21}$$

Using the solution outlined in Section 3 each agent invests iff

$$\epsilon_t^i \geq \omega_0^i-\omega_1 s_{t-1}^i-v_t \tag{22}$$

where the distribution function of ω_0^i is $\Gamma(.)$ with support set U and is determined from the distribution of the cost parameter, δ_0^i. Increases in the degree of structural heterogeneity are equivalent to mean-preserving spreads of $\Gamma(.)$. The law of motion for S_t is now

$$\begin{aligned} S_t&=(1-S_{t-1})\{\int_U(1-G(\omega_0^i-v_t))d\Gamma(\omega_0^i)\}+S_{t-1}\{\int_U(1-G(\omega_0^i-\omega_1-v_t))d\Gamma(\omega_0^i)\} \\ &=\{\int_U(1-G(\omega_0^i-v_t))d\Gamma(\omega_0^i)\}+S_{t-1}\{\int_U\{G(\omega_0^i-v_t)-G(\omega_0^i-\omega_1-v_t)\}d\Gamma(\omega_0^i)\} \end{aligned} \tag{23}$$

Applying the same two successive Taylor expansions as in the last subsection, our global measure of persistence becomes

$$P\approx\int_U [G(\omega_0^i)-G(\omega_0^i-\omega_1)]d\Gamma(\omega_0^i)\approx\omega_1\int_U g(\omega_0^i)d\Gamma(\omega_0^i) \tag{24}$$

Corollary 4: If g(.) is convex (concave), mean preserving spreads of $\Gamma(.)$ increase (decrease) P.

---

[8] The intertemporal increasing returns parameter, δ_1, can similarly be made individual specific.

Therefore, an increased dispersion of agents in the form of greater structural heterogeneity can interact with stochastic heterogeneity to increase the persistence generated through the cyclical component. The intuition here is that if g(.) is convex, the averages of neighboring points will be higher than g(.) itself. Since the higher is g(.) the higher is the measure of agents in the critical region where investment is only profitable when costs are low, in the case of convex g(.), the presence of structural heterogeneity leads to a more serially correlated process for $\{S_t\}$. We can therefore see that a mean-preserving spread (an increase in structural heterogeneity), by extending the cross sectional distribution of ω_0^i to regions of g(.) that are higher, increases persistence in the economic state. Conversely when g(.) is globally concave, an increase in structural heterogeneity will reduce persistence. The message from this section is clear, far from removing the non-linear nature of the business cycle, heterogeneities have a key determining influence on economic fluctuations.

5. Regime Shifts in Economic Fluctuations

One of the attractions of fixed cost models is that the discrete individual behavior they imply can explain the sharpness of business cycle turning points. While our general model implies that the cyclical component is always serially correlated, it does not impose any conditions on the nature of turning points. A number of studies (e.g. Hamilton (1989), Acemoglu and Scott (1994), Suzanne Cooper (1994), Diebold and Rudebusch (1994)) have achieved empirical success in modelling output fluctuations by assuming the business cycle to be characterized by regime shifts, that is by abrupt moves from recession to expansion (or vice versa). In the next subsection we investigate the factors which determine the nature of turning points in our model, establishing the circumstances under which models with regime shifts serve as a good approximation.

(i) The Nature of Turning Points

To analyze this issue we use our basic model with only stochastic heterogeneity, and focus on the probability of a given S_t occurring conditional upon S_{t-1}, which is :

$$\rho(S|S')=\frac{h(v)}{(1-S')g(\omega_0-v)+S'g(\omega_0-\omega_1-v)} \tag{25}$$

where $\rho(S|S')$ denotes the density of S_t conditional on $S_{t-1}=S'$ and v_t is written as an implicit function of S_t by (16).

The extreme case of regime shifts corresponds to the case where $S_t=0$ or 1 so that $\rho(S|S')$ has its mass concentrated at two particular points. Similarly in more general characterizations, if $\rho(S|S')$ has marked peaks, then transitions between different states take on the character of regime shifts. A necessary (but not sufficient) condition for $\rho(S|S')$ to have its mass concentrated in two different intervals is that it should be non-unimodal. Thus, if we can establish $\rho''(S|S')$ is positive at $\rho'(S|S')=0$, we will have located a local minimum which implies $\rho(S|S')$ cannot be single peaked.

From (25) we have:

$$sign(\rho''(S|S'))=sign[\frac{((1-S')g''(\omega_0-v)+S'g''(\omega_0-\omega_1-v))}{|(1-S')g'(\omega_0-v)+S'g'(\omega_0-\omega_1-v))|}-\frac{h''(v)}{|h'(v)|}] \tag{26}$$

While no general statement can be made regarding the transition between states we have the following:

Proposition 3: The conditional density of S_t is non-unimodal if g(ε) is more concave than h(v) in the neighborhood of ρ'(S|S')=0.

Proposition 3 suggests turning points tend to be abrupt when g(.) is locally more concave than h(.). Though this is not a necessary condition, naturally when g(.) has its mass concentrated at a particular point, it will tend to be more concave. This can be seen in Fig.1. If the idiosyncratic shock is uniformly distributed, G(.) is a straight line and output growth, though persistent, is distributed uniformly along the continuum $(\alpha_0,\alpha_0+\alpha_1)$. However, if the distribution function is concentrated in the middle (as in Fig.1) economic states become more distinct in the sense that the conditional distribution of $\{S_t\}$ is concentrated in particular intervals of (0,1). As a consequence, turning points are more likely to approximate the discreetness of regime shifts.

To investigate further this point, we utilize a simulations approach. Recall that Fig.3 showed the case where $\delta_1/\delta_0=1/2$ and the aggregate and idiosyncratic shocks were equally uncertain with a variance of 0.25. Fig. 4 maintains the degree of increasing return at the same level but increases the variance of the idiosyncratic shock to 2.5. The very different cyclical patterns in the two figures are readily apparent: in Fig.3, there are never less than 60% of agents investing so that even in recession large numbers of agents are still active, and turning points are extremely sharp, particularly the observations at around 37 and 73. In contrast, the greater importance of idiosyncratic shocks adds a considerable amount of noise to the cyclical indicator in Fig.4. And while the turning points at observations 37 and 73 can still be detected, they represent only two of several observations where the cyclical component changes direction.

The above results and discussion show again that the form of heterogeneity is crucial, this time in determining the sharpness of turning points and they also suggest that the limiting case where aggregate uncertainty is much more important than idiosyncratic uncertainty is useful to analyze as a benchmark.

(ii) A Model of Regime Shifts

Proposition 3 and related simulations suggest that the more concentrated the idiosyncratic distribution the more appropriate is a regime shift characterization. Further, Corollary 2 implies that increases in heterogeneity may reduce the persistence of cyclical fluctuations by reducing $g(\omega_0)$. This suggests that the business cycle will display extreme regime shifts and possess a highly persistent cyclical component in a version of our model with no heterogeneity. In this subsection we focus on such a model by setting $u_t^i = v_t$. The particular interest in this special case is that it offers a theoretical justification for the widely used discrete Markov state space models.

Because the model only contains an aggregate shock, the laws of motion for the aggregate economy are the same as those for the individual firm. From our results of Section 3(ii) and introducing the notation

$$Prob[S_t=1\,|S_{t-1}=1]=p \qquad Prob[S_t=0\,|S_{t-1}=0]=q \tag{27}$$

we have

Proposition 4: The stochastic process for the change in aggregate output is

$$\Delta Y_t = -\alpha_0 + (\alpha_1 + v_t) S_t \tag{28}$$

where S_t is a Markov chain with transition matrix

$$T = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{29}$$

and

$$p = \int_{\omega_0-\omega_1}^{\infty} dH(v)\ , \quad q = \int_{-\infty}^{\omega_0} dH(v) \tag{30}$$

As in the model with heterogeneity, cyclical fluctuations are the result of random shocks hitting the economy. These shocks are propagated by intertemporal increasing returns in a manner which requires a state space formulation for output growth. The distinguishing feature of (28)-(30) is that fluctuations take the form of shifts between distinct economics states. In each of these states the economy behaves differently, not only does the growth rate differ ($\alpha_1 \neq 0$) but if $p \neq 1-q$ then so do the durations of booms and recessions, features common with the model of Hamilton (1989) which closely corresponds to (28)-(30).

This regime shift model also shares a number of similarities with the model of Durlauf (1993) who also explicitly models the transition between states of the business cycle and focuses on technology adoption. In both models this choice involves a non-convexity and intertemporal increasing returns to scale with the end result being that white noise productivity shocks are converted into serially correlated output fluctuations. Both models also suggest a strong role for path dependence, in (28)-(30) shocks which shift the economy from one state to another are highly persistent as they affect the way the economy responds to future shocks. However, a key difference between ours and Durlauf's work comes in the form that the intertemporal non-separability takes. In Durlauf's model the intertemporal linkage arises through localized technological spill-overs; in other words firms have a higher propensity to invest if their neighbors invested in the recent past. The impact of this externality is the focus of Durlauf's paper. Due to the externality Durlauf's model generates multiple long run equilibria in the sense that the stochastic process for output is non-ergodic. However, in our model, because increasing returns to scale are internal, and this not only implies that the equilibrium path is uniquely determined, i.e. given v_t and $\{S_{t-1},S_{t-2},..\}$, we know with certainty in which state the economy will be, but also that $\{\Delta Y_t\}$ is always ergodic.

6. Econometric Evidence

In this section we use the econometric results of other researchers as well as maximum likelihood estimates of our own general model of Section 4 to assess what scale of intertemporal increasing returns and what form of heterogeneity are necessary to account for the business cycles in the U.S. We also use the analysis of Sections 3 and 5 to infer from our estimates the exact form of U.S. business cycle asymmetries and the relative importance of idiosyncratic and aggregate sources of uncertainty.

(i) Regime Shift Models

Hamilton (1989) estimates a two state discrete Markov model for US GNP which is nearly identical to (28)-(30) and finds p=0.9, q=0.76 and α_1=1.52. To calculate the implied degree of intertemporal increasing returns we solve the equations in (A3) in the appendix using these estimated parameter values and making assumptions regarding the distribution and variance of aggregate shocks plus the magnitude of the learning-by-doing effects (δ_1/δ_0)[9]. Fig.5 shows different combinations of δ_1/δ_0 and σ^2_v which generate p=0.9 and q=0.76 when aggregate shocks are normally distributed. To obtain the same persistence in S_t, a higher variance requires more increasing returns. The reason for this is that the greater the variance of shocks, the more likely agents are to receive a future shock less than ω_0-ω_1. This lessens the expected future benefits from increasing returns and makes agents both less likely to invest and less likely to remain investing once they have done so. Thus with increasing returns of 22% (δ_1/δ_0=0.22) to produce the appropriate values of p and q we require σ^2_v=0.05, but when σ^2_v=0.01 we need only δ_1/δ_0=11%. As well as providing persistence, intertemporal increasing returns generate considerable amplification of the productivity shock. For the case where δ_1/δ_0=0.22 and σ^2_v=0.05, even though σ^2_v/α_1 = 0.03, the variance of ΔY_t is equal to 0.634. Relying *only* on a

---

[9] α_0 does not influence the values of ω_0 and ω_1 and so (without loss of generality) it is set to ensure the model matches mean US output growth.

single productivity shock to drive output fluctuations, we need implausibly large learning-by-doing effects of 53% to explain Hamilton's results. But with an additional additive disturbance in (1), as assumed by Hamilton and all econometric implementation of unobserved component models, his results can be explained with very small amounts of intertemporal increasing returns and aggregate uncertainty.

Results from alternative studies confirm this finding that only small scale intertemporal increasing returns are necessary to generate empirically observed regime shift behavior. Suzanne Cooper (1994) uses monthly industrial production from 1931 to estimate a transition matrix similar to (28). To match her estimates of the transition probabilities (p=0.55 and q=0.46) while also matching the variance of US GNP growth (again without resorting to any additional productivity disturbances other than a unique aggregate shock), we need the saving in fixed costs arising from learning-by-doing to be only around 3% (i.e. $\delta_1/\delta_0=0.03$). Diebold and Rudebusch (1994) estimate equations analogous to (28) using US industrial production (allowing for a time dependent **T**). Assuming only one disturbance and using their estimated standard error for ΔY_t to calibrate σ_v^2 (an underestimate as this implicitly sets $S_t=1$ for all t), we find $\delta_1/\delta_0=0.8\%$ is sufficient to explain their results.

(ii) *The General Unobserved Components Model*

In this subsection we use (15) and (16) to obtain estimates of our structural parameters. These equations represent a general state space model and require assumptions regarding the distribution of idiosyncratic shocks before they can be estimated. We examine these equations assuming first that idiosyncratic shocks are uniformly distributed and then that they are normally distributed. As well as enabling us to assess the robustness of our estimates of structural parameters these assumptions allow us to examine the importance of asymmetries in U.S. output fluctuations. As discussed above, the assumption of uniform idiosyncratic shocks removes any asymmetry in the state equation. Estimating (15) and (16) under different distributional assumptions therefore enables a simple heuristic test of the importance of asymmetries in U.S. business cycles.

Assuming that idiosyncratic shocks are uniformly distributed over the range [-a,a], (15) and (16) can be written as

$$\Delta Y_t=b_0+b_1S_t+b_2v_t+b_3v_t^2$$

$$\begin{aligned}S_t&=[\frac{a-\omega_0}{2a}]+\frac{\omega_1}{2a}S_{t-1}+\frac{v_t}{2a}\\&=c+TS_{t-1}+u_t\end{aligned}\qquad\textbf{(31)}$$

where the coefficients b_i are functions of the structural parameters ω_0, ω_1, a, α_0 and α_1 and aside from the squared disturbance in the measurement equation this is the standard return to normality model (e.g. Harvey (1989), ch.3). Various versions of this model have been used to estimate the cyclical component of U.S. GNP, with Watson (1986) and Clark (1987) both estimating models similar to (31)[10]. If instead we assume that idiosyncratic shocks are distributed normally, then[11] (taking second-order Taylor approximations for all higher order terms, full details available from the authors)[12];

$$\Delta Y_t = b_0 + b_1 S_t + b_2 v_t + b_3 v_t^2 + b_4 v_t S_{t-1}$$

$$\begin{aligned} S_t &= (1-G(\omega_0)) + [G(\omega_0)-G(\omega_0-\omega_1)]S_{t-1} + g(\omega_0)v_t + \\ &\quad (g(\omega_0-\omega_1)-g(\omega_0))v_t S_{t-1} \\ &= c + TS_{t-1} + Ku_t S_{t-1} + u_t \end{aligned} \tag{32}$$

Equations (31) and (32) show that a simple test for asymmetry in either the state or measurement equation is to test the significance of the $v_t S_{t-1}$ term.

We estimate (31) and (32) using the growth rate of quarterly U.S. real GNP for the period 1954:1 and 1987:4. Because the equations contain only one shock, we augment the measurement equation in each case with a normally distributed additive measurement error. Estimation was performed using maximum likelihood via the Kalman filter. We also ignored the squared disturbance term which considerably simplified the estimation. In neither case did our estimates reveal any indication of heteroscedasticity, suggesting that dropping the squared term was not an important omission. Because in (31) and (32) the measurement and state equation have a correlated disturbance a simple alteration is required to the standard recursions of the Kalman filter (see Harvey (1989), p.112). Table 1 contains our estimation results, and Figure 6 shows the (smoothed) estimates of S_t that emerge from the different assumptions on idiosyncratic disturbances. Both versions of the model suggest that a persistent cyclical component successfully accounts for serial correlation in U.S. output growth.

Examining the uniform distribution case (equation (31)), we find that the cyclical component is persistent, with an autoregressive coefficient of 0.52 and the model has a R^2=0.595, thus explains just under 60% of the variation in U.S. GNP growth. The estimates of the structural parameters that we uncovered were as follows: the variance of the aggregate shock, $\sigma_v{}^2$, was 0.06, the variance of the idiosyncratic shock, $\sigma_\epsilon{}^2$, was 1.45. This implies that the

---

[10] Watson and Clark estimate their model using the level of the U.S. real GNP allowing for a trend, a cycle and an irregular component. Thus our comparison is purely with their specification of the cyclical component, our equation (16). They both actually estimate the cyclical component as an AR(2) process. Acemoglu and Scott (1993) show this can easily be allowed for by letting intertemporal increasing returns operate with longer lags.

[11] Equation (32) is an approximation which is valid for any distributional assumption regarding idiosyncratic shocks. Different assumptions regarding G(.) lead to different estimates of the model's structural parameters.

[12] The exact expression for b_4 is as follows:

$$b_4 = [G(\omega_0)-G(\omega_0-\omega_1)][1-\frac{[\int_{\omega_0-\omega_1}^{\omega_0} ug(u)du]\times\{g(\omega_0-\omega_1)-g(\omega_0)\}}{[G(\omega_0)-G(\omega_0-\omega_1)]^2}] + (\omega_0-\omega_1)g(\omega_0-\omega_1) - \omega_0 g(\omega_0).$$

This term disappears when idiosyncratic shocks are uniform.

variance of idiosyncratic shocks is around 25 times that of aggregate shocks, thus heterogeneity being relatively much more important than aggregate uncertainty. This result is in line with those of Schankerman (1991) and Davis and Haltiwanger (1992) from micro data that idiosyncratic variations are large compared to comovement across all agents. Given our results in the last section, this finding also explains why statistical tests of Hamilton's (1989) model (e.g. Hansen (1992) and Garcia (1992)) reject the notion of regime shifts between distinct states as an appropriate characterization of U.S. business cycles. We also recovered α_1 to be 0.0201, which implies that when the business cycle indicator S_t is at its peak, GNP growth is 2% higher compared to a trough.

Finally, using our estimated structural parameters and assuming a real interest rate of 4% per annum we can use equations (A3) to calculate estimates for δ_0 and δ_1, the learning-by-doing parameters. We find that $\delta_0 = 1.96$ and $\delta_1 = 0.54$ implying intertemporal increasing returns (δ_1/δ_0) of around 27.6% (as a proportion of fixed costs).

The estimates of equation (32) which allows for asymmetries reveal considerably greater persistence in the cyclical component -- the autoregressive coefficient in the state equation is now 0.67 as opposed to 0.52 in the uniform case. More importantly, R^2 in this case goes up to 0.712, thus the model with normally distributed idiosyncratic shocks explains over 70% of the variation in U.S. business cycles. The superior performance of this version of the model is also significant as equation (31) is essentially what Clark, Harvey and Watson have estimated, thus the more general model suggested by our theory that allows for asymmetries outperforms the conventional statistical model.

Further, the interaction term between v_t and S_{t-1} is significant in both the state and the measurement equations, strongly so in the former, confirming the asymmetric nature of U.S. business cycles and thus indicates where the improved performance of (32) comes from. Figure 6 shows the estimated sequence for the cyclical indicator arising from assuming normal and uniform idiosyncratic distributions. The strong correspondence between the two sequences is encouraging since it implies that we are basically uncovering the same underlying component in both exercises. However, there is a major difference between the two series: the asymmetric model enables much sharper downward swings into recession and it is this feature of the data which accounts for the success of the more general model. These sharper downswings are captured by the positive coefficient on the $v_t S_{t-1}$ term and are in line with the findings of Neftci (1984) who used a different statistical methodology than the one here.

Again uncovering our structural estimates, we find that $\alpha_1 = 0.017$ which is less than the estimate of the growth difference between boom and recession that we obtained from (31). However, note that this does not imply growth to be higher by only 1.7% at the peak of the cycle because in contrast to (31), we also have the added flexibility of having the asymmetry term $v_t S_{t-1}$. The presence of this term implies that as long as we remain in a boom (i.e. positive values of v_t) we obtain an added growth effect and this is of the order of 0.2%, thus at the peak of the cycle growth is higher by 1.9%. But also this term implies that when there is a downturn (v_t<0), this can happen rather sharply. Returning to the rest of the estimates, σ_v^2 was 0.035 and σ_ϵ^2 was 0.1207, so our estimate of the ratio of the variance of idiosyncratic shocks to aggregate shocks is around 3.5, considerably less than the approximate estimate of 10 from Schankerman (1991) and Davis and Haltiwanger (1992), though still giving a substantial role to idiosyncratic disturbances in generating uncertainty. The fact that our asymmetric model, (32), leads to a large estimate of the relative variance of aggregate shocks may be due to the absence in our model of any interactions across individual investment decisions. As a result, what may be firm specific effects translated into co-movements by economic interactions across agents will be estimated as aggregate shocks in our specification. Examining whether this is actually the case clearly requires a richer data set than the aggregate time series we use here. However, our estimates show that even without spillover effects we can have a model where idiosyncratic shocks are the major source of uncertainty yet output growth shows a clear cyclical pattern with sharp swings into recession. Resorting again to (A3) we find that our asymmetric model leads to estimates of δ_0 and δ_1 of 1.81 and 0.63 respectively, so that intertemporal effects of around 1/3 are required to explain U.S. fluctuations.

It is also interesting to note that the increased complexity of (32) implies that our structural parameters are overidentified (see for instance the expression in footnote 11) whereas they were exactly identified in (31). This gives us two overidentifying restrictions with which to test the plausibility of our model. Performing a Wald test on these restrictions gives a test statistic of 4.1 which is asymptotically distributed chi-squared with two degrees of freedom, thus comfortably accepting the overidentifying restrictions. This result is promising as it is another piece of evidence that our model provides a good match to the cyclical dynamics of U.S. GNP.

7. Conclusion

We have outlined a theory of economic fluctuations based on internal intertemporal increasing returns in a model of discrete investment choice. This model is motivated by microeconometric findings of persistence in firm level investment as well as the emphasis placed in the management science and economics of technology literature on the importance of learning-by-doing in the adoption of new technologies. Incorporating these effects into a model of a firm's investment choice leads to a tractable model of business cycle fluctuations.

This tractability enables us to fully analyze the determinants of aggregate economic fluctuations. Our theoretical findings are: (i) Intertemporal increasing returns naturally lead to temporally agglomerated and asymmetric economic fluctuations, in the sense of persistent periods of low and high growth separated by business cycle turning points (ii) Cross-sectional considerations play a key role in determining the non-linear and asymmetric nature of fluctuations and distributional issues are found to be crucial in determining the sharpness of turning points and deciding whether aggregate output experiences regime shifts. (iii) Although heterogeneity plays a key role, all the business cycle relevant information is captured in one variable, *the average number of active agents*, thus our model enables a synthesis between representative agent models and those stressing the importance of heterogeneity.

A major attraction of our model is its ability, under certain simplifying assumptions, to justify the popular unobserved component models of Watson (1986), Clark (1987) and Hamilton (1989) all of which place a special emphasis on an underlying cyclical indicator. An additional attraction is the model's ability to offer a general formulation for cyclical components, and provide a simple test for the presence of business cycle asymmetries. The data suggest that our general model that incorporates asymmetric effects provides a good fit to U.S. business cycles; that these are characterized by asymmetries, with particularly fast downturns that cannot be captured by a linear model; that the cyclical component is highly persistent; and, that despite the asymmetries, business cycles are not characterized by sharp swings such as required by regime shift models. In our context, this latter fact corresponds to the aggregate sources of uncertainty being dominated by idiosyncratic shocks, which is what investigations on micro data also find. Estimates of the degree of internal intertemporal increasing returns necessary to account for U.S business cycles vary between 3 and 53%, suggesting that under some circumstances only very modest amounts of intertemporal increasing returns are needed. Adding other sources of uncertainty to the model or allowing for spillovers between agents would serve to reduce even further the extent of internal increasing returns required. Assessing the relative importance of internal and external intertemporal increasing returns is an obvious topic of further research. At this stage, our results lead us to conclude that intertemporal increasing returns may be an important channel of persistence, amplification and asymmetries in economic fluctuations.

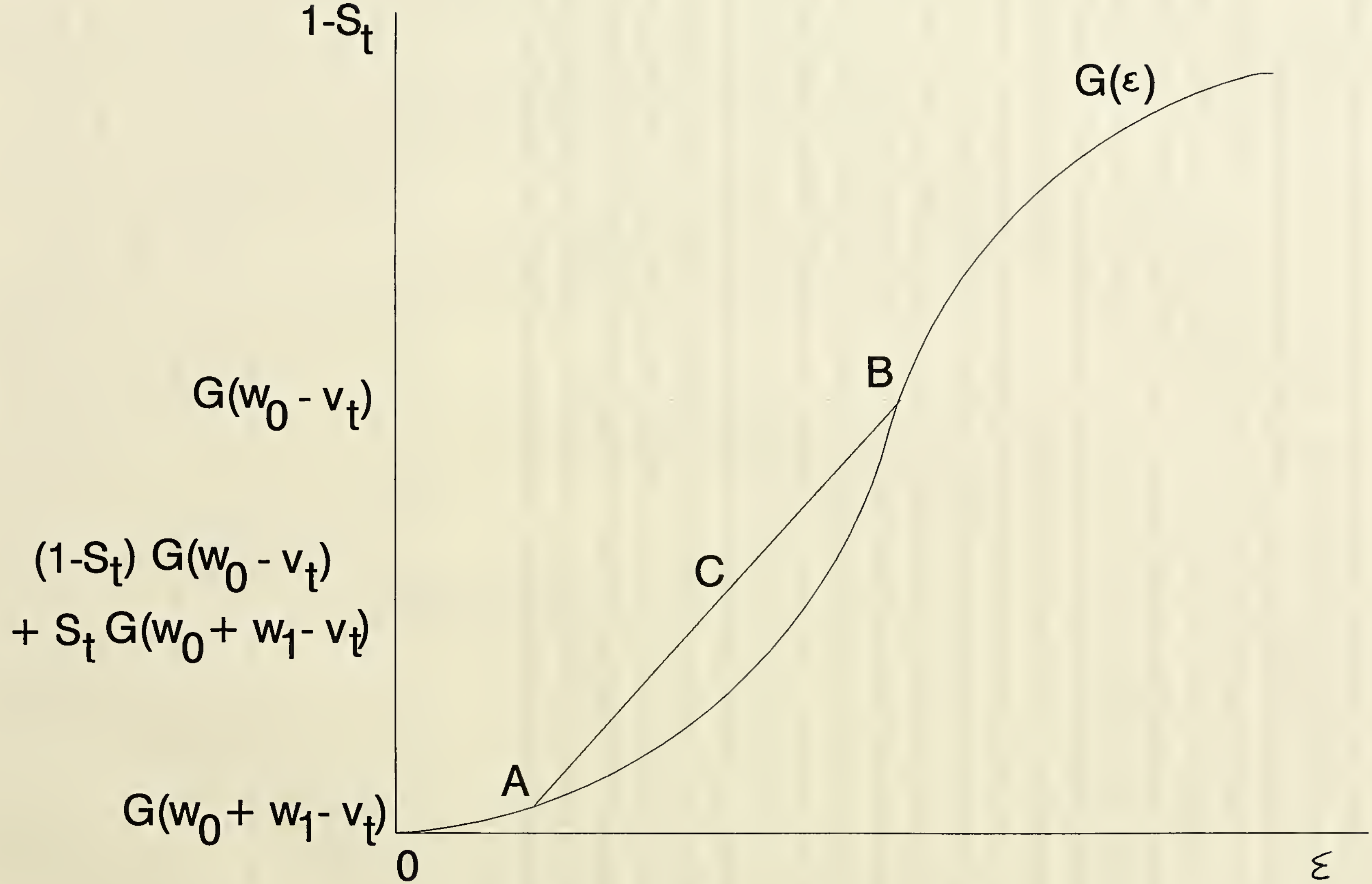

Figure 1

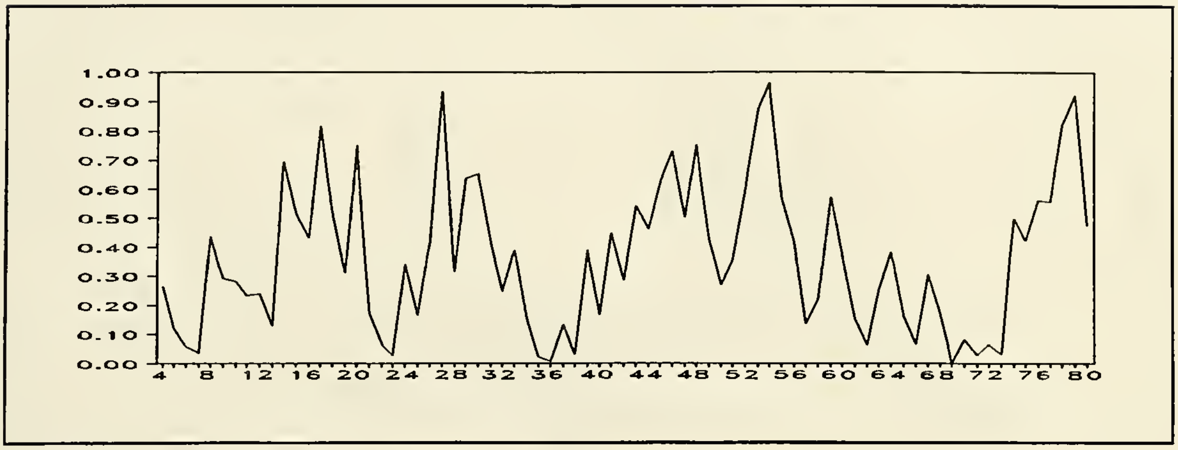

Figure 2 : $\delta_1/\delta_0=0.33$ $\sigma^2_v=0.25$ $\sigma^2_\varepsilon=0.25$ $\sigma^2_S=0.064$ $\sigma^2_{\Delta y}=0.841$ $\sigma^2_\lambda=0.025$

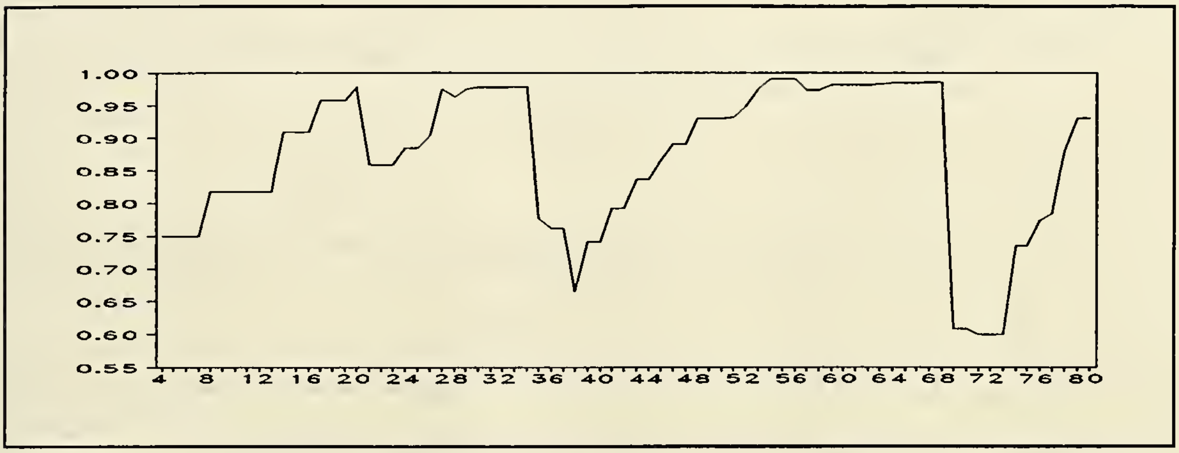

Figure 3 : $\delta_1/\delta_0=1/2$ $\sigma^2_v=0.25$ $\sigma^2_\varepsilon=0.25$ $\sigma^2_S=0.014$ $\sigma^2_{\Delta y}=0.541$ $\sigma^2_\lambda=0.032$

Figure 4 : $\delta_1/\delta_0=1/2$ $\sigma^2_v=0.25$ $\sigma^2_\varepsilon=2.5$ $\sigma^2_S=0.012$ $\sigma^2_{\Delta y}=4.855$ $\sigma^2_\lambda=0.001$

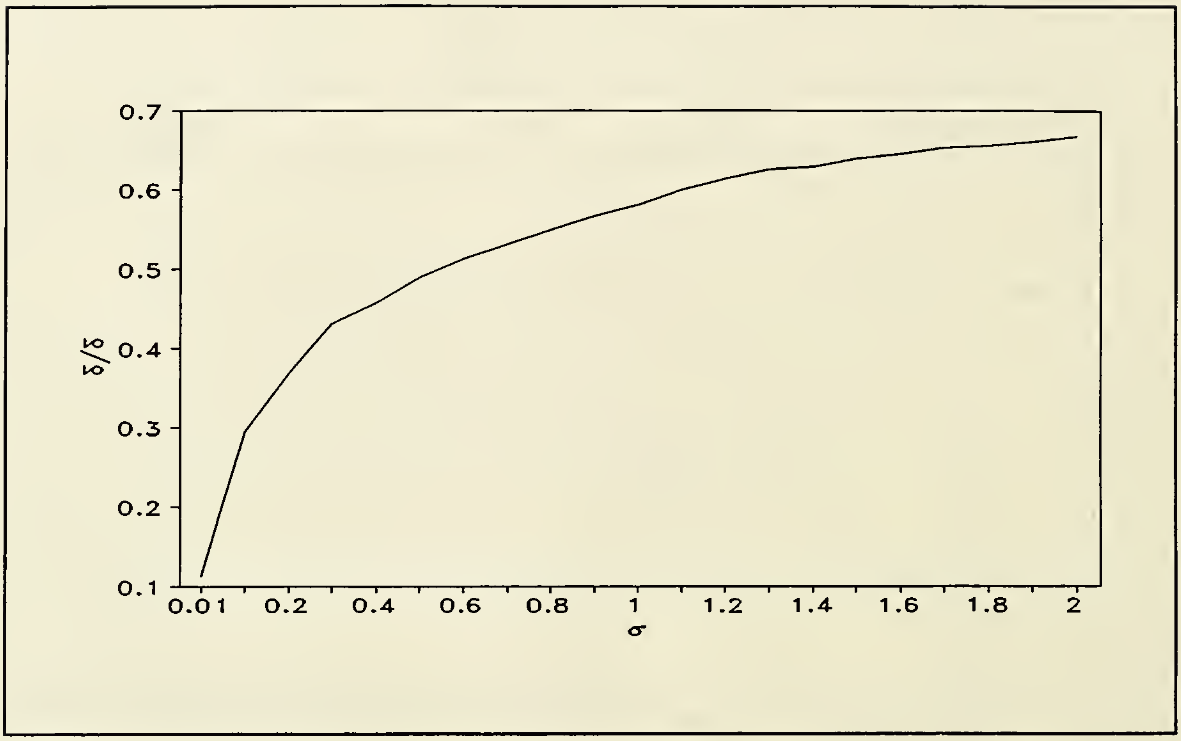

Figure 5 : Learning-by-Doing and Volatility of Aggregate Shocks

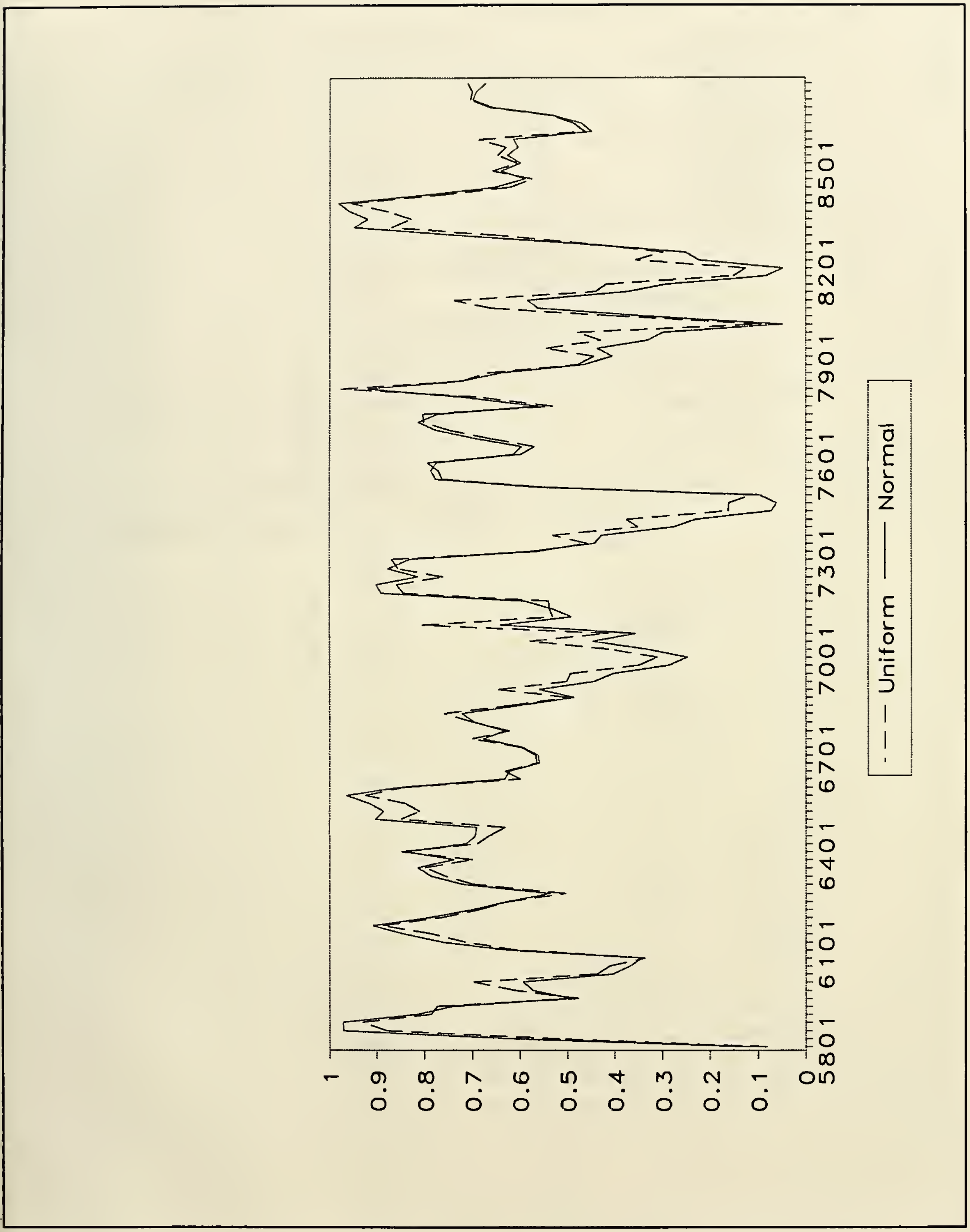

Figure 6 : The Cyclical Component of U.S. GNP Growth

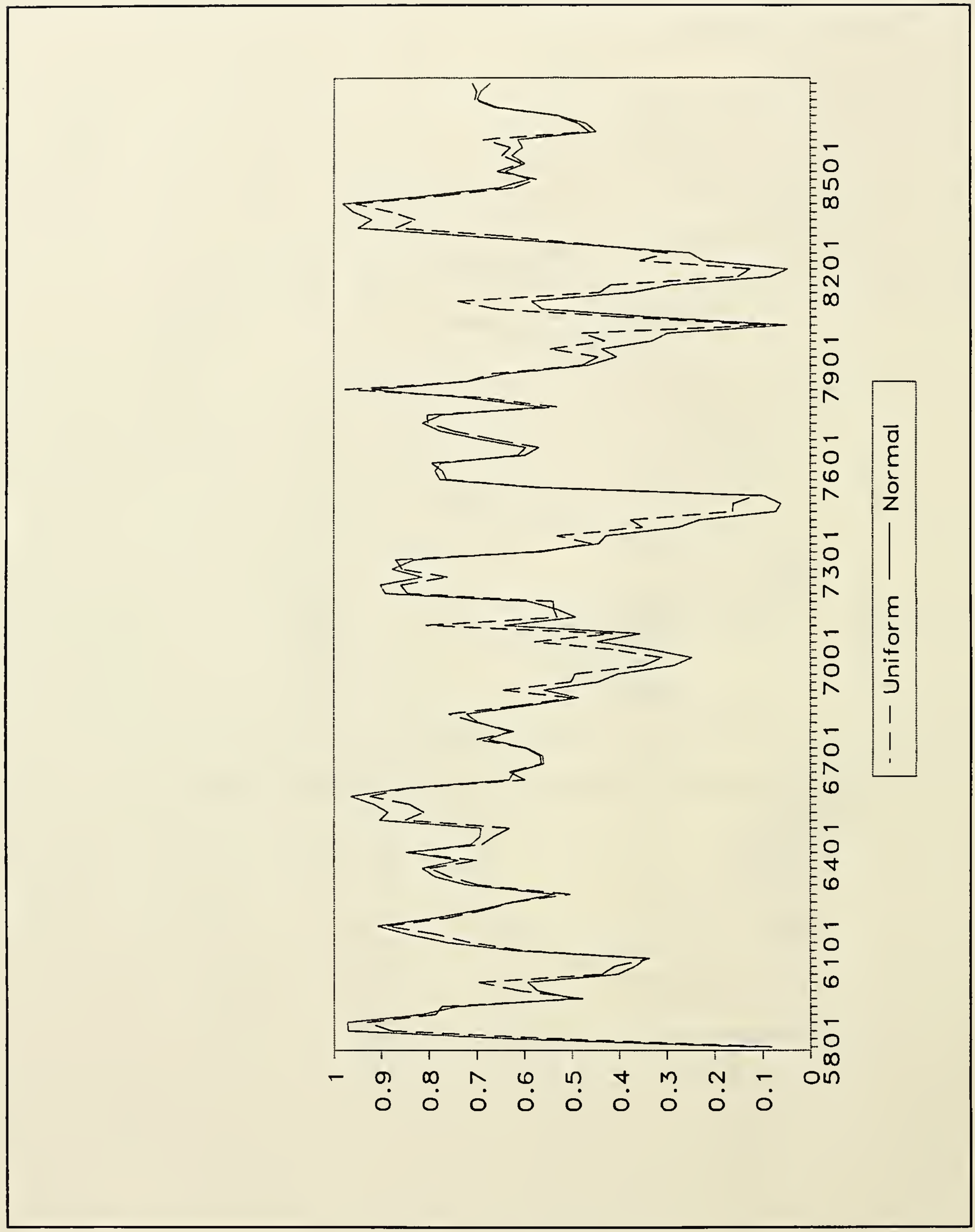

Figure 6 : The Cyclical Component of U.S. GNP Growth

Table 1 : Estimates of General and Uniform Model

	Uniform Case		Normal Case	
Parameter	Estimate	Tstatistic	Estimate	Tstatistic
b_0	0.0004	1.691	0.0020	1.957
b_1	0.0201	3.107	0.0173	3.284
b_4			0.1027	1.985
c	0.1826	2.232	0.1769	2.173
T	0.5213	5.669	0.6653	6.339
K			0.3229	3.496
$\sigma_u{}^2$	0.0070	2.703	0.0065	2.842
R^2	0.595		0.712	

Sample period 1957:1,1987:4. Data = Quarterly growth in U.S GNP. Estimation by Maximum likelihood via Kalman filter.

Appendix

Proof of Proposition 1

We need to establish that (i) the value function defined in (8) satisfies (7) for particular values of ω_0 and ω_1 and ϕ's (ii) the value function satisfies a transversality condition.

To establish (i) we substitute (8) into (7) to derive:

$$\begin{aligned}
&\textit{If } u_t \geq \omega_0-\omega_1 s_{t-1}, \textit{ then} \\
&\phi_o+\phi_y y_{t-1}+\phi_s s_{t-1}+\phi_u u_t = y_{t-1}+\alpha_0+\alpha_1+u_t-(\delta_0-\delta_1 s_{t-1}) \\
&+\beta\Big[\int_{\omega_0-\omega_1}^{\infty}(\phi_0+\phi_y(y_{t-1}+\alpha_0+\alpha_1+u_t)+\phi_s+\phi_u u_{t+1})dF(u_{t+1}) \\
&\quad+\int_{-\infty}^{\omega_0-\omega_1}(\phi_0'+\phi_y(y_{t-1}+\alpha_0+\alpha_1+u_t))dF(u_{t+1})\Big]
\end{aligned} \tag{A1}$$

where a similar equation applies for $v_t < \omega_0-\omega_1 s_{t-1}$. Equating coefficients gives

$$\begin{aligned}
&\phi_\alpha=\phi_u=\frac{1}{1-\beta} \qquad \phi_s=\delta_1 \\
&\phi_0-\phi_0'=\frac{\frac{1}{1-\beta}\alpha_1-\delta_0+\beta\delta_1\int_{\omega_0-\omega_1}^{\infty}dF(u_{t+1})+\frac{\beta}{1-\beta}\int_{\omega_0-\omega_1}^{\omega_0}u_{t+1}dF(u_{t+1})}{1-\beta\int_{\omega_0-\omega_1}^{\omega_0}dF(u_{t+1})}
\end{aligned} \tag{A2}$$

To find expressions for ω_0 and ω_1 we use (8) to see under what conditions the left hand side of (5) is greater evaluated at $s_t=1$ than evaluated at $s_t=0$. This gives

$$\begin{aligned}
&\omega_1=\delta_1(1-\beta) \\
&\omega_0=\Big[1-\beta\int_{\omega_0-\delta_1(1-\beta)}^{\omega_0}dF(u_{t+1})\Big]^{-1}\times\Big[(1-\beta)\delta_0-\alpha_1- \\
&(1-\beta)\beta\delta_1\int_{\omega_0-\delta_1(1-\beta)}^{\infty}dF(u_{t+1})-\beta\int_{\omega_0-\delta_1(1-\beta)}^{\omega_0}u_{t+1}dF(u_{t+1})\Big]
\end{aligned} \tag{A3}$$

Defining the right hand side of the definition of ω_0 as $\tau(\omega_0)$ we need to prove $\tau(\omega_0)$ has a fixed point. Defining z(u) as τ(u)-u we have

$$\lim_{\omega\to+\infty} z(\omega)=-\infty \quad \textit{and} \quad \lim_{\omega\to-\infty} z(\omega)=+\infty \tag{A4}$$

Continuity of z(u) follows from continuity of $\tau(.)$ and by the intermediate value theorem z(u) must have a zero. Thus a fixed point, ω_0, of $\tau(.)$ exists.

(ii) The value function defined in (8) satisfies a transversality condition iff

$$\lim_{t\to\infty}\beta^t\int_{-\infty}^{\infty} V(\alpha_{t-1},s_{t-1},u_t)dF(u_t)=0 \tag{A5}$$

since V(.) is linear condition is satisfied for ß<1.

Establishing uniqueness of ω_0 requires proving $z(\omega_0)$ has a unique fixed point. $z(\omega_0)$ is everywhere differentiable and its derivative equals

$$\begin{aligned} z'(\omega_0) &= -\beta[\omega_0 f(\omega_0)-\omega_0 f(\omega_0-(1-\beta)\delta_1)]\times[1-\beta\int_{\omega_0-(1-\beta)\delta_1}^{\infty} dF(u_{t+1})]^{-1} \\ &+[(1-\beta)\delta_0-\alpha_1-\beta(1-\beta)\delta_1\int_{\omega_0-(1-\beta)\delta_1}^{\infty} dF(u_{t+1})-\beta\int_{\omega_0-(1-\beta)\delta_1}^{\omega_0} u_{t+1}dF(u_{t+1})] \\ &\times\beta[f(\omega_0)-f(\omega_0-(1-\beta)\delta_1)][1-\beta\int_{\omega_0-(1-\beta)\delta_1}^{\omega_0} dF(u_{t+1})]^{-2}-1 \end{aligned} \tag{A6}$$

Substituting $z(\omega_0)=0$, we have $z'(\omega_0)=-1$ which establishes that $z(\omega_0)=0$ can only be true at a unique value of ω_0.

We now establish by contradiction that there can be no non-linear solutions to the recursion (7). Let $W(y_{t-1},s_{t-1},u_{t-1})$ be a solution to (7) and for given values of the state variables let s_t^w be the optimal choice of s_t. Thus

$$\begin{aligned} W(y_{t-1},s_{t-1},u_t) &= y_{t-1}+\alpha_0+(\alpha_1+u_t-\delta_0-\delta_1 s_{t-1})s_t^w \\ &+\int_{-\infty}^{\infty} W(y_{t-1}+\alpha_0+(\alpha_1+u_t)s_t^w,s_t^w,u_{t+1})dF(u_{t+1}) \end{aligned} \tag{A7}$$

Observation 1: $s_t^w(y_{t-1},s_{t-1},u_t)$ cannot depend on y_{t-1} as returns from a higher y_{t-1} accrue under both $s_t^w=0$ and $s_t^w=1$.

We know by assumption that both V(.,.,.) in (8) and W(.) satisfy (7). Take a value of u_t where V(.) gives $s_t^v=0$. The difference between (5) evaluated at W(.) and V(.), for given u_t, is

$$W(y_{t-1},s_{t-1},u_t)-\phi_0-\phi_y y_{t-1}=(\alpha_1+u_t-\delta_0-\delta_1 s_{t-1})s_t^w+$$
$$\beta\{E_t W(y_{t-1}+\alpha_0+(\alpha_1+u_t)s_t^w,s_t^w,u_{t+1})$$
$$-\int_{\omega_0}^{\infty}(\phi_0+\phi_y(y_{t-1}+\alpha_0)+\phi_u u_{t+1})dF(u_{t+1})-\int_{-\infty}^{\omega_0}(\phi_0'+\phi_y(y_{t-1}+\alpha_0))dF(u_{t+1})\} \tag{A8}$$

From Observation 1 this holds for all y_{t-1} for fixed u_t, implying that

$$W(y_{t-1},s_{t-1},u_t)-\beta E_t W(y_{t-1}+\alpha_0+(\alpha_1+u_t)s_t^w,s_t^w,u_{t+1})=K_1+k_1 y_{t-1} \tag{A9}$$

Thus W(.,.,.) is linear in y_{t-1} with a constant coefficient. To see this note that we can repeat this exercise for $s_{t-1}=0$ or 1 and get the same coefficient k_1. Similarly we can repeat the exercise for values of u_t at which $s_t^v=1$ and get the same relationship.

Observation 2: If

$$E_t\sum_{j=0}^{\infty}\beta^j r(y_{t+j-1},s_{t+j-1},u_{t+j}|s_t=1)>E_t\sum_{j=0}^{\infty}\beta^j r(y_{t+j-1},s_{t+j-1},u_{t+j}|s_t=0) \tag{A10}$$

is true for $u_t=u$, it is also true for $u_t=u+\Delta$, such that $\Delta>0$. The left-hand side of (A10) is increasing in u_t whereas the right-hand side does not depend on u_t. This implies there exists a value of u_t, $u^w(s_{t-1})$ (independent of α_{t-1} by Observation 1) such that $u_t\geq(<)u^w(s_{t-1})$ implies that $s_t^w=1$ (=0). Suppose $u^w(s_{t-1})\geq\omega_0-\omega_1 s_{t-1}$ (the argument for the reverse inequality is analogous).

First, consider u_t in the interval $(-\infty,\omega_0-\omega_1 s_{t-1})$. For all u_t in this interval, $s_t^w=s_t^v=0$. Varying u_t in this interval gives:

$$W(y_{t-1},s_{t-1},u_t)-\beta E_t W(y_{t-1}+\alpha_0,0,u_{t+1})=K_2 \tag{A11}$$

where K_2 does not depend upon u_t. As there is no investment we see from Observation 2 that W(.) is independent of u_t in the interval $(-\infty,\omega_0-\omega_1 s_{t-1})$.

Consider a value of u_t in the interval $(u^w(s_{t-1})),+\infty)$. This implies $s_t^v=s_t^w=1$. Considering variations of u_t in this interval gives

$$W(y_{t-1},s_{t-1},u_t)-\beta E_t W(y_{t-1}+\alpha_0+\alpha_1+u_t,1,v_{t+1})=K_3+(\phi_u-\beta\phi_\alpha)u_t \tag{A12}$$

However W(.,.,.) is linear in α_{t-1}, which is linear in u_{t-1} with coefficient ϕ_u, thus

$$W(y_{t-1},s_{t-1},v_t)=K_4+\phi_v v_t \tag{A13}$$

Thus over this range too W(.,.,.) is linear in u_t, with coefficient $\phi_u=1/(1-\beta)$.

We also need to consider the interval $(\omega_0-\omega_1 s_{t-1},v^w(s_{t-1}))$ in which $s_{t-1}^w=0$ and W_t does not depend on u_t over $(\omega_0-\omega_1 s_{t-1},u^w(s_{t-1}))$. Thus the coefficient of u_t of the value function W(.,.,.) has the same form as the value function in (7)-(8).

Finally we need to check how W(.) is influenced by s_{t-1}. However this variable only takes the values 0 and 1. Thus any general function $h(s_{t-1})$ can be written as $k_4 s_{t-1}$, implying any value function that satisfies (7) must have the same form as V(.). QED

References

Abernathy, W. (1980) *The Productivity Dilemma*, Baltimore John Hopkins

Acemoglu, D. and Scott, A. (1993) "A theory of economic fluctuations: Increasing Returns and Temporal Agglomeration" Centre For Economic Performance Discussion Paper No.163

Acemoglu, D. and Scott, A. (1994) "Asymmetries in the cyclical behavior of UK labor markets", *Economic Journal* November 1994

Arrow, K. (1974) *The Limits of Organization*, New York, Norton.

Bahk B. H. and M. Gort (1993) "Decomposing learning by doing in new plants" *Journal of Political Economy* Vol 101, pp 561-583.

Bertola, G. and R. Caballero (1990) "Kinked adjustment costs and aggregate dynamics" *NBER Macroeconomic Annual* MIT Press.

Bond, S., Elston, J, Mairesse, J and Mulkay, D (1994) "A comparison of empirical investment equations using company panel data for France, Germany, Belgium and the UK", Oxford University mimeo

Bond, S. and Meghir, C. (1994) "Dynamic investment models and the firm's financial policy", *Review of Economic Studies*, 61, 197-222

Caballero, R. and Engel, E (1991) "Dynamic (S,s) Economies", *Econometrica*, 59, 1659-1686

Caballero, R. and Engel, E (1994) "Explaining investment dynamics in US manufacturing: A generalized Ss approach" MIT Mimeo

Clark, P. K. (1987) "The cyclical component of U.S economic activity", *Quarterly Journal of Economics*, 102, 703-735

Cooper, R. and Haltiwanger, J (1993) "The aggregate implications of machine replacement: Theory and evidence", *American Economic Review*, 83, 360-82

Cooper, R, Haltiwanger, J and Power, L (1994) "Machine replacement and the business cycle : Lumps and Bumps", Boston University mimeo

Cooper, S. (1994) "Multiple regimes in U.S. output fluctuations", Harvard University mimeo

Davis, S.J and Haltiwanger, J (1992); "Gross Job Creation, Gross Job Destruction, and Employment Reallocation", *Quarterly Journal of Economics* Vol 107, 819-864

Diebold, F. X. and Rudebusch G. D. (1990) "A non-parametric investigation of duration dependence in the American business cycle", *Journal of Political Economy*, 98, 596-616

Diebold, F.X and Rudebusch, G.D (1994) "Measuring business cycles: A modern perspective", Working Paper, University of Pennsylvania

Doms, M. and T. Dunne (1993); "Capital adjustment patterns in manufacturing plants" Mimeo

Durlauf, S. N. (1991) "Multiple equilibria and persistence in aggregate fluctuations", *American Economic Review, papers and proceedings*, 81, 70-74

Durlauf, S. N. (1993) "Nonergodic economic growth", *Review of Economic Studies*, 60, 349-366

Freeman, C. (1982) *The Economics of Industrial Innovation*, Cambridge, MIT Press

Garcia, R (1992) "Asymptotic null distribution of the likelihood ratio test in Markov switching models", CRDE, Montreal, mimeo

Hall, R. E. (1991) *Booms and Recessions in a Noisy Economy*, Yale:Yale University Press

Hamilton, J. D. (1989) "A new approach to the economic analysis of non-stationary time series and the business cycle", *Econometrica*, 57, 357-384

Hansen, B.E (1992) "The likelihood ratio test under nonstandard conditions : testing the Markov switching model of GNP", *Journal of Applied Econometrics*, 7, S61-82

Harvey, A. C. (1985) "Trends and cycles in macroeconomic time series", *Journal of Business and Economic Statistics*, 3, 216-227

Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press

Hirsch, W. Z. (1952) "Manufacturing progress functions" *Review of Economics and Statistics* Vol 34, pp 143-155.

Lahiri, K. and Moore, G. (1991) *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge: Cambridge University Press

Lieberman, M. B. (1984) "The learning curve and pricing curve in chemical processing industries" *Rand Journal of Economics* Vol 15, pp 213-28

Myers, S. and D. G. Marquis (1969) *Successful Industrial Innovations*, Washington DC, National Science Foundation.

Neftci, S. N. (1984) "Are economic time series asymmetric over the business cycle?", *Journal of Political Economy*, 92, 307-28

Nelson, R. R. and S. G. Winter (1982) *An Evolutionary Theory of Economic Change*, Cambridge, MIT Press.

Pennings, J. M. and A. Buitendam (1987); *New Technology as Organizational Innovation; The Development and Diffusion of Microelectronics* eds, Ballinger Publishing Company

Scarf, H. E. (1959) "The optimality of (S,s) policies in the dynamic inventory problem" in K.Arrow, S.Karlin, P.Suppes (eds) *Mathematical Methods in Social Sciences*, Stanford University Press

Schankerman, M. (1991); "Revisions of Investment Plans and the Stock Market Rate of Return", NBER Working Paper No.3937

Stock, J. H. and Watson, M. (1989) "New indexes of coincident and leading economic indicators", in Blanchard, O and Fisher, S (eds) *NBER Macroeconomics Annual 1989*

Tushman, M. L. and P. Anderson (1986);"Technological discontinuities and organizational environments" *Administrative Science Quarterly* Vol 31, pp 439-465.

Watson, M. (1986) "Univariate detrending methods with stochastic trends", *Journal of Monetary Economics*, 18, 49-761