Are new firms an important source of innovation?

Evidence from the PC software industry

Thomas J. Prusa and James A. Schmitz, Jr.

State University of New York, Stony Brook, NY 11794, USA

We examine data from the PC software industry to determine if new forms play an important role in the advance of technology in this industry over the period 1982–1987. We find that new firms have a comparative advantage (over established firms) in *creating* new software categories, while established firms have a comparative advantage in developing subsequent improvements in *existing* categories.

1. Introduction

In their important book, Jewkes, Sawers and Stillerman (1958) (henceforth, JSS) studied the sources of the important inventions of the 19th and 20th centuries. A primary goal was to determine the importance of the small, independent entrepeneur relative to the large industrial research laboratory in developing these inventions. They found, contrary to the popular wisdom of the day, that the small, independent entrepeneur continued to play an important role in technological change during the 20th century. The purpose of this paper is to examine some of the same issues studied by JSS but for a more recent experience. In particular, we examine the importance of 'new' versus 'established' firms in generating innovations in the PC software industry. Similar to JSS, we find that new firms contribute a very large share of industry innovation.

The question of whether new (or small) firms play an important role in technological progress is an issue which has attracted considerable theoretical attention [see e.g. Holmstrom (1989), Williamson (1985) and Aron and Lazear (1990)]. Clearly more empirical evidence is needed on this issue. While this paper explores only a single industry, data similar to ours should be available for many more, from sources such as trade groups and for-profit firms (the source of our data).

2. Description of data

The data was obtained from Dataquest, Inc., an information services company. For each of 206 software products we have the following information:

(1) the introduction date of the program (month, year),
(2) the software category of the program (Dataquest has defined 17 categories which will be given below),

(3) the operating system on which the program runs,
(4) the firm marketing the program, and
(5) the annual unit sales of the program for the years 1982–1987.

Of course, most programs were introduced sometime during the period 1982–1987 and for those programs sales figures are available for the introduction year and beyond. Dataquest obtains its estimates for unit sales by surveying industry vendors (i.e. wholesalers and retailers)[1].

There were obviously many more software programs over this period than the ones in the data set. While the data set has only a few of the many programs available in this period (all programs in the data set sell at least 1000 units annually), the program in the data set account for a large majority of industry sales. In particular, the Software Publishers Association, the industry trade group, estimates that industry revenue was $2.7 billion in 1987. The revenue from the sales of programs on the data set in 1987 was $2.5 billion.[2]

3. New and established firms' share of category innovations

We define a firm to be 'new' at date t if the firm first appears on the data set at time t, while a firm is said to be an 'established' firm if it has appeared on the data set prior to time t. We also need a means to rank innovations. Ideally, we would like to use the social surplus created by a product as the means to rank innovations [as done by Trajtenberg (1989)]. Social surplus created by a program cannot be calculated from the data set but two reasonable proxies can be employed. Let us initially say a program is an important innovation if the program is the first program in a category. There is good reason to define a program which creates a new category as an important innovation. By showing the possibility of a new type of software, the program will obviously generate large 'spillovers'.

There are a total of seventeen software categories as defined by Dataquest (see table 1). In fifteen of these categories there is a unique firm which 'opened up' the software category. In two categories there was more than one firm; in one category four firms introduced programs 'simultaneously' (i.e., same month and year), while in the other category two firms introduced programs simultaneously. There are therefore a total of 21 products which introduced software categories. Of these 21 products, 18 (or 86%) were introduced by new firms and 3 were introduced by established firms. The type of firm introducing the first product in each category are given in the first column of table 1. The fourth column presents the number of established software firms in the data set at the time a category was introduced. This column defines the set of established firms which could have developed the category-opening product. The third column reports the total number of programs in each category, which might be thought of as a measure of the 'spillover' created by the first product.

These numbers do not imply, however, that established firms are not innovating. They are. First, of the 206 products introduced over this period, 108 (52%) were introduced by established firms, while 98 were introduced by new firms. Second, of the 108 products introduced by established firms, 47 (44%) were in categories in which the firm did not have a previous product. So nearly half the products introduced by established firms are in 'new' (for the firm) software categories. Third, these firms have developed a more equal share of important innovations under the second proxy for social surplus created.

---

[1] The reader might be interested in how the results below depend on the influence of the 'industry standard' operating systems, Microsoft DOS, Apple DOS, Apple Macintosh DOS, and Digital CP/M. The 206 programs actually do not include these programs which are bundled with hardware; if we had included them none of the results below would change.
[2] Dataquest does not report revenue data by program but by firm (from software sales). We use this firm revenue data in the comparison above.

Table 1
Software category innovators.

Category	Type of firm introducing first program [a]	Type of firm introducing category's product with greatest cumulative sales [a]	Total number of programs in category	Number of firms in existence prior to category's opening
Database management	(1,0)	(1,0)	23	0
Accounting	(1,0)	(1,0)	17	1
Word processing	(1,0)	(1,0)	26	1
Spreadsheets	(0,1)	(1,0)	14	6
File management	(1,0)	(1,0)	11	8
Graphics	(0,1)	(0,1)	33	8
Integrated	(1,0)	(0,1)	12	8
CAD	(1,0)	(1,0)	8	13
Languages	(1,0)	(1,0)	3	18
Personal finance	(4,0)	(1,0)	4	19
Utilities and accessories	(1,0)	(0,1)	23	19
Project management	(1,0)	(0,1)	4	29
Communications	(1,0)	(1,0)	8	30
Desktop publishing	(1,0)	(1,0)	8	47
Operating environments and operating systems	(1,0)	(0,1)	3	47
Development tools	(1,0)	(0,1)	7	53
Information Management	(1,1)	(0,1)	2	88
Total	(18,3)	(10,7)	206	

[a] The pair (x,y) in parentheses are defined as follows: x denotes the number of firms new firms while y denotes the number of established firms.

As the second proxy for the important innovation in a category we use the program which has the greatest cumulative sales in the category. Under some reasonable conditions (e.g., development costs approximately equal across programs), sales should also be a decent proxy for social surplus created by a program. We find that new and established firms provided roughly the same number of programs which had the greatest cumulative sales in each category (column 2, table 1). Comparing columns 1 and 2 of table 1 suggests that new firms have a comparative advantage in developing *new* categories, while established firms have a comparative advantage in developing subsequent improvements within existing categories.

In table 2 we consider the possibility that many firms were developing category-opening programs independently and introduced them a few months apart. If this were the case then we would want to count these other products as category 'openers'. In particular, suppose it takes k months to imitate a new program. Then all programs which were introduced within k months of a category-opening product were developed independently of the first program and should also be judged as *creating* the new category. In table 2 we present the types of firms which independently developed category-opening products based on different assumptions of how long it takes to imitate a program. We see that the result in column 1 of table 1 does not change much for the different imitation lags. The result that new firms provide a large share of category opening products is a robust one.

Table 2
Alternative specifications for 'limitation lag'.

Imitation lag value (months)	Type of firm introducing first program [a] (Total)
0	(18,3)
3	(20,4)
6	(22,6)
9	(22,6)
12	(23,10)

[a] The pair (x, y) in parentheses are defined as follows: x denotes the number of new firms while y denotes the number of established firms.

References

Aron, Debra and E. Lazear, 1990, The introduction of new products, American Economic Review Paper and Proceedings 80, 421–426.

Holmstrom, B., 1989, Agency costs and innovation, Journal of Economic Behavior and Organization, 305–327.

Jewkes, J. and D. Sawers and R. Stillerman, 1958, The sources of invention, (Macmillan, London).

Trajtenberg, M., 1989, The welfare analysis of product innovation, with an application to computed tomography scanners 97.

Williamson, O., 1985, The economic institutions of capitalism (The Free Press, New York).