NBER WORKING PAPER SERIES

LEARNING, COMPLEMENTARITIES AND ASYNCHRONOUS USE OF TECHNOLOGY

Boyan Jovanovic
Dmitriy Stolyarov

Working Paper 5870

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1997

This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Learning, Complementarities and
Asynchronous Use of Technology
Boyan Jovanovic and Dmitriy Stolyarov
NBER Working Paper No. 5870
January 1997
JEL No. L23
Productivity

ABSTRACT

This paper deals with processes that require several complementary inputs subject to improvements in quality. If after a quality upgrade one of these inputs requires a period of learning before it can be used effectively, then in general it will pay to purchase the inputs at different dates -- the purchases will be asynchronous. That is so because it is wasteful to tie up funds in the other inputs which will be underutilized until the date learning is over.

We provide evidence that technology has been used asynchronously in the automobile industry, in the television broadcasting industry, in electricity supply, and in railways, and we argue that our model helps explain this evidence.

Boyan Jovanovic
Department of Economics
New York University
269 Mercer Street
New York, NY 10003
and NBER
jovanovi@fasecon.econ.nyu.edu

Dmitriy Stolyarov
Department of Economics
University of Pennsylvania
Philadelphia, PA 19104
stolyar@econ.sas.upenn.edu

Learning, Complementarities and Asynchronous Use of Technology

Boyan Jovanovic
New York University
jovanovi@fasecon.econ.nyu.edu

Dmitriy Stolyarov
University of Pennsylvania
stolyar@econ.sas.upenn.edu

December 4, 1996

Abstract

This paper deals with processes that require several complementary inputs subject to improvements in quality. If one of these inputs requires a period of learning after a quality upgrade before it can be used effectively, then in general it will pay to purchase the inputs at different dates – the purchases will be asynchronous. That is so because it is wasteful to tie funds up in the other inputs which will be underutilized until the date learning is over.

1 Introduction

In a typical industry, the typical user of a technology has not developed it himself. Instead, he has adopted someone else's invention. Most users therefore improve their technologies not by invention, but by *adoption*. Zeckhauser's (1968) prototype adoption model assumes that the user must use one technology at a time. That is, to adopt a new technology the user must *drop* the old one. This assumption holds in a variety of situations. Many types of equipment (cars, computers, machine tools) are hard to modify – e.g., one can only use one type of chip in a computer.

The typical new technology also involves a period of *learning* and productivity growth. Output can fall right after adoption takes place, and it takes a while before productivity under the new method exceeds peak productivity of the old one. This foregone output is an implicit cost of switching, and in Zeckhauser's model, this cost of acquiring the technology-specific expertise prevents users from upgrading at a maximal rate. There are also explicit costs of switching technologies. To use a new method, one typically must buy some capital goods specific to that technology. Zeckhauser does not consider such costs, but in his model that would be a routine extension.

We extend Zeckhauser's model by assuming that the operation of a technology requires not one, but two *complementary* inputs, which we think of as two capital goods – e.g., hardware and software, or designs and manufacturing equipment. Our model contains human and physical capital costs of technological switching.

1.1 Explanations of synchronized behavior with complementarities

Many technologies divide into components each of which is embodied in a different durable good. But can we, for most purposes, still treat them as a single composite capital good? Since these inputs are complementary, can't we aggregate them and treat them as one input? The literature might indeed lead us to conclude that the answer is "Yes". When inputs in production are strong complements, it has been argued that these quantities will move together over time and space – in sync:

(i) Milgrom and Roberts (1990) argue that when a plant retools, or when a firm reorganizes, its complementary inputs should be readjusted at the same time.

(ii) Cooper and Haltiwanger (1993) say that complementarity causes agents' actions to be positively correlated, and investment to be "bunched" at certain dates.

(iii) Matsuyama (1995) argues that complementarity implies that various sectors of an economy will tend to develop at the same time.

(iv) Aghion and Howitt (1996) say that the adoption of general purpose technologies will be delayed until complementary technologies become available.

(v) Kremer (1993) says that when different tasks are complements in production, plants will segregate on the basis of the qualities of all their inputs.

(vi) Sheshinski and Weiss (1992) say that when two *products* are complements in the firm's profit function, their costly price changes will be synchronized.

The general thrust of these claims may be valid, but our paper will show that there is an important exception: When an input's productivity grows through learning or training, it may pay to economize on other inputs until the input in question is "up and running". That is, *learning creates asynchronization.*

1.2 Evidence of asynchronous use of technology

Here are two examples of complementary inputs upgraded asynchronously:

(a) Designs and manufacture can be treated as complementary inputs: A given manufacturing process is capable of supporting only up to a certain number of designs in production. The quality of output is measured by the variety of the models offered to the consumer. In contrast to the old mass-production technology, an efficient lean-manufacturing automobile factory can quickly shift to a new model without absorbing a productivity drop. But it can take 10 years for a plant to master the lean manufacturing technique, and U.S. plants that have adopted it are not yet at the same level of mastery as the Japanese plants: The U.S. plants started switching to lean manufacturing in the mid 1980's, but as late as 1990, the Japanese auto industry was still producing twice as many models as the Americans did (Womack pp. 119-120). It apparently will not pay U.S. firms to produce that many models until they master the new technique.

(b) Programming quality and broadcast quality are the two complementary inputs in the TV industry. The quality of output can be measured by its entertainment value. Before electronic TV there was mechanical TV which could not support high-quality programming because of low resolution. The development of electronic TV began in the early 1930's, when mechanical TV was largely abandoned – a technological *switch* took place even before the new technology was perfected. TV stations *chose* not to invest in high-quality programming. Instead, they switched to audio broadcasting (lower output) and used their video equipment for research. Apparently it did not pay to raise programming quality until high definition electronic TV was fully developed. (Udelson, 1982, p. 78).

The same logic works for asynchronized introduction of complementary *products*. Learning implies that it does not pay to introduce the second complementary product until the first one has been learned. Again, learning creates asynchronization. Here are two more examples, but this time of the asynchronous introduction of complementary products:

(c) Street lighting (supplied at night) and industrial electrification (supplied during the day) are complementary products for an electric utility. A big component of cost is their shared capacity – the power plant – which serves to produce both outputs. During the early years of electrification, utilities took a long time to gain expertise in efficiently transporting electricity to the user. In the late 1890's and early 1900's, they supplied mainly street lighting. Only by 1910 did the utilities became efficient enough to start supplying factories too. (Nye 1995, pp. 235 - 236). So street lighting service and factory electrification were introduced at different dates. It could not pay utilities to supply electricity to factories until they had learned to do it efficiently.

(d) Passenger service (needed during the day) and freight service (which can be supplied at night) are complementary products for a railroad. A big component of cost is their shared capacity – the railway track or the locomotive – which serves to produce both outputs. Freight train operation by itself is harder to master (Pollins 1971, p. 61, Sherrington 1928, p. 172). As a result, English railroads in the early 19th century primarily transported passengers until around 1840, by which time the technology had improved enough to accommodate freight. It did not pay railroads to move freight until they had raised their efficiency.

1.3 Other explanations of asynchronous behavior

We will use learning to explain asynchronous use of technology. But there are other explanations of asynchronous behavior generally, and possibly even of asynchronous use of technology. Not all are formally worked out in the literature, but they deserve mention.

1) In a two-sector model, Benhabib and Nishimura (1985) find that if the consumption-goods sector is more capital-intensive than the capital-goods sector, the sectors move asynchronously. If we have a lot of capital goods today, the resulting abundance of capital next period will make it optimal to produce a lot of the (capital-intensive) consumption good tomorrow. But then there will be less capital the day after. It then becomes optimal to focus on capital goods, and so on. The force here is not learning, but rather an off-diagonal weighting in the input-output matrix. This theory works at the aggregate level, however, and it can not explain the above examples of asynchronous behavior.

2) Internal costs of rapid adjustment *other than* implicit learning costs can also make it optimal to stagger the introduction of complementary inputs. This argument may in part explain example (a), where one might argue that a team of designers can either design new models, *or* help with the layout of the factory, but not both. But it is hard to argue this for the TV example, in which it seems instead that the costs of adjusting the two inputs were unrelated, and that costs depend on quality, and not on the relative timing of introduction. It is even harder to argue this for examples (c) and (d), where creating the capacity to produce one product in fact clearly *lowers* the cost of supplying its complement, though not significantly enough to warrant immediate production.

3) Financing constraints can make it optimal to stagger the introduction of complementary, but expensive capital goods. This hypothesis says that profitable investments are staggered – i.e., delayed – until they can be financed. The argument implies, however, that a big company would move earlier to introducing profitable opportunities that are, for financial reasons, unaffordable to smaller companies. That is, a big firm should adopt technologies synchronously. But this fails in all four examples. All automobile manufacturers are big companies, yet they all moved asynchronously. Railroads, electric utilities, and some TV companies are all big companies, and yet they too all moved asynchronously.

4) Time to build is similar to our learning hypothesis in that there is a waiting time between the date at which investment in a new technology begins, and the date at which it can yield (higher) output. In such a model, it is possible, and optimal to continue using the old technology while the new one is under installation. Then when the new technology is ready, it is optimal to invest in its complementary inputs. This argument works well at the aggregate level, but not so well at the micro level when

simultaneous use of two different technologies is infeasible. For instance, in an automobile plant, it is not possible to run mass production alongside lean manufacturing. In the TV example, programming stopped around 1933, and the video equipment was used instead for research. And it is hard to think what (other than expertise) it took decades to build in the electricity example. The hypothesis may explain part of the railway example, however, in that the movement of freight requires not only track and locomotives, but also different cars, warehouses and other local infrastructure.

No doubt that these hypotheses, and other considerations besides, may illuminate some aspect or other of each example. At this stage, we wish only to establish that learning is an additional hypothesis that can improve our understanding of asynchronous behavior in various contexts.

1.4 Plan of the paper

Throughout the paper, we shall model the optimal policy of a firm that produces *one* output with *two* inputs. This means that formally, we will be addressing examples (a) and (b) only. However, the same force -- learning -- will also explain the phenomena in examples (c) and (d). To show this, we shall, at the end of the paper, show how with minor modifications, the model can be transformed into one with one input and two outputs. One of the inputs is subject to a learning curve after it is installed, and its purchased quality is bounded above by frontier quality that grows exogenously. The other input can be upgraded any time, without limit. When learning is relatively unimportant, it is optimal to upgrade both inputs at the same time, in sync. But when learning matters enough, it is optimal to stagger the upgrading of the two inputs, as examples (a) and (b) show, and as management scientists emphasize.[1]

We will show that whenever upgrades make output drop, it is better to upgrade the inputs asynchronously. Baloff (1970) argues that output drops are especially prevalent in machine-intensive manufacturing. Furthermore, for a non-degenerate subset of parameter values, the producer optimally *chooses* to incur the drop in output, deriving benefit from rapid growth associated with more frequent upgrading.

Our other result concerns aggregation. When the two inputs are optimally upgraded together, in sync, then the two input problem degenerates to the one-input case. But when the parameters are such that the optimal upgrading policy is asynchronous, then the two-input case is genuinely different.

The next section presents the model. Section 3 contains the aggregation result, section 4 compares synchronous and asynchronous upgrading policies, and section 5 discusses optimal policies. Section 6 asks if our conclusions are robust to various changes in the assumptions. In particular, it introduces the two-output variant of the model. Finally section 7 contains the Bellman equation specification of the model.

[1] For instance, Attewell (1992, p. 1) advances the thesis that "Firms delay in-house adoption of complex technology until they obtain sufficient technical knowhow to implement and operate it successfully"

2 The Two-Input Model

Until section 6, we shall deal exclusively with the two-input model. A firm's production function is Leontief with two inputs x and z:

$$y(t) = \min\{x(t), z(t)\}.$$

We think of x and z as the services provided by two complementary capital goods – say, computer hardware and software. We assume that there is a fixed input of labor, and that $y(t)$ is labor productivity. Under this interpretation, a worker is assigned to the two machines, and machines can not be shared. We assume that the size of the firm is exogenous so that the labor input is equal to 1. Then x and z are the qualities of the two inputs that the firm's only worker is using to produce output.

2.1 Feasible input choices

Only one unit of each input can be used in production at any time, and x, z measure quality, rather than quantity of inputs. A new quality of each input can be purchased at any time. It is not possible to add to the old levels of x or z – only one unit of each may be used at a time, and when a new input is purchased, the old one is useless.

The producer can purchase any quality of input x. More formally, suppose that new x is purchased at times s_j, $j = 0, 1, ...$, and the quality of x purchased at time s_j is x_j. Likewise, let new z be purchased at times τ_i and the quality be z_i. Then the time paths for input qualities $x(t)$ and $z(t)$ will be fully determined by the sequences of (nonnegative) numbers$\{x_j, s_j\}_{j=0}^{\infty}$ and $\{z_i, \tau_i\}_{i=0}^{\infty}$ respectively:

$$x(t) = x_j, t \in [s_j, s_{j+1}), j \in \mathbf{N},$$

$$z(t) = \begin{cases} \theta z_i & t \in [\tau_i, \tau_i + T) \\ z_i & t \in [\tau_i + T, \tau_{i+1}) \end{cases} \quad i \in \mathbf{N},$$

where $z_i \leq Z(\tau_i) = Z(0)e^{g\tau_i}$ for every i, where Z and θ will be defined presently.

Note two asymmetries in what is assumed about feasible choices of inputs x and z:

1) While any quality of input x is a feasible choice, the quality of input z is constrained by the frontier Z, which grows exogenously at rate g: $Z(t) = Z(0)e^{gt}$. This exogeneity premise is appropriate at the micro level, where most innovation consists of adopting someone else's invention. But it is not appropriate at the very aggregate level: Because the advance of the frontier is exogenous to the decision unit, this unit must be small in relation to the sector that produces the growth in Z. A team, a plant, a firm, or even a small open economy all satisfy this criterion.

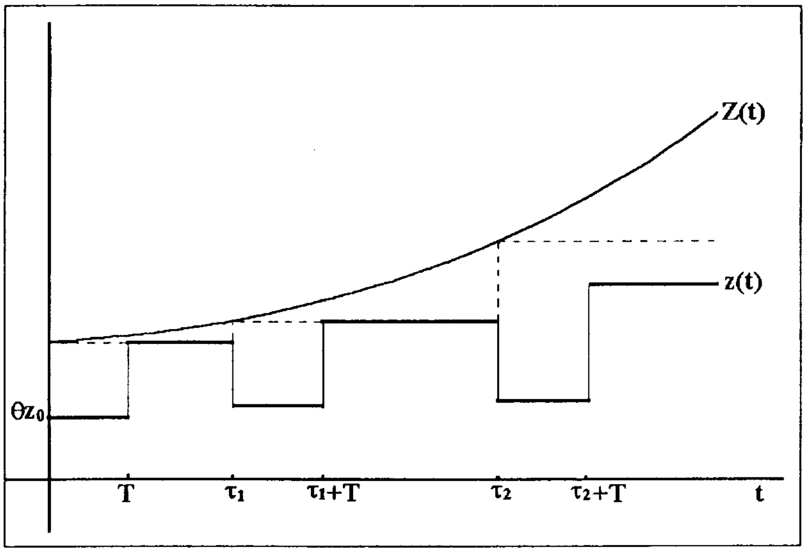

Figure 1: The time path for the frontier $Z(t)$ and input productivity $z(t)$

2) For T periods after each upgrade of z, only a fraction $\theta \in [0,1]$ of its quality can be used in production. This captures the idea that after each purchase of new z the producer has to learn how to use it, and while he is learning, his productivity is lower. In contrast, x has no learning.

These assumptions simplify the analysis but do not affect the main results. The following subsection addresses learning in more detail.

2.2 Nature of learning

To explain the nature of the learning process, we portray a typical time path for input z in Figure 1. Here input z is purchased at times $0, \tau_1$,and τ_2. Levels z_0 and z_1 are shown to be at the frontier, but z_2 is below the frontier. Productivity of z_i for the first T periods after its purchase is θz_i.

The time path $z(t)$ consists of a sequence of learning curves. Each learning curve is L-shaped so that there is a well defined learning period, or "start-up" period of the kind emphasized by management scientists (Baloff 1970), and found in plant-level productivity by Bakh and Gort (1993). L-shaped learning is an extreme form of the S-shaped learning curves that can arise when an activity is complex (see Jovanovic and Nyarko 1995 figure 1).

Another source of L-shaped learning is analyzed by Stolyarov (1996). When learning draws effort away from production, and when the returns to effort are linear, it is optimal to use a "bang-bang" policy: All effort goes to learn the technology for some initial time interval $[0,T]$, and then effort switches to production. During the start-up phase $[0,T]$, output will be low (say $\theta < 1$), and after that, output will be high (say 1).

L-shaped learning vastly simplifies the analysis. Our result that asynchronous upgrading sometimes dominates synchronous upgrading, by continuity, remains valid when L-shaped learning is approximated closely enough by continuous learning. We now note some specific properties of this learning process:

(i) Learning is measured by θ and T. Relative to peak productivity, output "lost" to learning is $(1-\theta)T$. If θ is close to 1 or if T is small, learning is unimportant.

(ii) After an upgrade in z, output may drop.

(iii) Learning is not transferable between technologies: Learning the current technology does not depend of how many grades were learned previously. In this respect, our model is like Zeckhauser's and not like Parente's (1994).

(iv) Once in place, a technology's productivity depends only on time, and can not be hastened by spending or by speeding up production. As Bakh and Gort (1993, p. 563) put it, "...learning may simply depend on time...[because of] intertemporal substitution in which costs are reduced by later delivery."

(v) Learning z is not interactive with x. For instance if z is a machine tool and x is a raw material, this says that you can debug the tool using any quality material.

(vi) If a factory can make only one product, one can interpret T as "lead time" – the time elapsed from the start of work on the new product to market introduction.

(vii) This is not a "T-period time to build z" model even when $\theta = 0$. In such a model, one could order the new z input T periods before dropping the old z, and hence suffer no fall in output. Here one must install the new z before learning can start.

The capacity constraint – only one x and only one z per user is more realistic at the micro level. At the aggregate level, many technologies are used simultaneously, and our model does not apply.

2.3 Prices and the maximization problem

Let the input prices per unit quality be constant and equal to p_x and p_z, respectively. The price of output is constant and normalized to unity. Then, lifetime profit is given by:

$$V(\{x_j, s_j\}, \{z_i, \tau_i\}) = \int_0^\infty e^{-rt} \min\{x(t), z(t)\} dt - \sum_{i=0}^{\infty} e^{-r\tau_i} p_z z_i - \sum_{j=0}^{\infty} e^{-rs_j} p_x x_j \quad (1)$$

The first term is present value of revenue, the second term gives the total cost of input z. The third term represents the total cost of input x. The producer maximizes his lifetime profit, given by (1), with respect to $x(t)$ and $z(t)$, with the initial conditions $x(0) = \hat{x}$, $z(0) = \hat{z}$:

$$\max_{\{z_i,\tau_i\},\{x_j,s_j\}} V(\{x_j, s_j\}, \{z_i, \tau_i\})$$
$$s.t.\ \forall i,j \quad z_i \leq Z(\tau_i)$$
$$x_j \geq 0, z_i \geq 0$$
$$s_j \geq 0, \tau_i \geq 0$$

Note that we assume that prices of inputs are unchanged only *relative to output.* This helps in two ways: First, it enables us to discuss the issue of aggregation in Section 3. And second, it leads to a stationary solution, which shortens the proofs without losing any insight.

3 Aggregation of Inputs

In general, inputs x and z cannot be aggregated into one capital good. However, this can be done for a special case of perfectly synchronized upgrading, and this section provides the necessary and sufficient conditions for input aggregation.

Let $\tilde{z}(t)$ be the time path for the level (i.e., potential productivity) of input z

$$\tilde{z}(t) = z_i,\ t \in [\tau_i, \tau_{i+1}).$$

Note that for any level time path $\tilde{z}(t)$ the productivity time path $z(t)$ is uniquely defined by

$$z(t) = \delta(t)\tilde{z}(t),$$

$$\delta(t) = \begin{cases} \theta, & t \in [\tau_i, \tau_i + T) \\ 1, & t \in [\tau_i + T, \tau_{i+1}) \end{cases}, i \geq 0$$

(if $\theta = 1$ or $T = 0$, the two are identical), so that the pair $(x(t), \tilde{z}(t))$ fully determines the producer's policy. For a particular policy, the time path for the cost is given by

$$C(t) = \begin{cases} p_x x_j, & t = s_j \text{ and } \forall i\ s_j \neq \tau_i \\ p_z z_i, & t = \tau_i \text{ and } \forall j\ \tau_i \neq s_j \\ p_x x_j + p_z z_i, & t = \tau_i \text{ and } \exists j : \tau_i = s_j \\ 0, & \text{otherwise} \end{cases}$$

Definition: We will say that the inputs x and z are *aggregable* with respect to the policy $(x(t), \tilde{z}(t))$ if there exists a price index $P = P(p_x, p_z)$, a quantity index $X(t) = G(x(t), \tilde{z}(t))$ and a production function $F(X(t), t)$, such that the resulting time paths for cost and output, expressed through $(x(t), \tilde{z}(t))$ and through $X(t)$, are identical. That is to say, for every t

$$C(t) = P(p_x, p_z)X(t)$$
$$y(t) = F(X(t), t).$$

Three remarks may be made about this definition:

(a) It is motivated by Varian (1992), where the counterpart of our policy $(x(t), \tilde{z}(t))$ (i.e. the firm's demand for *intermediate* goods) is consumers' demand for *final* goods.

(b) This is a relatively weak concept of aggregability, because the functions P, G and F are allowed to depend on a particular policy. And yet we will show that even such a weak condition will generally not be met in our model.

(c) By allowing the production function F to depend on time, we allow the possibility of learning the aggregated input X.

Proposition 1: Inputs x and z are aggregable with respect to the policy $(x(t), \tilde{z}(t))$, if and only if for every $i \geq 0$ $s_i = \tau_i$ and $x_i = \alpha z_i$ for some positive constant α.

This states that we can aggregate if and only if all upgrades are *perfectly synchronized* and the levels of inputs are *proportional*.

Proof: Since we can arbitrarily rescale the units of measurement, it suffices to prove the statement for $\alpha = 1$, which is assumed to be the case without loss of generality. First, observe that when x and z are both constant, X must also be constant. Therefore, X should change only when at least one of the inputs is purchased. That is to say that

$$X(t) = G\left(x(t), \tilde{z}(t)\right) = X_k, \ t \in \left[t_k, t_{k+1}\right), k \geq 0$$

where t_k is the time, when at least one of the inputs is purchased, defined recursively by

$$\begin{aligned} t_0 &= \min\left\{s_0, \tau_0\right\}, \\ t_{k+1} &= \min\left\{\min_j \left\{s_j > t_k\right\}, \min_i \left\{\tau_i > t_k\right\}\right\}. \end{aligned}$$

Then, the cost time path for input X must be given by

$$C(t) = \begin{cases} PX_k, & t = t_k \\ 0, & \text{otherwise} \end{cases}.$$

Sufficiency: Assume that for every $i \geq 0$ $s_i = \tau_i$ and $x_i = z_i$. Then, set $G\left(x(t), \tilde{z}(t)\right) = \tilde{z}(t)$, $P = p_x + p_z$ and $F(X,t) = \delta(t)X$. Checking that output and cost paths for $X(t)$ and $(x(t), \tilde{z}(t))$ are identical is now straightforward.

Necessity: Assume first that the inputs are not always upgraded in synch, and when they are, the levels are not equal. That is to say that there exist i and j, such that $s_j < \tau_i$ and there are no upgrades in between. Set $t_k = s_j$. It follows that

$$\begin{aligned} X_k &= X(t_k) = G\left(x_j, z_{i-1}\right), \\ C(t_k) &= C(s_j) = p_x x_j = PX_k \end{aligned}$$

But then, we have

$$p_x x_j = G\left(x_j, z_{i-1}\right),$$

which means that either $G(x,z)$ does not depend on z or $x_j = z_{i-1}$. This can be expressed as $p_x x_j = P\tilde{G}(x_j)$. Since there are no upgrades between s_j and τ_i, $t_{k+1} = \tau_i$. It follows that

$$X_{k+1} = \begin{cases} \tilde{G}(x_{j+1}), & \text{if } s_{j+1} = \tau_i \\ \tilde{G}(x_j), & \text{if } s_{j+1} > \tau_i \end{cases}$$

$$C(t_{k+1}) = C(\tau_i) = PX_{k+1} = \begin{cases} p_x x_{j-1} + p_z z_i, & \text{if } s_{j+1} = \tau_i \\ p_z z_i, & \text{if } s_{j+1} > \tau_i \end{cases}$$

For x and z to be aggregable, it is necessary that either $p_x x_{j+1} + p_z z_i = P\tilde{G}(x_{j+1})$ or $p_z z_i = P\tilde{G}(x_j)$. The first equality cannot hold, because $x_{j+1} \neq z_i$. Then the second equality must hold, but if it does,this implies $p_z z_i = P\tilde{G}(x_j) = p_x x_j$, which means that such P does not exist. It is left to consider the case, when upgrades are synchronous, but the levels are not equal. The proof directly follows from the above argument. **Q. E. D.**

Evidently, then, we can not aggregate most of the policies, which strongly suggests that the two input case has possibilities that are absent in the one input case.

4 "S-policies" vs "A-policies"

This section contrasts synchronous (S) and asynchronous (A) upgrading behavior. When the following two assumptions are met, synchronous behavior is not optimal.

Assumption 1

$$p_z \geq \frac{\theta}{r}\left(1 - e^{-rT}\right).$$

This assumption says that the revenue during the learning period is too low to cover the purchase cost of input z.

Assumption 2

$$\theta + e^{-rT} \leq 1.$$

This assumption says that learning effects are important (low θ, big T).

Consider the set S of policies where x and z are upgraded together every time. That is,

$$S = \{x(t), z(t) : \forall i \in \mathrm{N}\ s_i = \tau_i\}.$$

These policies describe the most extreme form of synchronous behavior. Define *the S-policy* as the optimal policy in the set S. The S-policy is fully characterized by the following result.

4.1 Inferiority of the S-policy

Lemma 1 Let the initial level of inputs at time zero be $x(0) = \hat{x}$, $z(0) = \hat{z}$ and let Assumption 1 hold. Then, beginning from time $\tau_0 \geq 0$, the S-policy has both inputs upgraded periodically, with a constant period τ. The level of both inputs at the time of upgrading is equal to the frontier at that time. That is, for every $i \geq 0$, $s_i = \tau_i = \tau_0 + \tau i$, $z_i = x_i = Z(\tau_0 + \tau i)$. The value of the S-policy is

$$V(\cdot) = V_0(\hat{z}, \hat{x}, \tau_0) + e^{-(r-g)\tau_0} V_S(\tau) =$$

$$= \frac{\min\{\hat{x}, \hat{z}\}}{r}\left(1 - e^{-r\tau_0}\right) + e^{-(r-g)\tau_0} Z(0) \frac{\frac{\theta}{r}\left(1 - e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\tau}\right) - p_z - p_x}{1 - e^{-(r-g)\tau}}$$

Proof: In the Appendix.

Theorem 1 Under Assumptions 1, 2 any policy $(x(t), z(t)) \in S$ is strictly dominated from any initial conditions.

Proof: To prove the theorem it is enough to show that the S-policy is strictly dominated beginning from time τ_0. Then, without loss of generality, assume for the rest of the proof that $\tau_0 = 0$.[2]

Suppose that $\theta e^{g\tau} > 1$. That is to say that a brand-new z technology immediately becomes more productive than the new one. Consider an alternative policy, which still has $z_i = Z(\tau i) = Z(0)e^{g\tau i}$, for every i. Set $x(t)$ to be identically equal to $z(t)$. That is to say that

$$x(t) = \begin{cases} \theta z_i & t \in [\tau i, \tau i + T) \\ z_i & t \in [\tau i + T, \tau(i+1)) \end{cases} \quad i \in \mathbf{N}.$$

This policy has the same output, but lower cost. Its value for any fixed τ is given by

$$V(\tau) = Z(0) \frac{\frac{\theta}{r}\left(1 - e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\tau}\right) - p_z - p_x(\theta + e^{-rT})}{1 - e^{-(r-g)\tau}} > V_S(\tau).$$

Consider the case, when $\theta e^{g\tau} \leq 1$. This is the case, when a brand-new z technology is initially less productive than the old one it has replaced. Define another alternative policy, which we will call *the A-policy* (where "A" stands for "asynchronous") by setting $z_i = Z(\tau i) = Z(0)e^{g\tau i}$ and

$$x(t) = \begin{cases} \theta z_0 & t \in [0, T) \\ z_i & t \in [\tau i + T, \tau(i+1) + T), i \geq 0 \end{cases}.$$

This policy also gives the same output as the S-policy, but is less costly. Its value is

$$V_A(\tau) = Z(0) \frac{\frac{\theta}{r}\left(1 - e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\tau}\right) - p_x(\theta + e^{-rT}) + p_x\theta e^{-(r-g)\tau} - p_z}{1 - e^{-(r-g)\tau}} > V_S(\tau).$$

Since the S-policy is always dominated, so is any policy in S. **Q. E. D.**

Theorem 1 applies if θ is sufficiently small, or if T is sufficiently large. That is, if the effect of learning on productivity is sufficiently strong, upgrading both inputs at the same

[2] Given the same initial conditions, an alternative policy can mimic the S-policy for $t \in [0, \tau_0)$, and yield strictly bigger profit on $[\tau_0, \infty)$. Since τ_0 is finite, as shown in proof of Lemma 1, this is sufficient for the S-policy to be strictly dominated.

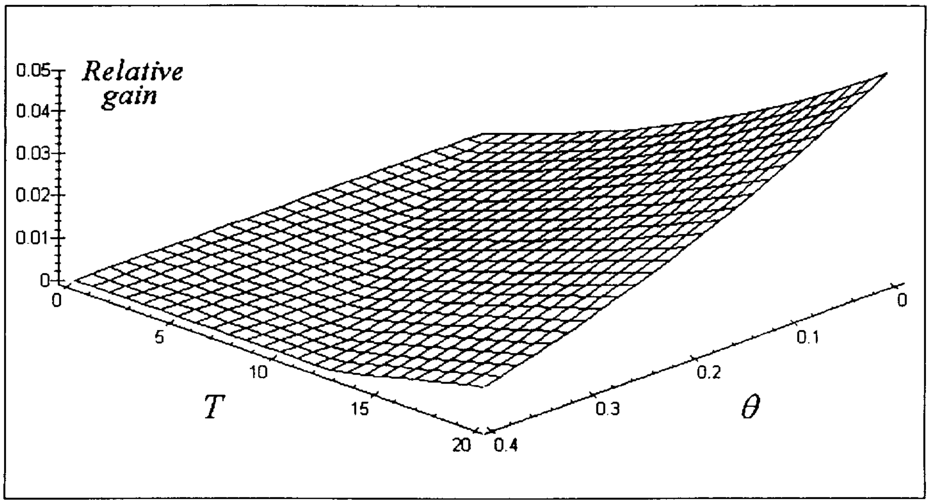

Figure 2: Relative gain from use of the A-policy instead of the S-policy

time every time is not optimal. When $\theta e^{g\tau} > 1$, the dominant policy has perfect comovement between the inputs, and seemingly argues in favor of total synchronization. However, the initial productivity of the brand-new technology need not exceed that of the one it has replaced. That is, if every upgrade of z involves a drop in output, the condition $\theta e^{g\tau} < 1$ is satisfied, implying that the A-policy is dominant. As we will show in the following section, when θ is small, the producer's optimal choice will be to incur the drop in output. Indeed, with small θ it takes bigger values of τ to satisfy $\theta e^{g\tau} > 1$, and if it does not pay to delay the growth (i.e. increase τ) for that long, then the A-policy dominates both the S-policy and the policy with perfect comovement.

4.2 The Gain from the A-policy

To measure the gain from the A-policy relative to the S-policy, we can use the difference in total costs relative to the (same for both policies) lifetime output. Let C_S and C_A be the total cost for the S-policy and the A-policy, respectively, and Y be lifetime output (the same for them both). Then, we have the following result:

Proposition 2

$$\frac{C_S - C_A}{Y} > \frac{rp_x\left(1 - e^{-rT} - \theta\right)}{\theta\left(1 - e^{-rT}\right) + e^{-rT}}$$

Proof: In the Appendix.

Figure 2 plots on its vertical axis the fraction of output saved from using the A policy. This is just the left-hand side of the inequality in Proposition 2. The figure is drawn an the assumption that $r = 0.04$, and $p_x = 1$. Amortized, the cost of x is then equal to r % of sales, which, if anything, is on the low side given that the share of capital is some 25% of sales.

The gain is positive when Assumption 2 is met. The gain is proportional to p_x, because the A-policy saves on the cost of input x. But the gain depends also on how much learning matters, i.e., θ, and T. Generally, T often *is* big and θ often *is* small: Bakh and Gort (1993, p. 577) find that the productivity of capital grows significantly until the fifth or sixth year after the birth of a plant. Now let us see what examples (a) and (b) of the introduction indicate about the possible magnitude of this gain:

(a) According to Toyota executives at the Georgetown, Kentucky plant (Womack 1990), T is about 10 years. If U.S. plants make more models (higher x) before they have mastered the new lean manufacturing technique (z), they will lose money.

(b) Mechanical TV was abandoned by 1933, whereas electronic TV was commercially introduced in the 1940's, T was about 10 years. In the interim, there was no TV broadcasting, so θ must have been small. (Of course, had there been no war, and no depression, T may have been lower). If TV stations had invested in programming quality (higher x) before high definition electronic TV (z) was fully developed, they would have lost money.

These examples show that when the new technology is complex, T is likely to be large, and θ is likely to be small, so that the relative gain from using the A policy will be significant.

4.3 Investment Patterns

Under the S-policy investment is (i) more lumpy, because both inputs are purchased at the same time, and (ii) negatively correlated with output, since when investment is made, output is below its mean. Under the A-policy, investment may be positively or negatively correlated with output, depending on prices of inputs. Under the A- and S- policies, $y(t) = z(t)$. Detrended output is given by

$$y'(t+\tau i) = z'(t+\tau i) = \frac{z(t+\tau i)}{Z(\tau i)} = \begin{cases} \theta, & 0 \le t \le T \\ 1, & T \le t \le \tau \end{cases}, i \in \mathbf{N}.$$

For the S-policy, detrended investment is

$$i_S(t) = \begin{cases} p_x + p_z, & t = \tau i \\ 0, & t \neq \tau i \end{cases}, i \in \mathbf{N}.$$

Under the A-policy it is

$$i_A(t) = \begin{cases} p_z + \theta p_x, & t = 0 \\ p_x, & t = \tau i + T \\ p_z, & t = \tau(i+1) \\ 0, & \text{otherwise} \end{cases}, i \in \mathbf{N}$$

Remark: Investment is more spiked under the S-policy. Clearly, $i_S(t)$ is more spiked, since there the producer pays for two inputs at a time.

In the S-policy all investment occurs at the point in time when the firm's post-investment output is below its trend. As a result, $i_S(t)$ is negatively correlated with output. On the other hand, $i_A(t)$ can be either positively or negatively correlated with output. This is the content of the next proposition.

Proposition 3

(i) Synchronous investment is negatively correlated with output and $cov(y', i_S) = -(1-\theta)(p_x + p_z)\left(1 - \frac{T}{\tau}\right) < 0$.

(ii) Asynchronous investment's correlation with output depends on relative prices of inputs and $cov(y', i_A) = (1-\theta)\left(p_x \frac{T}{\tau} - p_z\left(1 - \frac{T}{\tau}\right)\right)$.

Proof: In the Appendix.

These results are intuitive since investment leads output by more when we use the S-policy. Let us compare the results of this proposition to related ones in the literature.

Since Zeckhauser's single capital good also implies switching when the post-investment output is below trend, part 1 of the proposition merely replicates his one-input case, and is really implied by the aggregation result of section 4.

Caballero and Hammour (1996) also generate technological switching that is negatively correlated with productivity. But the reason is different. In their model, a technological switch must be accompanied by intervening unemployment of labor. Firms want to switch when productivity is low because the opportunity cost of the unemployed resources is then the smallest.

Klenow (1993) generates switching that is *positively* correlated with output volume in a one-capital good model because he assumes learning that is a function not of time, but of cumulative output. Firms like to switch technologies when the demand shock is high because high demand affords faster learning.

Benhabib and Nishimura (1985) get an asynchronous activity of sectors in an optimal two-sector growth model. This corresponds to our A-policy under which investment alternates between the two inputs.

5 Optimal Policies

Section 4 has contrasted A-policies and S-policies, and shown that when learning is sufficiently important and when upgrades in z make output drop A-policies do better than S-policies. However, there may be some other policy which is even better than the A-policy. To determine whether this is true, we will state conditions on the primitives of the model, under which the A-policy is globally optimal.

Without any additional statements, we can provide a simple intuitive explanation of why synchronous upgrades cannot be optimal when θ, the fraction of z utilized during the learning period, is small. Consider the extreme case, when $\theta = 0$. Since $\theta = 0$, the output during $[\tau_i, \tau_i + T)$ is equal to zero, regardless of the level of x and z. Then if we were to purchase x at the same time as z but before the firm has learned to use z, the x would sit idle throughout the learning period. That is, by purchasing x_j at time $\tau_i + T$ instead of τ_i, we delay the cost of x_j leaving output unchanged, thereby increasing profit. This shows that when $\theta = 0$, the optimal solution will not involve synchronous upgrading.

To prove the global optimality of the A-policy, we need additional assumptions. In particular, assume that θ is small, so every upgrade in z involves a drop in output. That is to say that in the optimum for every i

$$\frac{z_i}{z_{i+1}} > \theta \tag{2}$$

Although this condition involves endogenous variables, it will automatically hold in the optimum for a subset of primitives, which will be defined later.

Assumption 3

$$e^{-gT} \leq \frac{1}{2}.$$

This assumption is made to ensure that the growth rate is sufficiently large, enough that it does not pay to keep inputs constant for a long time.[3] Then we have the following result on relative frequency of upgrades:

Lemma 2 Let $x(t)$, $z(t)$ be the optimal solution to (1). Then, under Assumptions 1, 3 and condition (2), on any interval between two successive upgrades of x, there is at most one upgrade of z. That is to say, for every j there exists at most one i, such that

$$s_j < \tau_i \leq s_{j+1}. \tag{3}$$

Proof: In the Appendix.

5.1 Optimality of the A-policy

Finally, using the result of Lemma 2, we can provide sufficient conditions for the global optimality of the A-policy, given by

Theorem 2 Let Ω be the set of all $(g, r, \theta, T, p_x, p_z)$ satisfying

(i). $e^{-gT} \leq \frac{1}{2}$

(ii). $0 \leq p_x \leq \frac{1}{r}\left(1 - e^{-rT}\right)$

(iii). $0 \leq \theta < e^{-2rT} / \left(1 - e^{-rT} + \frac{r}{r-g}\right)$

(iv). $0 \leq p_z - \frac{\theta}{r}\left(1 - e^{-rT}\right) \leq \frac{1}{r}e^{-rT} - p_x(\theta + e^{-rT}) - \frac{\theta}{r-g}$.

Then, for any $(g, r, \theta, T, p_x, p_z) \in \Omega$ the A-policy is the optimal solution to (1) with initial conditions $x(0) = 0$, $z(0) = 0$.

Proof: First note that any vector of parameters from Ω satisfies Assumptions 1-3, so we can apply all the results proved above (condition (2) will be satisfied automatically for

[3] This assumption can be relaxed. See Section 6.

any point in Ω). Let $\tau_0 = \min\{t : z(t) > 0\}$ be the time of the first purchase of z. If τ_0 is infinite, the lifetime profit must be equal to zero. Since we assume (and further on, check) that positive profit is feasible, this cannot happen. Then, it is optimal to set $\tau_0 = 0$. As before, let $\{\tau_i\}$ be the sequence of times, when new z's are purchased. If $x(t) = 0$ for $t \in [0, \tau_1)$, then the profit on this interval is strictly negative (output is zero, and cost is $p_z z_0 > 0$), which cannot happen in the optimum. Then x must either be purchased at time 0, or at time T, or both.

First, prove that x is always purchased at time T. If not, it must be purchased at time 0, i.e. $s_0 = 0$. By Lemma 2, $s_1 \in \{\tau_1, \tau_1 + T\}$. Under condition (2), $\max_{[0,\tau_1)} z(\cdot) = \max_{[0,\tau_1+T)} z(\cdot) = z_0 \leq Z(0)$. Therefore, by Claim 2, it must be true that $x_0 \leq Z(0)$. Since $x(t)$ is constant on $[0, \tau_1)$, Claim 3 gives $x_0 = z_0$. Note that there are no gains from purchasing x at time τ_1, which is why $s_1 = \tau_1 + T$. Since $\theta + e^{-rT} < 1$ (Assumption 2), it is always less costly to purchase $x_0 = \theta z_0$ at time 0 and $x_1 = z_0$ at time T.

Since x is always purchased at time T, then any x_0 purchased at time $s_0 = 0$, will be used on $[0, T)$ only. From Claim 2 it follows that $x_0 = 0$ or $x_0 = \theta z_0$. Let x_1 be the level of x purchased at time T, and let $s_1 = T$. By Lemma 2, $s_2 \in \{\tau_1, \tau_1 + T\}$. Since it must also be true that $x_1 \leq z_0$, the profit on $[0, \tau_1)$ is given by

$$\max\left\{0, \frac{\theta z_0}{r}\left(1 - e^{-rT} - rp_x\right)\right\} - p_z z_0 + \frac{x_1}{r}\left(e^{-rT} - e^{-r\tau_1} - rp_x e^{-rT}\right) =$$

$$= -\theta z_0 p_x + \frac{z_0}{r}\left(\theta\left(1 - e^{-rT}\right) - rp_z\right) + \frac{x_1}{r}\left(e^{-rT} - e^{-r\tau_1} - rp_x e^{-rT}\right).$$

The last equality here is implied by (ii). Because the profit on $[0, \tau_1)$ must be positive (otherwise, $z_0 = 0$), the second term in the above expression is strictly positive, which gives $x_1 = z_0$. Since $z_0 = \max_{[T,\tau_1+T)} z(\cdot)$, there are no gains from purchasing x at time τ_1, so $s_2 = \tau_1 + T$. Repeating the above argument on the interval $[\tau_i, \tau_{i+1})$, we get the following expression for the profit on this interval:

$$e^{-r\tau_i}\left[\frac{z_i}{r}\left(\theta\left(1 - e^{-rT}\right) - rp_z\right) + \frac{x_{i+1}}{r}\left(e^{-rT} - e^{-r\Delta\tau_i} - rp_x e^{-rT}\right)\right].$$

Similarly, we get that $x_{i+1} = z_i$ and $s_{i+1} = \tau_i + T$ for every $i \geq 0$. Summing over i, we obtain the expression for the lifetime profit:

$$V(\cdot) = -\theta p_x + \sum_{i=0}^{\infty} e^{-r\tau_i}\frac{z_i}{r}\left[\theta\left(1 - e^{-rT}\right) - rp_z + \left(e^{-rT} - e^{-r\Delta\tau_i} - rp_x e^{-rT}\right)\right].$$

In the optimum every term of this sum must be strictly positive, which implies that $z_i = Z(\tau_i) = Z(0)e^{g\tau_i}$ for every i. This allows the lifetime profit to be rewritten as

$$V(\cdot) = -Z(0)\theta p_x + Z(0)\sum_{i=0}^{\infty} e^{-(r-g)\tau_i}\frac{1}{r}\left[\theta\left(1 - e^{-rT}\right) - rp_z + \left(e^{-rT} - e^{-r\Delta\tau_i} - rp_x e^{-rT}\right)\right]$$

Using the same argument as in Lemma 1, we can now show that the maximum value of lifetime profit will only depend on $\Delta\tau_0 \equiv \tau$. Therefore, the optimal policy is given by

$$z_i = Z(\tau i), \; t \in [\tau i, \tau(i+1)), i \geq 0$$

$$x(t) = \begin{cases} \theta z_0 & t \in [0, T) \\ z_i & t \in [\tau i + T, \tau(i+1) + T) \end{cases}, i \geq 0.$$

By definition, this is the A-policy.

We will now show that condition (2) is always satisfied. Since $z_i = Z(\tau i)$ and $z_{i+1} = Z(\tau(i+1))$ for every i, condition (2) is just

$$e^{-g\tau} > \theta,$$

which we must check for $\tau^* = \arg\max V_A(\tau)$. It is sufficient to prove that condition (2) is satisfied for some $\bar{\tau} > \tau^*$. The lifetime profit under the A-policy is given by:

$$V_A(\tau) = Z(0)\frac{\frac{\theta}{r}\left(1 - e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\tau}\right) - p_x e^{-rT} - p_z}{1 - e^{-(r-g)\tau}} - Z(0)\theta p_x$$

and

$$\frac{d}{d\tau}V_A(\tau) = \frac{Z(0)e^{-(r-g)\tau}}{1 - e^{-(r-g)\tau}}\left(e^{-g\tau} - \frac{V_A(\tau)}{Z(0)}(r-g) - (r-g)p_x\theta\right).$$

If we set $\bar{\tau}$ to satisfy

$$\exp(-g\bar{\tau}) = \frac{V_A(\bar{\tau})}{Z(0)}(r-g)$$

then the derivative $dV_A(\bar{\tau})/d\tau$ will be negative, implying that $\bar{\tau} > \tau^*$. Since $V_A(\tau)$ is decreasing for all $\tau > \tau^*$, it follows that

$$\exp(-g\bar{\tau}) = \frac{V_A(\bar{\tau})}{Z(0)}(r-g) > \lim_{\tau \to \infty} \frac{V_A(\tau)}{Z(0)}(r-g).$$

If the RHS of this expression is bigger than θ[4], condition (2) will hold for $\bar{\tau}$ as well as for τ^*. This can be rewritten as

$$\frac{1}{r}e^{-rT} - p_x(\theta + e^{-rT}) - \frac{\theta}{r-g} \geq p_z - \frac{\theta}{r}\left(1 - e^{-rT}\right),$$

which is the same as (iv).

It is left to see why the set Ω is not empty. Fix $r > 0$ and $g < r$. Next, fix $T > 0$, satisfying (i) for the given g. With r, g, T fixed, the RHS of inequalities (ii), (iii) are strictly positive numbers, so we can easily find r, g, T, θ, p_x, satisfying (i)-(iii). It is left to show that

[4] This also enough to ensure that the profit on every interval $[\tau i, \tau(i-1))$ is indeed strictly positive and that τ^* is finite.

(iv) defines a non-empty interval for p_z. Evaluating the RHS of (iv) at the biggest admissible value of p_x, and taking (iii) into account, we have:

$$\frac{1}{r}e^{-rT} - p_x(\theta + e^{-rT}) - \frac{\theta}{r-g} \geq \frac{1}{r}e^{-rT} - \frac{1}{r}\left(1 - e^{-rT}\right)(\theta + e^{-rT}) - \frac{\theta}{r-g} =$$

$$\frac{1}{r}\left(e^{-2rT} - \theta\left(1 - e^{-rT} + \frac{r}{r-g}\right)\right) > 0.$$

It follows that the interval for p_z is not empty, and thus Ω is not empty. **Q. E. D.**

5.2 Stability of the A-policy

We will now show that the A-policy is uniformly asymptotically stable.

Let $(x(t), z(t))$ be the optimal solution to (1) from the initial conditions $\hat{x} = x(0)$, $\hat{z} = z(0)$. We will say that the A-policy is *uniformly asymptotically stable*, if for any $(\hat{x}, \hat{z})$ there is a finite time $\hat{t}(\hat{x}, \hat{z})$, such that for every $t \geq \hat{t}(\hat{x}, \hat{z})$ $(x(t), z(t))$ follows the A-policy.

First, we prove the important corollary to Theorem 2.

Corollary: If $(x(t), z(t))$ is the optimal solution to (1) from the initial conditions $x(0) = \hat{x}$, $z(0) = \hat{z}$ and there exists $\hat{t}$, such that

1. $\hat{t} = \tau_i$ for some $i \geq 1$
2. $x(\hat{t}) = z_{i-1}$
3. $x(\hat{t}) > \theta Z(\hat{t})$,

then the optimal solution follows the A-policy beginning from time $\tau_i + T$.

Proof: Since $x(\tau_i) = z_{i-1}$ and $x(\tau_i) > \theta Z(\tau_i)$, there are no gains from purchasing x at time τ_i. Therefore, by Lemma 2, x must be purchased at time $\tau_i + T$. Now we can directly apply the argument presented in the proof of Theorem 2 to get

$$z(t) = \begin{cases} \theta Z(\tau_i + \tau k), & t \in [\tau_i + \tau k, \tau_i + \tau k + T)) \\ Z(\tau_i + \tau k), & t \in [\tau_i + \tau k + T, \tau_i + \tau(k+1)) \end{cases}, k \geq 0$$

$$x(t) = Z(\tau_i + \tau k),\ t \in [\tau_i + \tau k + T, \tau_i + \tau(k+1) + T), k \geq 0$$

By the same argument, the present value of the policy $(x(t), z(t))$ beginning at time τ_i is

$$e^{-(r-g)\tau_i} Z(0) \sum_{k=0}^{\infty} e^{-(r-g)\tau k} \left[\frac{\theta}{r}\left(1 - e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\tau}\right) - p_z - e^{-rT} p_x\right] =$$

$$= e^{-(r-g)\tau_i} Z(0) \left[\frac{V_A(\tau)}{Z(0)} + \theta p_x\right]$$

Q. E. D.

This result implies that if the above conditions 1-3 are satisfied, then the A-policy is stable.

Proposition 4: If conditions (i)-(iv) of Theorem 2 are satisfied, the A-policy is stable.

Proof: The idea of the proof is to show that for any initial conditions the optimal solution must reach the point $\hat{t}$, where conditions 1-3 of the Corollary are satisfied. Condition 1 is satisfied if z is purchased at least twice.[5] Suppose that condition 2 is not satisfied. This means that for every i either $x(\tau_i) < z_{i-1}$ or $x(\tau_i) > z_{i-1}$. If the former inequality holds, then purchasing z at time τ_i is suboptimal. If the latter inequality holds for every i, take $\tau_i > s_0$[6] and observe that Claim 2 is not satisfied. Therefore, there exists a finite time τ_i, such that $x(\tau_i) = z_{i-1}$.Finally, condition 3 holds automatically in the optimum. Indeed, if it holds, the lifetime profit can be expressed as

$$V(\cdot) = \int_0^{\tau_i} e^{-rt} \min\{x(t), z(t)\}\, dt + e^{-(r-g)\tau_i} Z(0) \left[\frac{V_A(\tau)}{Z(0)} + \theta p_x\right].$$

The F.O.C. for maximization with respect to τ_i is

$$x(\tau_i) = e^{g\tau_i} Z(0) \left[\frac{V_A(\tau)}{Z(0)} + \theta p_x\right](r-g).$$

The RHS can be rewritten as

$$e^{g\tau_i} Z(0) \left[\frac{V_A(\tau)}{Z(0)} + \theta p_x\right](r-g) = Z(\tau_i) \left[\frac{V_A(\tau)}{Z(0)} + \theta p_x\right](r-g) = Z(\tau_i) e^{-g\tau} > \theta Z(\tau_i),$$

where the last equality follows from the F.O.C. for maximization with respect to τ, presented in the proof of Theorem 2. This shows that in the optimum $x(\tau_i) > \theta Z(\tau_i)$. **Q. E. D.**

6 Robustness of Results

6.1 Incremental Upgrades

Here we will focus on the case of incremental upgrades of input x only, because doing so for input z would entail an entirely new specification of the learning process.

Allowing input x to be upgraded incrementally enables us to make a stronger case for the A-policy in Theorem 1. Indeed, with incremental upgrades τ will fall,[7] and, all other things being equal, the inequality $\theta e^{g\tau} \leq 1$ will be easier to satisfy, making the A-policy the dominant one for a bigger set of parameters.

[5] It is shown in the proof of Lemma 1 that for any initial conditions z is purchased at least once, at time τ_0. Repeating the same argument for $\hat{x} = x(\tau_0)$ and $\hat{z} = z(\tau_0)$, it can be shown that z is purchased at least twice, and, in fact, countably many times.

[6] If s_0 is infinite, then $x(t) \equiv \hat{x}$, and by Claim 3, the inequality will be voilated beginning from some finite time.

[7] Incremental upgrades have the same effect as a decrease in p_x, since now the input is not scrapped after an upgrade. Quite intuitively, the decrease in p_x will make the upgrades more frequent. Note that the

Incremental upgrades also enable us to interpret the optimality result of Section 5 more broadly. Indeed, it may seem that Assumption 3 is overly restrictive, and the values of T satisfying it are so big that the optimality result cannot apply on the level of the individual firm. However, with incremental upgrades we no longer need Assumption 3 to prove Lemma 2. Intuitively, the producer now can make upgrades more frequently without "repurchasing" the same level of input over and over again. Since he is interested in delaying the costs, he has nothing to gain from creating the idle stock of input x to be used later, which implies that he follows the upgrading pattern described in Lemma 2. Moreover, Assumption 3 is not needed for the results in section 4 on the superiority of the A-policy over the S-policy.

6.2 The two-output model

This subsection modifies the two-input model. To explain the phenomena in examples (d) and (e) of the introduction, we now transform it into a *two output* model. The products will be complements in the firm's profit function.

There are two outputs with quality levels x and z, which can be adjusted at any time. Only one quality level of each output can be produced. The producer can choose to make any quality of output x. More formally, suppose that producer switches to new x at times s_j, $j = 0, 1, ...$, and the corresponding quality is x_j. Likewise, z is adjusted at times τ_i and

---

optimal frequency of upgrading is implicitly defined by

$$\frac{dV_A}{d\tau} = 0$$

which is equivalent to

$$e^{-g\tau} - \left(\frac{V_A(\tau)}{Z(0)} + \theta p_x\right)(r-g) = 0$$

Performing comparative statics on this equation is quite simple:

$$\frac{d\tau}{dp_x} = \frac{1}{g}e^{g\tau}\left(-\frac{1}{Z(0)}\frac{\partial V_A}{\partial p_x} - \theta\right)(r-g) = \frac{1}{g}e^{g\tau}\left(\theta + \frac{e^{-rT}}{\left(1-e^{-(r-g)\tau}\right)} - \theta\right)(r-g) > 0$$

$$\frac{d\tau}{dp_z} = \frac{1}{g}e^{g\tau}\left(-\frac{1}{Z(0)}\frac{\partial V_A}{\partial p_z}\right)(r-g) = \frac{1}{g}e^{g\tau}\left(\frac{1}{\left(1-e^{-(r-g)\tau}\right)}\right)(r-g) > 0$$

So, as inputs get cheaper, upgrades become more frequent.

$$\frac{d\tau}{d\theta} = \frac{1}{g}e^{g\tau}\left(-\frac{1}{Z(0)}\frac{\partial V_A}{\partial \theta} - p_x\right)(r-g) = \frac{1}{g}e^{g\tau}\left(p_x - \frac{\left(1-e^{-rT}\right)}{r\left(1-e^{-(r-g)\tau}\right)} - p_x\right)(r-g) < 0$$

$$\frac{d\tau}{dT} = \frac{1}{g}e^{g\tau}\left(-\frac{1}{Z(0)}\frac{\partial V_A}{\partial T}\right)(r-g) = \frac{1}{g}e^{g\tau}\frac{-\theta e^{-rT} + e^{-rT} - rp_x e^{-rT}}{1-e^{-(r-g)\tau}}(r-g) >$$

$$> \frac{1}{g}e^{g\tau}\frac{e^{-rT}\left(1-e^{-rT}-\theta+e^{-r\tau}\right)}{1-e^{-(r-g)\tau}}(r-g) > 0$$

Also, when learning is easier (bigger θ, smaller T) upgrades become more frequent.

the quality is z_i. Then the time paths for $x(t)$ and $z(t)$ will be fully determined by the sequences of (nonnegative) numbers$\{x_j, s_j\}_{j=0}^{\infty}$ and $\{z_i, \tau_i\}_{i=0}^{\infty}$ respectively:

$$x(t) = x_j, t \in [s_j, s_{j+1}), j \in \mathbf{N},$$

$$z(t) = \begin{cases} \theta z_i & t \in [\tau_i, \tau_i + T) \\ z_i & t \in [\tau_i + T, \tau_{i+1}) \end{cases} i \in \mathbf{N},$$

where $z_i \leq Z(\tau_i) = Z(0)e^{g\tau_i}$ for every i.

As before, the quality of z is constrained by the frontier $Z(t) = Z(0)e^{gt}$ and for T periods after each upgrade of z, only a fraction $\theta \in [0, 1]$ of its quality can be used in production, because of learning.

The cost of quality adjustment is asymmetric. Adjusting the quality of z is always costly. It costs $c_z z_i$ to switch to quality level z_i. In contrast, quality of x can be costlessly adjusted to any level at or below the current quality of the other output $z(t)$. But if the producer chooses a quality level of x above $z(t)$, he pays c_x per each additional unit. That is, the cost function for x is given by

$$C(x_j, s_j) = c_x \max\{0, x_j - z(s_j)\}.$$

The motivation for this cost function is that the two products share the production capacity. For example, once the power plant is built for providing street lighting at night, capacity is automatically created for industrial lighting during the day. Or, once the railway line is built for transporting passengers during the day, capacity is automatically created for hauling freight at night. The form of this cost function is the reason why the two products are complements in the profit function. In terms of quantity of output, production capacity is restricted to one unit of each output per unit of time. So for example, there is a single railway line and one train, and the speed of service represents quality. The capacity is shared, because a better engine, say, will haul transport passengers faster during the day, and it will haul freight faster at night. This is the analog of the assumption that only one unit of each input can be used in the two-input model, but that the quality of each input is variable.

Let p_x and p_z be producer's markup above marginal costs of x and z, respectively. Then, lifetime profit is given by

$$V(\{x_j, s_j\}, \{z_i, \tau_i\}) = \int_0^{\infty} e^{-rt} (p_x x(t) + p_z z(t))\, dt - \sum_{i=0}^{\infty} e^{-r\tau_i} c_z z_i - \sum_{j=0}^{\infty} e^{-rs_j} C(x_j, s_j) =$$

$$= \sum_{i=0}^{\infty} \left(\int_{\tau_i}^{\tau_{i+1}} e^{-rt} p_z z(t) dt - e^{-r\tau_i} c_z z_i \right) + \sum_{j=0}^{\infty} \left(\int_{s_j}^{s_{j+1}} e^{-rt} p_x x_j dt - e^{-rs_j} c_x \max\{0, x_j - z(s_j)\} \right).$$

Assume that $p_x < rc_x$. Then choosing $x_j > z(s_j)$ yields strictly negative profit. That is to say, adjustments of x will be made only when they are costless. A costless adjustment can be made only if $x_{j-1} < z(s_j)$. In the optimum, s_j must be the earliest time when this is possible, i.e.,

$$s_j = \min\{s : z(s) > x_{j-1}\}.$$

Quality increases are asynchronous: The optimal quality of x will be $x_j = z(s_j)$ for every j. Then, if $z(t)$ drops after each upgrade (the analog of condition 2 in Section 5), quality adjustments in x and z are necessarily asynchronous: It does not pay to adjust the quality of x when it is above that of z, but once z is fully learned, x is adjusted immediately. Formally,

$$s_0 = \tau_0, \; s_j = \tau_{j-1} + T, \; j \geq 1.$$

The rest of the problem is solved in the usual way: positive optimal profit implies that z will always be upgraded to the frontier, and upgrading frequency will be constant.

6.3 Changing Prices

Our model assumes that the price per efficiency unit of each input is constant relative to the price of the output. That is, the cost function is linear in x and z. If the relative input prices changed over time, we would not have a constant τ policy in the limit.

Two kinds of departures from the constant price assumption are possible. First, p_x and p_z can be declining functions of x and z, and indeed of time as well. In this case, the frequency of input purchases rises over time under both S- and A- policies. But producers still choose to upgrade z to the frontier.

Second, p_x and p_z can be *increasing* functions of x, z, and time. The dependence of prices on x and z means that the cost function becomes convex. In this case the results change a bit more. A producer now may not want to upgrade to the frontier, and the frontier ceases to be a binding constraint, which also happens in Parente's model.

If p_x and p_z depended on time but if their ratio remained fixed, our results on the comparison of A-policies and S-policies should remain basically unchanged.. But if the inputs prices were to vary relative to each other, synchronous policies would be less likely to be optimal.

7 Dynamic Programming Representation

The maximization problem can also be represented as a dynamic programming problem. The state space is:

a - the present age of input z

x - the present stock of input x-technology

z - the present stock of input z, $z \leq Z$

Z - the "base" frontier, i.e., the frontier at the time of the last upgrade in z. Z is not a function of time. The current frontier is given by Ze^{ga}.

The control variables are:

x', z' - future stocks of inputs

s - waiting time until the next upgrade.

If the amount z of input z is used at time t, its productivity is given by $z\delta(a+t)$, where $\delta(\cdot)$ represents the learning curve:

$$\delta(t) = \begin{cases} \theta, & t < T \\ 1, & t \geq T \end{cases} .$$

Let $V(a,x,z,Z)$ be the optimal value function. Given the states, one of the following happens after time s, whichever is more profitable:

1) x is upgraded, giving the continuation value of $e^{-rs}\left(V(a+s,x',z,Z)-p_x x'\right)$

2) z is upgraded, giving the continuation value of $e^{-rs}\left(V(0,x,z',Ze^{g(a+s)})-p_z z'\right)$

3) both x and z are upgraded, giving the continuation value of $e^{-rs}\left(V(0,x',z',Ze^{g(a+s)})-p_x x'-p_z z'\right)$.

Then, the Bellman equation is

$$V(a,x,z,Z) = \max\left\{\max_{x',s}\left(\int_0^s e^{-rt}\min\left[x,\delta(a+t)z\right]dt + e^{-rs}\left(V(a+s,x',z,Z)-p_x x'\right)\right);\right.$$

$$\max_{z',s}\left(\int_0^s e^{-rt}\min\left[x,\delta(a+t)z\right]dt + e^{-rs}\left(V(0,x,z',Ze^{g(a+s)})-p_z z'\right)\right);$$

$$\left.\max_{x',z',s}\left(\int_0^s e^{-rt}\min\left[x,\delta(a+t)z\right]dt + e^{-rs}\left(V(0,x',z',Ze^{g(a+s)})-p_x x'-p_z z'\right)\right)\right\}.$$

Value function V is homogenous of degree 1 in x,z,Z. Indeed, the operator in the RHS of the Bellman equation, maps linear homogenous functions into linear homogenous. If there is a fixed point of this operator (which is V), it should also be linear homogenous. Then we can define the reduced state space $\left(a,\frac{x}{Z},\frac{z}{Z}\right)=(a,u,w)$, with controls (s,u',w') and the optimal value function $v(a,u,w)$, given by

$$v(a,u,w) = \frac{1}{Z}V(a,x,z,Z) = V\left(a,\frac{x}{Z},\frac{z}{Z},1\right).$$

The Bellman equation for $v(a,u,w)$ will take the form

$$v(a,u,w) = \max\left\{\max_{u',s}\left(\int_0^s e^{-rt}\min\left[u,\delta(a+t)w\right]dt + e^{-rs}\left(v(a+s,u',w)-p_x u'\right)\right);\right.$$

$$\max_{w',s}\left(\int_0^s e^{-rt}\min\left[u,\delta(a+t)w\right]dt + e^{-rs}e^{g(a+s)}\left(v(0,ue^{-g(a+s)},w')-p_z w'\right)\right);$$

$$\left.\max_{u',w',s}\left(\int_0^s e^{-rt}\min\left[u,\delta(a+t)w\right]dt + e^{-rs}e^{g(a+s)}\left(v(0,u',w')-p_x u'-p_z w'\right)\right)\right\}$$

As shown in Corollary to Theorem 2, there is a subset A of the (reduced) state space and a subset of the set of parameters Ω, such that for $(a,u,w)\in A$ and $(g,r,\theta,T,p_x,p_z)\in\Omega$, all the upgrades are asynchronous. That is to say that upgrading x only or z only is always preferred to upgrading both inputs. Formally,

$$\max\left\{\max_{u',s}\left(\int_0^s e^{-rt}\min\left[u,\delta(a+t)w\right]dt+e^{-rs}\left(v(a+s,u',w)-p_x u'\right)\right);\right.$$

$$\left.\max_{w',s}\left(\int_0^s e^{-rt}\min\left[u,\delta(a+t)w\right]dt+e^{-rs}e^{g(a+s)}\left(v(0,ue^{-g(a+s)},w')-p_z w'\right)\right)\right\}\geq$$

$$\geq\max_{u',w',s}\left(\int_0^s e^{-rt}\min\left[u,\delta(a+t)w\right]dt+e^{-rs}e^{g(a+s)}\left(v(0,u',w')-p_x u'-p_z w'\right)\right)$$

$$\text{for } (a,u,w)\in A, (g,r,\theta,T,p_x,p_z)\in\Omega.$$

To characterize the set A, we must express conditions of the Corollary in terms of the state variables. Condition 1 translates into $a\geq T$,[8] condition 2 gives $u=w$, and condition 3 will hold automatically for all parameter values from Ω. Thus

$$A=\{(a,u,w):a\geq T,u=w\}.$$

As an example, consider value function representation of the A-policy. Let τ be the optimal waiting time until z is purchased, that is, the maximum age of z. Then, since the producer follows the A-policy, we can write

$$v(T,1,1)=\int_0^{\tau-T}e^{-rt}dt+e^{-r(\tau-T)}e^{g\tau}\left(v(0,e^{-g\tau},1)-p_z\right)$$

$$v(0,e^{-g\tau},1)=\int_0^T\theta e^{-rt}dt+e^{-rT}\left(v(T,1,1)-p_x\right).$$

Solving for $v(T,1,1)$, we have

$$v(T,1,1)=\frac{e^{rT}}{1-e^{-(r-g)\tau}}\left(\frac{1}{r}\left(e^{-rT}-e^{-r\tau}\right)+\frac{\theta e^{-(r-g)\tau}}{r}\left(1-e^{-rT}\right)-p_x e^{-rT}e^{-(r-g)\tau}-p_z e^{-(r-g)\tau}\right).$$

This, not surprisingly, is equal to the value of the A-policy beginning at time T.

8 Conclusion

We have analyzed processes that require two complementary inputs subject to improvements in quality. We have shown that if one of these inputs requires a period of learning after a quality upgrade before it can be used effectively, it can then pay to purchase the inputs at

[8]This is in fact stronger than condition 1. However, since x is not upgraded while z is constant, condition 1 is implied by $a\geq T$ and $u=w$.

different dates. The purchases are asynchronous because money is wasted when it is tied up in the other inputs which are then underutilized until the date learning is over.

The intuition is simple, and we believe that this logic sheds light on a number of phenomena, some of which we mentioned in the paper. In particular, we gave examples of asynchronous behavior in a variety of contexts and our analysis suggests that the learning hypothesis can help us understand such behavior, especially when the inputs or outputs in question are complementary.

On a formal level, we have extended Zeckhauser's model by assuming that a technology requires not one, but two inputs. We showed that in most of the cases the two inputs cannot be aggregated in one, so that the two input case is qualitatively different from the one input case, especially when purchases are asynchronous.

9 Appendix

The following claims provide necessary conditions for $(x(t), z(t))$ to be the optimal solution to the profit maximization problem.

Claim 1: If $x(t), z(t)$ is an optimal solution to the profit maximization problem, then $\{x_j\}_{j=0}^{\infty}$ and $\{z_i\}_{i=0}^{\infty}$ are strictly increasing sequences.

Proof: To establish that $\{x_j\}_{j=0}^{\infty}$ is a strictly increasing sequence, we must prove that for every j

$$x_j > x_{j-1}$$

Suppose this is not true. Then, if the producer leaves x at its previous level x_{j-1} on the interval $[s_j, s_{j+1})$, he thereby saves $e^{-rs_j}p_x x_j$ in cost, and there is no loss of output. Therefore, downgrading x cannot be optimal. The argument for $\{z_i\}_{i=0}^{\infty}$ is quite similar. **Q.E.D.**

Claim 2: If in the optimum x is upgraded at time s_j to the level $x_j > 0$, then it must be the case that

$$\min_{[s_j, s_{j+1})} z(\cdot) \leq x_j \leq \max_{[s_j, s_{j+1})} z(\cdot)$$

Proof: If the right inequality does not hold, then on $[s_j, s_{j+1})$ z is a limiting factor in production. Then if we purchase ε less of x, output remains the same, but the cost is $e^{-rs_j}p_x\varepsilon$ less and upgrading to x_j cannot be optimal. Suppose now the left inequality does not hold. Then, on $[s_j, s_{j+1})$ x is a limiting factor in production and, on this interval, profit is a linear function of x_i. Then, in the optimum $x_i = 0$, which contradicts the assumption of the claim, or for every $x_j < \min_{[s_j, s_{j+1})} z(\cdot)$ an ε increase in x gives strictly bigger profit, and therefore x_j cannot be optimal. **Q.E.D.**

Claim 3: If $x(t), z(t)$ is an optimal solution and $x(\tau_i) \geq Z(\tau_i)$, then $z_i = Z(\tau_i)$. If $x(t)$ is constant on $[\tau_i, \tau_{i+1})$, $x(\tau_i) = x \leq Z(\tau_i)$ and Assumption 1 holds, then $z_i = x$.

Proof: Since $x(\tau_i) \geq Z(\tau_i)$, $z_i \leq Z(\tau_i)$ and $x(t)$ is non-decreasing (Claim 1), the present value of profit on $[\tau_i, \tau_{i+1})$ is given by

$$z_i e^{-r\tau_i}\left(\frac{\theta}{r}(1-e^{-rT}) - p_z + \frac{1}{r}\left(e^{-rT} - e^{-r\left(\tau_{i+1}-\tau_i\right)}\right)\right)$$

If the term in brackets is negative, the profit on $[\tau_i, \tau_{i+1})$ is negative, and the producer can do strictly better by not purchasing z at time τ_i. If the term in brackets is positive, then it is optimal to set z_i to its maximum possible level, $z_i = Z(\tau_i)$.

If $x(t)$ is constant on $[\tau_i, \tau_{i+1})$, and $x(\tau_i) = x \leq Z(\tau_i)$, the present value of profit on $[\tau_i, \tau_{i+1})$ is given by

$$e^{-r\tau_i}\left(\frac{\min\{x, \theta z_i\}}{r}(1-e^{-rT}) - p_z z_i\right) + \frac{\min\{x, z_i\}}{r} e^{-r\tau_i}\left(e^{-rT} - e^{-r\left(\tau_{i+1}-\tau_i\right)}\right)$$

Under Assumption 1 the first term in this expression is always non-positive, so it does not pay to set z_i strictly above x. This leaves us with the case $z_i \leq x$, for which we can directly apply the first part of the proof. This gives $z_i = x$. **Q.E.D.**

Corollary: In the optimum, for every i, $\tau_{i+1} - \tau_i > T$.

Proof: If this is not the case, then, by Assumption 1, profit on $[\tau_i, \tau_{i+1})$ is non-positive, no matter what $x(t)$ is. **Q.E.D.**

Claim 4: In the optimum, x is not upgraded in the points, where $z(t)$ is constant. That is, for every j, $s_j = \tau_i$ or $s_j = \tau_i + T$, for some i.

Proof: Suppose this is not true. Take an arbitrary interval $[t_1, t_2)$, where $z(t) = z = const$. Suppose there is $s_j \in (t_1, t_2)$. Then, the present value of profit on $[t_1, t_2)$ will be given by

$$\frac{\min\{x_{j-1}, z\}}{r}\left(e^{-rt_1} - e^{-rs_j}\right) + \frac{\min\{x_j, z\}}{r}\left(e^{-rs_j} - e^{-rt_2}\right) - p_x x_j e^{-rs_j} =$$

$$= e^{-rs_j}\left(\frac{\min\{x_j, z\}}{r} - \frac{\min\{x_{j-1}, z\}}{r} - p_x x_j\right) + \frac{\min\{x_{j-1}, z\}}{r} e^{-rt_1} - \frac{\min\{x_j, z\}}{r} e^{-rt_2}$$

This is a monotonic function of s_j, which is why setting $s_j = t_1$ or $s_j = t_2$ is always preferred to any $s_j \in (t_1, t_2)$. To finish the proof, observe that when we exclude all the open intervals, where $z(t)$ is constant, this leaves us with the points $\{\tau_i, \tau_i + T\}_{i=0}^{\infty}$. **Q.E.D.**

Proof of Lemma 1: Let τ_0 be the time, when (both) inputs are purchased for the first time. Then, for $t \in [0, \tau_0)$, $x(t) = \hat{x}$, $z(t) = \hat{z}$. The special form of policies in the set S allows us to dramatically simplify the expression (1) for lifetime profit:

$$V(\hat{z}, \hat{x}, \{z_i, \tau_i\}, \{x_j, s_j\})\Big|_{(x,z)\in S} = V_0(\hat{z}, \hat{x}, \tau_0) + V_S(\{z_i, x_i, \tau_i\}) =$$

$$= \int_0^{\tau_0} e^{-rt}\min\{\hat{x}, \hat{z}\}dt + \sum_{i=0}^{\infty}\left(\int_{\tau_i}^{\tau_{i+1}} e^{-rt}\min\{x_i, z_i\}dt - e^{-r\tau_i}\left(p_z z_i + p_x x_i\right)\right)$$

By Claim 2, for every i it must be the case that $\theta z_i \leq x_i \leq z_i$, which allows us to rewrite the expression for V_S as

$$V_S(\{z_i, x_i, \tau_i\}) = \sum_{i=0}^{\infty} e^{-r\tau_i}\left(\frac{z_i}{r}\left(\theta\left(1-e^{-rT}\right) - rp_z\right) + \frac{x_i}{r}\left(e^{-rT} - e^{-r\left(\tau_{i+1}-\tau_i\right)} - rp_x\right)\right)$$

Assuming that positive lifetime profit is feasible, every term of this sum, representing the profit on $[\tau_i, \tau_{i+1})$, must be positive in the optimum, because otherwise $\{\tau_i\}$ cannot be the optimal sequence of upgrading times (see proof of Claim 3). Combined with Assumption 1, this gives

$$e^{-rT} - e^{-r\left(\tau_{i+1}-\tau_i\right)} - rp_x > 0$$

which immediately implies that $x_i = z_i$. Then, from Claim 3 we have that $x_i = z_i = Z(\tau_i) = Z(0)e^{g\tau_i}$ for every i. Without loss of generality, assume that $\tau_0 = 0$. Define $\Delta\tau_i = \tau_{i+1} - \tau_i$. Substituting this into the expression for V_S, we have

$$V_S(\{\Delta\tau_i\}_{i=0}^{\infty}) = Z(0)\sum_{i=0}^{\infty}\exp\left\{-(r-g)\sum_{k<i}\Delta\tau_k\right\}\left(\frac{\theta}{r}\left(1-e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\Delta\tau_i}\right) - p_z - p_x\right) =$$

$$Z(0)\left(\frac{\theta}{r}\left(1-e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\Delta\tau_0}\right) - p_z - p_x\right) + e^{-(r-g)\Delta\tau_0}V_S(\{\Delta\tau_i\}_{i=1}^{\infty})$$

Taking the maximum of both sides and cascading the maximization operator gives

$$\max_{\{\Delta\tau_i\}} V_S = \max_{\Delta\tau_0}\left[Z(0)\left(\frac{\theta}{r}\left(1-e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\Delta\tau_0}\right) - p_z - p_x\right) + e^{-(r-g)\Delta\tau_0}\max_{\{\Delta\tau_i\}} V_S\right]$$

The above expression shows that the maximum value of V_S will only depend on $\Delta\tau_0$. Thus, beginning from time τ_0, the S-policy will have upgrades of inputs happening periodically, with the constant period $\tau \equiv \Delta\tau_0$. Now getting the expression for the value of the S-policy is straightforward:

$$V(\cdot) = V_0(\hat{z}, \hat{x}, \tau_0) + e^{-(r-g)\tau_0}V_S(\tau) =$$

$$= \frac{\min\{\hat{x}, \hat{z}\}}{r}\left(1-e^{-r\tau_0}\right) + e^{-(r-g)\tau_0}Z(0)\frac{\frac{\theta}{r}\left(1-e^{-rT}\right) + \frac{1}{r}\left(e^{-rT} - e^{-r\tau}\right) - p_z - p_x}{1 - e^{-(r-g)\tau}}.$$

It is left to show that τ_0 is finite. Indeed, F.O.C. for maximization over τ_0 gives:

$$\frac{dV}{d\tau_0} = e^{-r\tau_0}\left(\min\{\hat{x}, \hat{z}\} - e^{g\tau_0}V_S(\tau)\right).$$

Since we assume that $V_S(\tau) > 0$, there exists a finite τ_0, such that the RHS is strictly negative, implying that the optimal τ_0 is finite. **Q. E. D.**

Proof of Proposition 2:

$$\frac{C_S - C_A}{Y} = \frac{C_S\left(1-e^{-rT}\right) - p_x\theta Z(0)}{Y} = \frac{\sum_{i=0}^{\infty} e^{-(r-g)\tau i} p_x Z(0)\left(1-e^{-rT}\right) - p_x\theta Z(0)}{\sum_{i=0}^{\infty} e^{-(r-g)\tau i} Z(0)\left(\frac{\theta}{r}\left(1-e^{-rT}\right) + \frac{1}{r}\left(e^{-rT}-e^{-r\tau}\right)\right)} =$$

$$= \frac{p_x\left(1-e^{-rT}\right) - p_x\theta(1-e^{-(r-g)\tau})}{\frac{\theta}{r}\left(1-e^{-rT}\right) + \frac{1}{r}\left(e^{-rT}-e^{-r\tau}\right)} > \frac{rp_x\left(1-e^{-rT}-\theta\right)}{\theta\left(1-e^{-rT}\right)+e^{-rT}} \textbf{Q. E. D.}$$

Proof of Proposition 3: Both $y'(t)$ and $i_S(t)$ are periodic functions with period τ, and $i_A(t)$ is also periodic for $t \geq T$. We can use periodicity to define the covariance between y' and i as

$$cov(y', i) = \int_0^{\tau} \left(y'(t) - \bar{y}\right)\left(i(t) - \bar{i}\right) dt,$$

where

$$\bar{y} = \frac{1}{\tau}\int_0^{\tau} y'(t)dt = \theta\frac{T}{\tau} + \left(1 - \frac{T}{\tau}\right),$$

$$\bar{i}_A = \bar{i}_S = \frac{p_x + p_z}{\tau},$$

and is given by

$$cov(y', i_S) = \int_0^{T} \left(\theta - \bar{y}\right)\left(i_S(t) - \bar{i}\right) dt + \int_T^{\tau} \left(1 - \bar{y}\right)\left(i_S(t) - \bar{i}\right) dt =$$

$$= \lim_{\Delta t \to 0}\left[\int_0^{\Delta t} \left(\theta - \bar{y}\right)\left(\frac{p_x + p_z}{\Delta t} - \bar{i}\right) dt - \int_{\Delta t}^{T} \left(\theta - \bar{y}\right)\cdot \bar{i} dt - \int_T^{\tau} \left(1 - \bar{y}\right)\cdot \bar{i} dt\right] =$$

$$= \left(\theta - \bar{y}\right)\left(p_x + p_z - \bar{i}\Delta t - \bar{i}T + \bar{i}\Delta t\right) - \bar{i}\left(1 - \bar{y}\right)\left(\tau - T\right) =$$

$$= -(1-\theta)(p_x + p_z)\left(1 - \frac{T}{\tau}\right)$$

$$cov(y', i_A) = \int_0^{T} \left(\theta - \bar{y}\right)\left(i_A(t) - \bar{i}\right) dt + \int_T^{\tau} \left(1 - \bar{y}\right)\left(i_A(t) - \bar{i}\right) dt =$$

$$\lim_{\Delta t \to 0}\left[\int_0^{\Delta t} \left(\theta - \bar{y}\right)\left(\frac{p_z}{\Delta t} - \bar{i}\right) dt - \int_{\Delta t}^{T} \left(\theta - \bar{y}\right)\bar{i} dt + \int_T^{T+\Delta t} \left(1 - \bar{y}\right)\left(\frac{p_x}{\Delta t} - \bar{i}\right) dt - \int_{T+\Delta t}^{\tau} \left(1 - \bar{y}\right)\bar{i} dt\right] =$$

$$= \left(\theta - \bar{y}\right)\left(p_z - \bar{i}T\right) + \left(1 - \bar{y}\right)\left(p_x - \bar{i}(\tau - T)\right) = (1-\theta)\left(p_x\frac{T}{\tau} - p_z\left(1 - \frac{T}{\tau}\right)\right) \textbf{Q. E. D.}$$

Proof of Lemma 2: The proof is done by contradiction. If the statement of the lemma does not hold, it is always possible to construct an alternative policy, which will satisfy (4) and give strictly bigger profit. Consider the interval $[s_j, s_{j+1})$ for arbitrary j. Suppose there exists a number $n \geq 1$[9], such that

$$s_j < \tau_1 < \ldots < \tau_n < \tau_{n+1} \leq s_{j+1}$$

[9]Note that even if s_{j+1} is infinite, n must be finite, because once z_i reaches the level of x_j, it stays there.

where $\tau_1 = \min\{\tau_i : s_j < \tau_i\}$ and $\tau_{n+1} = \max\{\tau_i : \tau_i \leq s_{j+1}\}$, with $z(t) = z_i$ on $[\tau_i, \tau_{i+1})$. From Claim 4, we have that $s_j = \tau_0$ or $s_j = \tau_0 + T$ and $s_{j+1} = \tau_{n+1}$ or $s_{j+1} = \tau_{n+1} + T$. Without loss of generality, assume that $\tau_0 = 0$.

The fact that $x(t)$ and $z(t)$ is the optimal solution on $[s_j, s_{j+1})$ means that there are certain restrictions on the possible levels of z_i and x_j, implied by the necessary conditions for the optimum. In particular, note that $x(t) = x_j$ is constant for all $t \in [s_j, s_{j+1})$. From Claim 2 and condition (2) it follows that $x_j \leq Z(\tau_n)$. Then, by Claim 3, $z_n = x_j$. Also, it must be the case that $x_j > Z(\tau_{n-1})$, because otherwise $z_{n-1} = x_j$, and there is no gain from purchasing z_n at time τ_n. It follows that $x_j > Z(\tau_i)$, for all $i \leq n-1$. By Claim 3, then, $z_i = Z(\tau_i)$, $1 \leq i \leq n-1$. Collecting all of the above gives the following profile for $z(t)$ on $[\tau_1, \tau_{n+1})$:

$$z(t) = \begin{cases} \theta Z(\tau_i) & t \in [\tau_i, \tau_i + T) \\ Z(\tau_i) & t \in [\tau_i + T, \tau_{i+1}) \\ \theta x_j & t \in [\tau_n, \tau_n + T) \\ x_j & t \in [\tau_n + T, \tau_{n+1}) \end{cases} \quad i = 1, ..., n-1$$

The level x_j must be big enough to make the purchase of z_n at time τ_n worthwhile. That is to say that buying $z_n = x_j$ at time τ_n gives at least as much profit as not buying z at time τ_n and using z_{n-1} instead. This is the case if and only if

$$\frac{x_j}{z_{n-1}} \geq \frac{1 - e^{-r\Delta\tau_n}}{\theta(1 - e^{-rT}) + e^{-rT} - rp_z - e^{-r\Delta\tau_n}}.$$

The RHS of this inequality is decreasing in $\Delta\tau_n$ and positive, because profit on $[\tau_n, \tau_{n+1})$ must be positive. Therefore, it must be the case that

$$\frac{x_j}{z_{n-1}} \geq \min_{\Delta\tau} \left[\frac{1 - e^{-r\Delta\tau}}{\theta(1 - e^{-rT}) + e^{-rT} - rp_z - e^{-r\Delta\tau}}\right] = \frac{1}{\theta(1 - e^{-rT}) + e^{-rT} - rp_z}.$$

Consider an alternative policy $x^{(1)}(t)$, where

$$x^{(1)}(t) = \begin{cases} z_{n-1} & t \in [s_j, \tau_n + T) \\ x_j & t \in [\tau_n + T, s_{j+1}) \end{cases}$$

Since $\min\left\{x^{(1)}(t), z(t)\right\} = \min\{x(t), z(t)\}$, both policies give the same output and have the same cost of z. The policy $x^{(1)}(t)$ is less costly if and only if

$$p_x e^{-rT}(z_{n-1} + e^{-r\tau_n} x_j) < p_x e^{-rT} x_j$$

or

$$\frac{x_j}{z_{n-1}} > \frac{1}{1 - e^{-r\tau_n}}$$

For this inequality to hold, it is sufficient that

$$\frac{x_j}{z_{n-1}} \geq \frac{1}{\theta(1 - e^{-rT}) + e^{-rT} - rp_z} > \frac{1}{1 - e^{-r\tau_n}}$$

This is equivalent to

$$\theta(1 - e^{-rT}) - rp_z + e^{-rT} + e^{-r\tau_n} < 1$$

Taking into account Assumption 1 and corollary of Claim 3, this is satisfied if

$$2e^{-rT} < 1$$

which is the case under Assumption 3. On the next step, we can dominate $x^{(1)}(t)$ by

$$x^{(2)}(t) = \begin{cases} z_{n-2} & t \in [s_j, \tau_{n-1} + T) \\ x^{(1)}(t) & t \in [\tau_{n-1} + T, s_{j+1}) \end{cases}$$

By similar argument, the policy $x^{(2)}(t)$ is less costly if and only if

$$\frac{z_{n-1}}{z_{n-2}} = e^{g\Delta\tau_{n-2}} > \frac{1}{1 - e^{-r\tau_{n-1}}}$$

or

$$e^{-r\tau_{n-1}} + e^{-g\Delta\tau_{n-2}} < 2e^{-gT} \leq 1$$

On the i-th step,

$$x^{(i)}(t) = \begin{cases} z_{n-i} & t \in [s_j, \tau_{n-i-1} + T) \\ x^{(i-1)}(t) & t \in [\tau_{n-i+1} + T, s_{j+1}) \end{cases}$$

The n-th step yields

$$x^{(n)}(t) = \begin{cases} z_0 & t \in [s_j, \tau_1 + T) \\ z_i & t \in [\tau_i + T, \tau_{i+1} + T) \quad i = 1, ..., n-1 \\ x_j & t \in [\tau_n + T, s_{j+1}) \end{cases}$$

By construction, $x^{(n)}(t)$ satisfies property (3). **Q. E. D.**

References

[1] Aghion, Philippe, and Peter Howitt, "On the Macroeconomic effects of Major Technological Change," unpublished, Ohio State University, 1996.

[2] Attewell, Paul, "Technology Diffusion and Organizational Learning: The Case of Business Computing," *Organization Science* 3, no. 1 (February 1992): 1 - 19.

[3] Bakh, Byong-Hyong, and Michael Gort, "Decomposing Learning by Doing in New Plants," *Journal of Political Economy* 101, no. 4 (August 1993): 561 - 583.

[4] Baloff, Nicholas, "Start-up Management," *I.E.E.E. Transactions on Engineering Management* 17 (November 1970): 132 - 141.

[5] Benhabib, Jess, and Kazuo Nishimura,"Competitive Equilibrium Cycles," *Journal of Economic Theory* 35 (1985): 284 - 306.

[6] Caballero, Ricardo, and Mohamad Hammour, "On the timing and Efficiency of Creative Destruction," *Quarterly Journal of Economics* (August 1996): 805 - 844.

[7] Cooper, Russell, and John Haltiwanger, "The Aggregate Implications of Machine Replacement: Theory and Evidence," *American Economic Review* 83, no. 3 (June 1993): 360 - 382.

[8] Jovanovic, Boyan, and Yaw Nyarko, "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves," *Brookings Papers in Economic Activity* (Microeconomics Issue) 1 (1995): 247 - 305.

[9] Klenow, Peter, "Learning Curves and the Cyclical Behavior of Manufacturing Industries," University of Chicago, September 1993.

[10] Kremer "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics* (Aug.1993):551-75.

[11] Markoff, John, "Despite Rivals' Chip Speed, Intel Still Leads PC Market," *New York Times*, October 17, 1994.

[12] Matsuyama, Kimonori, "Complementarities and Cumulative Processes in Models of Monopolistic Competition," Discussion paper no. 1106. Department of Economics, Northwestern University, October 1994.

[13] Milgrom, Paul, and John Roberts,"The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review* 80, no. 3 (June 1990): 511 - 528.

[14] Nye, David, *Electrifying America*, Cambridge: MIT Press, 1995

[15] Parente, Stephen, "Technology Adoption, Learning-by-Doing, and Economic Growth," *Journal of Economic Theory* 63, no.2 (August 1994): 346 - 69.

[16] Pollins, Harold, *Britain's Railways: An Industrial History.* Totowa, N.J.: Rowman and Littlefield, 1971.

[17] Sherrington, C.E.R. *The Economics of Rail Transport in Great Britain*, vol. 2: Rates and Services. London: Edward Arnold, 1928

[18] Sheshinski, Eytan, and Yoram Weiss, "Staggered and Synchronized Policies under Inflation: The Multiproduct Monopoly Case," *Review of Economic Studies* 59, no. 2 (April 1992): 331 - 359.

[19] Stolyarov, Dmitriy, "A Model of Skilled Labor Supply with Learning," University of Pennsylvania 1996.

[20] Varian, Hal, *Microeconomic Analysis*, New York: Norton 1992.

[21] Womack, James P., Daniel T. Jones, and Daniel Roos. *The Machine that Changed the World.* New York : Rawson Associates, 1990.

[22] Zeckhauser, Richard, "Optimality in a World of Progress and Learning," *Review of Economic Studies* 35 (1968): 363 - 365.