The Missing Link: Technology, Productivity, and Investment

Laura Power*
U.S. Department of the Treasury

CES 95-12 October 1995

Center for Economic Studies, Room 1587, FB 3, Bureau of the Census, Washington, DC 20233-6101, (301-457-1882) or INTERNET address snguyen@info.census.gov.

Abstract

This paper examines the relationship between productivity, investment, and age for over 14,000 plants in the U.S. manufacturing sector in the 1972-1988 period. Productivity patterns vary significantly due to plant heterogeneity. Productivity first increases and then decreases with respect to plant age, and size and industry are systematically correlated with productivity and productivity growth. However, there is virtually no observable relationship between investment and productivity or productivity growth. Overall, the results indicate that plant heterogeneity and fixed effects are more important determinants of observable productivity patterns than sunk costs or capital reallocation.

Key Words: productivity, investment, technical change

Office of Tax Analysis, US Department of the Treasury, 1500 Pennsylvania Avenue, NW, Washington DC, 20220, USA, Phone 202-622-0238, Fax 202-622-0236. Work was done on site at the Center for Economic Studies of U.S. Census Bureau while author was research associate. The views expressed are the authors and

do not necessarily reflect the views of the Census Bureau or the Department of Treasury. I would like to thank Peter Cramton, Mark Doms, Tim Dunne, Curtis Grimm, David Joulfaian, Edward Montgomery, Dennis Mueller, Mark Roberts, Ploutarchos Sakellaris, Dylan Supina, and Ken Troske for many helpful comments, suggestions, and editorial remarks. A special thanks to Bill Evans for many helpful insights, and especially John Haltiwanger, for all of his thoughts and ideas.

I. Introduction

Substantial existing research attempts to unravel the mystery through which technological innovation generates economic progress. Within this theoretical spectrum, vintage models are one class of models which often share similar assumptions regarding the dissemination of technology, and its impact on growth. The basic premise is that technology improves over time, but a variety of barriers (eg., fixed or sunk costs) impede most plants from immediately acquiring the newest vintage.[1] Most often, technological improvements are manifested in the form of more productive machinery, a hypothesis which has been labeled machine embodied technical change (Greenwood, Hercowitz, and Huffman (1988)). Implicitly, machine embodied technical change suggests a strong correlation between high productivity and high recent levels of investment (Baily, Hulten, Campbell (1992)). Beyond its hypothetical convenience, the idea that there is a direct link between productivity and investment has been an important element of economic thought, as well as a cornerstone

[1]There is a large, related literature which focuses on nonconvexities in investment and lumpy investment behavior. See Cooper, Haltiwanger, Power (1995), Pindyck(1991), Caballero and Pindyck(1992), Doms and Dunne(1993).

of 20th century U.S. tax and fiscal policy (Cooley, Greenwood, Yorukoglu(1994)).

However, as Solow succinctly stated in 1962 "This (i.e. the notion that new investment embodies new technology) is certainly not literally true. No one knows whether it is more or less true than the exactly opposite assumption." Almost forty years later, the concept of machine embodied technical change has not actually been tested empirically. The primary purpose of this paper is therefore to test the premise of machine embodied technical change, and concomitantly the widespread belief that investment generates high productivity, through a detailed analysis of the relationships between productivity and investment.

To do so, I discuss the implications of combining plant heterogeneity and fixed costs, in the context of a simple vintage framework. The discussion illustrates how machine embodied technical change influences productivity and growth. Further, the plant level focus of the analysis highlights the influence of plant heterogeneity on observed productivity patterns.[2] All of these relationships are then tested empirically using a plant

---

[2] The importance of plant heterogeneity in determining observed economic patterns has since been the focus of a great deal of theoretical research. Many empirical examinations of the variation in economic variables with respect to observable plant characteristics has supported these theories. For theoretical models see Pakes and Ericson (1987, 1988), Lippman and Rummelt (1982) , Lambson (1989), and Dixit (1989). For empirical support, see Dunne, Roberts, and Samuelson (1989) , Evans (1987), Hall (1987) , Garen (1989), and Baily, Hulten, and Campbell (1993) offer empirical support of the predictions. In related empirical work, Davis and Haltiwanger (1992) investigate the empirical relevance of passive learning models for job reallocation rates. One facet of this is an examination of the implications of plant age for job reallocation patterns.

level data set that includes almost 14,000 plants.

The major findings of this analysis refute the hypothesis of machine embodied technical change. In particular, virtually no correlation exists between high productivity and high recent levels of investment. However, the importance of size, industry, and permanent plant characteristics suggest that systematic differences among plants play an important role in determining observed patterns. Thus, new investment is only one small component of productivity - other, plant specific influences such as management or location, play an even more important role. The results further suggest that the relationship between investment and productivity is not causal. Perhaps investment really only "pays off" for plants which are already productive due to other factors.[3]

The analysis also has broad implications for a second assumption concerning the dissemination of technology. The logic behind this hypothesis is that sunk costs provide new plants with an advantage in acquiring the latest technology. Therefore, technological dispersion occurs through the birth of new plants (see Campbell (1994)).[4] This notion, which is called plant

---

[3] To clarify, note the distinction between investment causing high productivity, and investment benefitting productive plants. The former is the basis for the machine embodied technical change assumption, and has been the impetus behind U.S. fiscal policy. It implies a strong correlation between high productivity and high recent levels of investment, as well as a decline in productivity with respect to investment age.

[4] Other methods of technological dispersion such as endogenous innovation (Andolfatto and MacDonald(1993)) have also been postulated.

embodied technical change, implies that across plants, productivity decreases with respect to plant age. My empirical results imply that, across plants, productivity initially increases, and later decreases, with respect to plant age. This is not inconsistent with the notions underlying plant embodied technical change, however, the overall results suggest that plant idiosyncracies play a more important role in determining observed productivity patterns than sunk costs or capital reallocation.

The structure of the paper is as follows. Section II outlines the framework and the testable implications. Section III describes the data set used to conduct the analysis, and provides the variable definitions. Section IV presents the results, and section V discusses the implications of the results and concludes.

II. The Theoretical Background:

As mentioned above, vintage models are a class of models which often incorporate hypotheses and predictions regarding the dissemination of technology. The common theme is that overall growth is determined through the combination of technological improvement and barriers to acquisition. From the theoretical framework, detailed predictions regarding productivity fluctuations (Campbell, (1994)), the balanced growth path (

Cooley, Greenwood, Yorukoglu(1994)), transitional dynamics (Krusell 1992)), and the behavior of economic variables (Greenwood, Hercowitz, and Krusell(1992)), have been derived. From a different perspective, models focusing on plant dynamics (eg. Pakes and Ericson (1987,1988), Lambson (1989), Dixit(1989)) also contain significant predictions regarding technology and the evolution of economic variables.[5]

The implications of both of these types of models are highly pertinent to the present study, and two specific studies are particularly relevant. The first is the machine replacement model of Cooper and Haltiwanger (1993), and the second is Jovanovic 's (1982) model. The former is a vintage capital model which is built around the investment decision. Investment is driven primarily by the fixed costs which generate lumpy investment behavior, and the resulting investment patterns can have aggregate implications.[6] The key to the latter model is the selection among plants which results from their systematically different cost structures (i.e. plant heterogeneity). The implications of these heterogenous cost structures include declining failure rates with respect to size, as well as a positive correlation between size and age. The present study

[5]See footnote 2.

[6]Cooper, Haltiwanger, Power (1995) investigates the timing of lumpy investment episodes at the plant level, and also analyzes the aggregate implications of these timing decisions.

relies heavily on many of these ideas developed in these models, and in fact builds on elements within them.

To better understand the implications of machine embodied technical change, it is useful to conceptualize plant heterogeneity in the form of productivity differences, and also to adopt a vintage framework. These features are synthesized and examined in a simple theoretical framework (Power (1994)). Because the focus of this analysis is the empirical test of machine embodied technical change, this structure is not explicitly detailed in the present paper. However, in order to illustrate the main points, its primary features are discussed below.[7]

Suppose that plants can differ randomly on two levels -their permanent productivity feature 2_{ip} (e.g. managerial ability), and their time variant idiosyncratic productivity feature 2_{imt} (i.e. outcome of investment). Suppose further that each attribute is randomly drawn from a distinct distribution, but jointly they determine each plants' profits over time. Thus, $B_{it} = f(2_{ip}, 2_{imt})$, where plants are indexed by i, time is index by t. Prior to entry, there is uncertainty as to the value of both the permanent

[7]The simple framework derived in Power (1994) combines features from both the Cooper Haltiwanger model and the Jovanovic model in order to highlight the interaction of fixed costs and plant heterogeneity. It is presented in complete, detailed, and explicit form in Power (1994). Its implications are also proven formally in Power (1994). However it is not a structural model; that is, it is not designed to describe the functional form of the relationship between productivity and these plant heterogeneities. Rather, it is a broad system designed to highlight the above mentioned factors.

and the time variant characteristics, but the value of the permanent component is revealed at the end of the first period after the plant's birth. Since this value is fixed, its impact can be captured by tracking productivity patterns with respect to plant age.[8]

Now suppose that there are fixed costs associated with the process of investment. This implies that plant level investment occurs in spurts, and the occurrence of these spurts can be labeled investment spikes. The outcome of these investment spikes is random, and thus, after each investment spike, each plant gets a new realization of 2_{imt}, which can be denoted 2_{imt}^*

The value of 2_{imt}^* is learned at the end of the period of investment, and its impact can be evaluated by tracking the pattern of productivity with respect to the time elapsed since the last investment spike. Thus, if this elapsed time is called the plant's investment age, then $i_t=(t-s)$; where i_t is investment age, t is the current time period, and s is the period in which the last investment spike occurred.

To incorporate the notion of technological change, assume that the mean μ_t of the distribution from which 2_{imt} is drawn is increasing over time at a constant rate (, which implies that $\mu_t=(\mu_{t-1}$, (>1. This essentially reflects the idea that new

---

[8]In order to track the pattern of productivity across plants using age, it must be assumed that cohorts are not systematically different regarding the relationship between productivity and age. This assumption was substantiated empirically.

technology is embodied in new machinery, and therefore implicitly incorporates the assumption of machine embodied technical change.

The actual value of 2_{imt} at time t also is influenced by depreciation. That is, the value depends on whether the plant invests in period t, and, if not, how long it has been since the plant last invested. For example, if 2_{imt}^{0} is the current value of the plant's last draw from the 2_{imt} distribution, then $2_{mit}^{0}=(D^{j-1}2_{mij}^{0})$, where j is the last period in which plant i had an investment spike, and the depreciation rate, D, is between 0 and 1.

Overall, each plant's decides to whether or not to invest in a given period by comparing its total value from the current period onward if it invests (V_i^I), with its total value if it does not invest (V_i^N). The total value of investing (not investing) is determined by the current period utility from investing (not investing), as well as by the entire future stream implied by the decision to invest (not invest). Thus, if c is the operational cost incurred during each period in which the plant operates, k is the cost of investing, and W_i^I is the current utility from investing, then $EW_i^I= E(2_{ip}+2_{im}^{*}-c-k)$, and $V_i^I=EW_i^I+ \$EV(D2_{im}^{*})$. Further, if W_i^N is the current utility from not investing, then $EW_i^N= E(2_{ip}+D2_{im}^{0}-c)$, and $V_i^N=EW_i^N+\$EV(D^2 2_{im}^{0})$. Finally, note that it is the fixed cost k which generates lumpy investment behavior, and that the discount rate, \$, must be between 0 and 1.

Framework Implications

The implications of this framework are the focus of the empirical analysis. The first is that the productivity of plants which have recently invested is higher than that of plants which invested long ago. This is intuitive, and arises because depreciation lowers the productivity of machines over time, and because improvements in technology and research and development enhance the productivity of new machinery at rapid rates. Note that this prediction derives directly from the assumption of machine embodied technical change; that is, since new machinery embodies the newest technology, on average, high recent investment is associated with higher productivity. Thus, empirical tests of this prediction implicitly test the assumption of machine embodied technical change.

The second is that failure rates decline with respect to plant age; in particular, old plants have a lower failure rate than young or medium plants. It is essentially uncertainty over permanent attributes precipitates this failure. Plants enter with the belief that they will be able to make positive profits. Based on their 2_{pi} and 2_{mit} draws, however, some learn that this is not possible.

Finally, the mean levels of productivity across plants increase with respect to plant age. In this simple theoretical context, the mean of productivity increases with respect to plant age because of selection effects: over time, the less productive plants exit, causing the overall mean to rise. Realistically, other effects can cause an increase in productivity with respect to age. For example, learning causes productivity to increase rapidly at young ages, and then more slowly over time. This implies a concave pattern of productivity with respect to age. Life cycle models hypothesize that, in older plants, managerial discretion inhibits profit maximization, and therefore the relationship between productivity and age is actually humped shaped.[9]

Note, however, that, in its strictest form, plant embodied technical change essentially implies a purely negative correlation between productivity and age. The driving force behind this assumption is the existence of sunk costs: existing plants cannot acquire the latest technology with the same relative ease as the new plants, because of their sunk costs. Thus, the newer plants are free to purchase the newest technology, and therefore they have highest productivity. The empirical analysis attempts to simultaneously examine the importance of sunk costs, learning, and selection, by examining the pattern of productivity with respect to plant age.

III. The Data

The data set is a pooled cross-section time-series extract

[9] See Grabowski and Mueller (1972)

from the Longitudinal Research Database (LRD). The LRD is a plant level panel data set containing information on more than 750,000 U.S. manufacturing establishments for the years 1963, 1967, and 1969-1988. The extract utilized for the regression analysis contains annual information on 13,936 large manufacturing plants from 1980-1988. In addition, information on these plants from 1972-1980 was used to construct some of the variables.

Table 1 reports the number of plants and total number of observations contributing to each industry. The distribution of plants in the data set across two digit industries in roughly comparable with that of total manufacturing, and all twenty industries are well represented. In total, there are 109,647 observations. Table 2 reports number of plants and total number of observations by plant size. Plants are assigned to size classes based on their average, size weighted employment over the entire sample period.[10] Although the data set excludes small plants, the 13,936 plants which are included comprise the universe of the large manufacturing plants in the United States. Thus, the distribution of plants across these five large size classes is identical to their distribution for the total manufacturing sector. It should be noted that, although there is a higher concentration of plants in the "smaller" size classes,

[10] $SIZE = (1/mnte)\{G_{i=72}^{i=88}(te_i)^2\}$, where mnte=mean(te72,...te88), and te is annual number of workers. This average size measure was chosen to capture long run size,and to avoid transitory fluctuations in size (Davis, Haltiwanger, Schuh (1995)).

all size classes are well represented. Finally, Table 3 reports the distribution of plants over time. While most plants are in existence for the entire sample period, the combination of entry and exit results in a small overall decline in the number of plants and number of observations over time.[11]

The Variable Definitions

To construct exact plant age, data from 1972-1980 are used. Starting in 1980 and looking backward to 1972, plant age is thus defined as the difference between the current year and the first year the plant is ever recorded in the data set, until the plant reaches age eight. After this, the plant is always assigned age eight. Therefore, the possible plant age categories are age 0 to age 7, and age 8+.[12] Table 4 reports the number of observations in each of the 9 age categories. The majority of observations in the data set are in the oldest category, although all ages have a substantial number of observations.

Investment age measures vintage as defined by the time

[11]In particular, from 1980-1988 969 plants are born, and 2,592 plants die. However, 238 of the births have missing productivity in their birth year, and are thus excluded from the regression analysis.

[12]Robustness checks were performed by running productivity age regressions on the subset of this data set which does not include the open ended age category (that is, the data set which includes only those 7,365 large plants born after 1973, and which contains 16 precise age measures). The patterns of productivity with respect to plant age for this subset were similar to those found using the open ended age category. Further, a test regression was run including a separate dummy variable for plant age 8, to see if there were significant productivity differences for age 8 plants; specifically whether the inclusion of the open ended age category was distorting the regression results. The coefficient of the age 8 dummy was insignificant, indicating that no strong bias exists.

elapsed since the occurrence of an extremely large investment. Thus, the definition of investment age requires the definition of an investment spike: in order to determine a plant's investment age, "lumpy" investment episodes must be identified, and then the time between these episodes tracked.

The concept of lumpy investment chosen for this analysis is a relative one; that is, a plant's investment is considered lumpy if it is large relative to that plant's other investments.[13] In particular, an investment spike is defined as an investment event ($) which is extremely large (") relative to each plant's own normal investment ($$_{norm}$).

$$\text{If}(\ _t) > (\ (\ _{norm}))\ \text{then an investment spike occurs in period } t \tag{1}$$

This definition has several merits. First, it effectively captures the intuitive notion of lumpy relative investment, because it attempts to identify periodic, large bursts of

[13] One could imagine an absolute definition of a lumpy investment, such as an x% change in total capital stock. This type of definition focuses on expansion, and might be more appropriate for analyzing the aggregate implications of investments as a whole. The correspondence between relative lumpiness and absolute lumpiness was tested, and it appears that many, but not most, absolute spikes are also relative, and vice versa. The absolute definition captures many smooth expansions which are ignored by the relative definition, and the relative definition captures many investments which are large relative to the plants other investments, but not large in any absolute sense.

investment.[14] Further, it is consistent with the notion that fixed costs generate lumpy investment. Finally, it ensures that relative investment spikes are lumpy in a general sense. It does so by imposing a uniform definition of large relative to the normal investment across plants; formally, " is restricted to be the same for all plants.

Thus, the choice of " must satisfy some concept of "large", and in an attempt to satisfy this criteria, " is set at 2.5. Although intuitively, it seems reasonable to identify an investment two and a half times as large; it is also admittedly ad hoc. Therefore, all analyses were also conducted for two alternative specifications of ": " =1.75 and "=3.25.[15] With one minor exception (to be noted later), all of the results are very similar. Therefore, in the interest of simplicity and clarity, only the " =2.5 results are presented.

Throughout the analysis, the rate of investment in period t - $\$_t$ - is defined as the ratio of the plant's nominal new machinery purchase (nm) in year t to its total nominal book value

---

[14]Essentially, the structure assumes the existence of a plant level investment distribution which has a high concentration of investments of small investments, but which has a long right hand tail, indicating the periodic occurrence of unusually large investments.

[15]In order to gain some sense of the breadth of the definition, several investment spike characteristics, including number of spikes, percent of spike observations, and total sample investment accounted for by spikes, are presented for each of the definitions of " in Table 7. For a much more complete analysis and discussion of the characterisitics of investment spikes, robustness checks, etc., see Power (1994).

of capital (mae) at the end of period t.[16] $\$_{norm}$ - each plant's normal investment - is defined as the median of its investments over the entire sample.

However, some investment projects are large enough that they might last more than one year, and thus a single annual accounting period need not necessarily reflect the total expenditures necessary to complete a project. Further, even a "year long" project need not begin at the start of the accounting year, nor end at close of the accounting year, which implies that a portion of the investment necessary to complete the project could be distributed over two consecutive years. In these instances, the true investment spike is obviously the total investment recorded in all of these years. In an attempt to capture such events, adjacent years of relatively intense investment activity are grouped into a single investment event. This grouping is labeled a multi-year spike, and is modeled using the following specification.[17]

---

[16]The nature of Census data is such that beginning of period assets do not exist in 1972, and book values are imputed in 1986, and 1988. Therefore, in order to utilize the most information possible, end of period assets, rather than the traditional beginning period of assets, are utilized as the measure.

[17]Although in principle this specification allows for a multi-year spike of any duration, approximately 90% of the multi-year spikes are three years or under.

If ($_t$)>(($_{norm}$)) and ... ($_{t\ i}$)>(($_{norm}$)),

where 0< < , then a multiyear investment (2)

spike occurs in periods t ... (t i)

Obviously, there are an infinite number of combinations through which an investment project could be portioned out over consecutive accounting years. The parameter O must attempt to capture all of these combinations, without losing the notion that the investment is intense or lumpy. As in the case of ", the designation must ultimately be somewhat ad hoc. Therefore, in order to ensure robustness of the results, nine alternative values of O were tested, and all provided quantitatively similar results.[18] Again , in the interest of simplicity and clarity, only the results for the value 2.25 (i.e. 90% of ") is presented.

Given this definition of an investment spike, investment age is defined in the following manner. The analysis is initialized in 1980, and the data from 1972-1980 are used to construct eight precise investment age categories. Starting in 1980 and looking backward to 1972, investment age is equal to plant age for all plants, until a plant has its first investment spike. Thereafter, investment age is defined as the difference between the current year and the year of a plant's most recent investment

[18]Again, for a more complete discussion and analysis of the characteristics of investment spikes and multiyear investment spikes, as well as their robustness, see Power (1994).

spike, until the plant reaches investment age eight.[19] After this, the plant is assigned age eight until it experiences another investment spike. Therefore, the possible investment age categories are age 0 to age 7 and age 8+. Plants which never have investment spikes have investment ages equal to their plant age for the entire sample period. The distribution of plants across investment ages is reported in Table 5. The distribution is fairly consistent, with a slight downward trend, until age 8.[20]

IV. The Empirical Analysis

The purpose of the empirical analysis is to test the hypotheses of machine embodied technical change, and more generally to shed some light on the relationship between productivity, investment, and age. Following a common practice of productivity analysis, I estimate a logarithmic specification.[21] However, to avoid imposing an arbitrary structure on the complex functional relationship between productivity and investment, a reduced form OLS specification is employed.

---

[19] The structure therefore implies that the variable is reset to zero during every period of major investment.

[20] Note the large number of observations in the investment age 8 is due to the open-ended nature of the category.

[21] See Baily, Hulten, Campbell(1992), Olley and Pakes (1990).

$$Y_{it} = \alpha + \beta \quad + \varepsilon$$

$$\log(lp_{it}) \ , \ lpdif_{it}$$

$$\log(lp_{it}) = \log(tvs_{it}) - \log(tph_{it})$$

$$tph_{it} = \quad ((sw_{it}) + (cw_{it})) / (ww_{it}))$$

$$pdif_{it} = \log(lp_{it}) - \log(lp_{it}$$