ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Master Thesis

Labor Mobility in the Disk Drive Industry:

Do Pioneering Firms Attract More Inventive Employees?

Economics & Business MSc Programme
Supervisor: Dr. A.S. Bhaskarabhatla

Name: John Wilken
Student Number: 324310
Contact: John.Wilken@web.de

Rotterdam, 16.08.2012, Reprint Prohibited

ABSTRACT

This thesis studies the causal relationship between shakeout periods and labor mobility. Two previously existing datasets containing patent-, inventor- and company-level observations of the disk drive industry are employed. A measure of innovative firms is created and a hypothesis formulated. The hypothesis states that inventive employees tend to move to companies that pioneer in product markets. All coefficients of the variable of interest turn out positive throughout the regressions and thus suggest the hypothesis to be true.

KEYWORDS: *labor mobility, Disk drive industry, Shake out, Innovation*

JEL CLASSIFICATION: J61, L63, O31

If we knew what it was we were doing, it would not be called research, would it?

Albert Einstein

ACKNOWLEDGEMENTS

I would like to express my special appreciation to my supervisor Dr. Ajay Bhaskarabhatla for supporting me in my work on this thesis. Not only has he introduced me to this topic and infected me with his seemingly infinite enthusiasm for research, but he also provided me with significant encouragement. I am particularly grateful for his aid in working out the correct commands in Stata, which saved me valuable search time. Consequently, I was able to focus more on the research matter.

Furthermore, I would like to thank Clay Christensen, Lee Fleming, Ronald Lai, Amy Yu, Alexander D'Amour and Ye Sun whose datasets made this research possible in the first place. Lee Fleming et al. in particular are role-models in providing their work publicly and thus help to enhance research effort and ease of study.

I also owe my gratitude to Archie Pollock and Fritz Rabeler for their input on content and language.

Lastly, I would like to thank my friends, colleagues, and family, who encouraged and inspired me throughout the writing process.

The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" but rather "hmm....that's funny..."

Isaac Asimov

Table of Contents

Introduction

In recent years, the availability of disaggregated data on industries, inventors, and employer-employee matches has provided an opportunity to study underexplored and important topics in industrial organization. Specifically the relatively young field of labor mobility has benefitted from this development. As Trajtenberg and Shiff (2008) mentioned in their 2008 publication "Identification and Mobility of Israeli Patenting Inventors" 'every economic phenomenon appears in a certain "location" in time and space. There was so far a lot of attention to the time dimension (e.g. discounting) but much less so to space'. The research into labor mobility may still be considered to be in its infancy despite the potential knowledge that can be gained in this field. Therefore, various opportunities for scientific progress remain and hence this thesis will focus on labor mobility.

The introduction of information technology has aided researchers greatly in the process of research[1]. Manually gathering reliable data, particularly on laborers moving between work places, has proven to be a very difficult, if not an impossible, assignment in the past. Furthermore, the internet provides a significant source for several aspects of labor mobility such as addresses, organization numbers or merger and acquisition information. Thus, the concern about data availability has depreciated in recent years.

When researching labor mobility, one should search for an industry that has experienced a strong influx and outflow of workers as well as a vigorous intra-industry exchange within the workforce in order to obtain a sufficient amount of data. An industry that exhibits these characteristics is the disk drive industry, which incidentally has a large, if not the main, impact on the information technology development. In particular inventors that work for technological companies have been identified to move freely amongst firms (Trajtenberg, 2008). The industry began to expand in terms of the number of firms in the mid-1970s, a period when many other relevant data, such as patents, had begun to be collected.

The disk drive industry is especially suited for labor mobility research because of its quick development through the industry's life-cycle for which it has been labeled with the term "fruit-fly industry" (Christensen, 1997). It had become a major industry during the 1980s and despite its young age has since developed extraordinarily fast and provides researchers with a vast amount of information, compressed into a 30-year life-cycle. In addition, the tenure of

[1] For an example, see Trajtenberg, Shiff, Melamed (2006) in the literature review

even the early cohorts of workers in the industry is longer than the shakeout period in the industry, which lasted the relatively short period of less than ten years. Consequently, the disk drive industry is a very suitable platform to conduct research into the field of labor mobility because of its quick development, availability of data, and high mobility of workers.

One additional advantage of the disk drive industry as a research platform over others is that it has moved through a shakeout period in its life-cycle. The shakeout in the disk drive industry has been identified to start at the end of the year 1983 (Lerner, 1997). A shakeout is defined as "a fast drop in the number of independent producers in a given industry. Most shakeouts are accompanied by growth in industry output and average firm size, and are preceded by a fast rise in the number of producers. According to Jovanovic and MacDonald (1994), the reason for these changes is a technological innovation that dramatically increases the efficient scale of production." (Financial Times Lexicon, 2012) This offers the opportunity to test the impact of a shakeout on labor mobility. I am testing the hypothesis that inventive employees tend to move to companies that pioneer in new product markets. With the aim of attaining satisfactory results I expect the coefficient estimates of the dependent variable of interest to turn out positive in all regression specifications. An inventive employee is defined as a staff member associated with at least one patent or co-patent in the dataset, who is employed at a company producing disk drives (or disk drive parts).

In order to test the hypothesis I develop in this thesis, I use two datasets. One dataset was developed by Clay Christensen and provides extensive data on patents and companies in the disk drive industry. The other dataset is available on the internet and was publicized by Lee Fleming and co-workers and contains detailed information on inventors in the disk drive industry.

After reorganizing and cleaning those datasets, a descriptive analysis gives a rough overview on the extent of labor mobility over the industry life-cycle; especially the time period after the beginning of the shakeout is of major interest. In order to test the main hypothesis, I build a multiple regression and deploy a negative binomial approach that fits the distribution of the data.

The remainder of this paper is structured as follows. Section one gives a rough overview on the development of the disk drive industry throughout its life-cycle so far. I consider this worthwhile for readers who are not that familiar with this industry; especially as it is somewhat unique in its speed of development and occurrence of a shakeout period triggered by disruptive innovations. This will already cover some part of the existing literature, especially the one concerning the history of this industry and its underlying factors.

In the second section more specific literature is reviewed thoroughly. The significant academic works selected for this research are rather limited because of the relative novelty of the string of labor mobility research. Furthermore, additional literature on the start of patent data usage in academic research will be included in the review.

Section three contains a description of the dataset used in the analysis of this paper. I will explain the origin of the data as well as my own effort in attaining a suitable dataset by altering and merging Christensen's and Fleming's data.

Section four covers the descriptive analysis of Fleming's and Christensen's datasets individually, as potentially valuable information can be gained by examining these datasets apart from each other.

Following this, section five explains the model used for the regression analysis and the results gained. Hereafter, robustness tests will strengthen the outcomes. The thesis will conclude with a discussion of the results and potential limitations to this research that may suggest opportunities for further research.

Section 1 - The Disk Drive Industry

The history of the disk drive Industry is short but yet all the more impressive. This section serves the purpose of introducing the common reader to this industry and its fast-paced development compared to other commerce.

Emerging from the necessity of reliable information storage the disk drive originated in the San Jose laboratories of IBM, Inc. (International Business Machines) in 1956 as part of a research project. Today this project has developed into a commerce creating 33.4 billion US dollars of revenue in 2010 with forecasts predicting an increase of up to 25 percent until 2015 (Shilov, 2011). Constant product innovation, disruptive technological change and emergence of new market segments within this industry have led to a shakeout period and continuous change in market leadership (Christensen, 1993).

IBM commercialized the first computer with a moving head hard disk drive (the IBM 305 RAMAC) with the idea to support accounting methods in electronics via RAMAC, the random access method of accounting and control (IBM video, 2007). The entire computer unit required a protective location demanding an area of 9 meter by 15 meter (Appendix: Picture 1 – The BM 305 RAMAC), where the disk drive itself necessitated 1.5 meters of space and could hold 5 megabytes of information (Levy, 2006). By means of this radical innovation IBM preceded its competitors and was the first company to enter into the disk drive market.

Prior to the RAMAC, magnetic drum technology was employed which yielded the invention of magnetic tape, becoming the dominant form of data storage. However, the shortcoming of magnetic tape was the long access time. Whenever information was retrieved, the tape had to be read from the beginning to the spot where the required information was stored. Furthermore, changing any information on the tape required a rewrite of the entire tape. The moving head of the RAMAC succeeded in overcoming this problem by randomly accessing the data (Christensen, 1993).

The innovation of the random access method allowed IBM to gain a monopolistic position in the disk drive market. Having been faced with several anti-trust law suits already, IBM introduced a protective method for their innovations called defensive disclosure which enabled them to maintain their monopoly until about 1976.

In 1964 IBM introduced the System 360 employing the 1311 model disk drive. This allowed a wider range of applications such as commercial and scientific use as well as applications for home-owners and on a business-scale. However, the demand and profitability drew competition into the market and IBM's 100 percent market share fell to 94 percent in 1970 (Bhaskarabhatla, 2012). Thus, the market structure of the disk drive industry began to change along with another antitrust case filed against IBM for anti-competitive behavior.

In 1973, IBM, still a major player, introduced its Model 3340 also called 'Winchester' which still remains the dominant design until today, even though it has been improved in several aspects over the course of time. The Winchester model allowed a large increase in information density while at the same time decreasing the cost price per piece (Christensen, 1993).

The disruptive innovation of plug-compatible storage systems (Interchangeability of components) commenced the entrance of new firms into the industry. Originally, eleven firms had pioneered alongside IBM, two of which were start-ups and nine had been industry-related companies. But between 1975 and 1989 about 500 new firms joined the market of which a minimum of 87 were original equipment suppliers (Christensen, 1993). Christensen decided upon the extensively growing minicomputer industry as the cause for the comprehensive growth in the original equipment industry for disk drives and categorized five kinds of entrants into the industry: start-ups, related-technology firms, related-market firms, forward integrators and vertically integrated computer manufacturers (Appendix: Picture 1 - Entrance and exit into and out of the industry between 1976 and 1989). While Japanese firms proved to be dangerous competition for the U.S.-based companies, European firms never represented a serious thread.

The rapid increase in competition between 1975 and 1989 in the disk drive industry led to significant improvements in all aspects of the technology such as recording density. Average data access time declined as well as an enormous drop in price per megabyte of memory (appendix: Picture 2 - Experience curve of the industry). Along with these improvements also the size of the design model was reduced: 14 inches to 8 inches in 1978; 8 to 5.25 inches in 1980; 5.25 to 3.5 inches in 1985; and 3.5 to 2.5 inches in 1989 (Christensen, 1993) (appendix: Picture 5 - Different sizes of disk drives). The size has now been reduced to 0.85 inches by Toshiba in 2004 with a maximum capacity of two gigabytes in 2004 (Farrance, 2006).

The year 1984 marks a special occurrence for the disk drive industry. The end of 1983 and 1984 had been predicted as the beginning of a shakeout period by contemporary literature (Sahlman, 1985) and been confirmed as such by later research (Lerner, 1997).

The first crisis for the disk drive industry did not occur as is often the case in technological industries due to demand shortages but was linked to the 1997 Asian financial crisis leading to an austere cyclical decline in the disk drive industry (Doner, 2001). However, this impediment did not persist and growth rates in the disk drive market continued to be impressive. While the world economic crisis starting in 2007 took its expected toll on the industry by lowering the demand in the disk drive market and thus forcing prices down, the year 2011 proved to be even more severe to manufacturers. Two natural disasters in Thailand, one of the main producing countries of hard drives, caused a major economical setback to the entire industry. The floods in Thailand led to extreme shortages in disk drive shipments (up to 25 percent less production) leading to a strong increase in prices. Even though some retailers welcomed this development others realized the need for equilibrium: “In many respects, the hard disk drive industry has collectively hit the 'reset' button,” said John Rydning, research vice president, Hard Disk Drives at IDC (IDC, 2012). Rydning refers to the opportunity to steady the previous disparity between demand and supply and anticipates a slow price decrease along with investment and continuous revenue creation.

Despite the supply shortage that will continue to affect customers well into 2013 (Connor, 2012) industry revenues have been predicted to approach the 50 billion U.S. dollar mark in 2016, provided that the hybrid solid state hard drive, introduced by Seagate and Samsung in 2007, will be commercially functional by then (IDC, 2012).

Section 2 - Literature Review

The disk drive industry has been used in several strings of diverse academic literature because of its ‘fruit-fly’-characteristic and many conclusions have been drawn from this research. However, the opportunities that lie within the subject of labor mobility have only rudimentarily been studied by a few researchers so far and there is a wide range of further research that can be accomplished with it. As mentioned previously, the disk drive industry is an excellent industry to test labor mobility due to its condensed life-cycle and existence of a shakeout (Sahlman, 1985).

Data on the mobility of all technical employees is difficult to obtain. Consequently, I use a dataset of inventors in the disk drive industry in order to determine firms pioneering in new product markets attract inventive employees; there is only a very limited amount of existing literature that can be considered useful. Because of the sparse information and methodology available, papers from somewhat less related areas of research have to be taken into account as well.

The first academic who developed the idea of using patent data for economic research on a large scale was Jacob Schmookler in 1966, followed by Scherer in 1982 and Griliches in 1984. However, labor mobility research requires a considerably large amount of data which was almost impossible to gather at that point in time due to the previously mentioned limitations in information technology.

Jaffe, Trajtenberg and Hall (2001) have been considered the forerunners in creating a sufficient dataset that could link patents over time and space. This paper was published by the National Bureau of Economic Research[2]. Since, more literature on labor mobility has slowly been introduced and developed but numerous research opportunities in combination with patent data remain.

Kim, Lee and Marschke published a paper in 2006 called “International Knowledge Flows: Evidence from an Inventor-Firm Matched Data Set” (Kim, 2006). In this paper they investigate the increase of employment and collaboration of U.S.-based firms in the pharmaceutical and semiconductor industry regarding researchers possessing foreign working experience. However, the most interesting aspect of this paper is the use of patents and patent

[2] This dataset is available online at http://www.nber.org/patents/

citations as a measure of scientists' research and development productivity. Furthermore, Kim et al. point out the inconvenience of firms patenting under different names due to mergers and acquisitions, firms' name changes or listings under either the firm's parent name or subsidiary's name. These concerns are also highly relevant for the dataset used in this paper in addition to firms listed multiple times under different company divisions' names.

One of the major sources on how to approach the topic of labor mobility in academic research is the paper "Identification and Mobility of Israeli Patenting Inventors" written in 2008 by Manuel Trajtenberg from the Tel Aviv University and Gil Shiff from the National Economic Council in Jerusalem. They examine the geographical mobility of Israeli inventors as well as their mobility across assignees followed by a measurement in change of innovation quality according to those moves. While patent quality is largely ignored in this thesis, the approach on following the inventors' geographic mobility is inspiring and respectable support in conducting valuable descriptive analysis is provided. Additionally, the 'who is who' problem as they call it is central to their paper. To solve this, they try to identify the same inventors with different spellings in their data set by applying a computerized matching procedure.

This computerized matching procedure was created by Trajtenberg, Shiff and Melamed in 2006. In their paper "The 'Names Game': Harnessing Inventors' Patent Data for Economic Research" Trajtenberg et al. develop a methodology and corresponding algorithm that allows researchers to overcome confusion caused by different spellings of names and identify unique inventors across datasets. This paper is imperative in the sense that it allows researchers across the globe to create reliable datasets, which may be applied to all sorts of economic research.

These papers give a good indication as to how to approach research about labor mobility. Furthermore, they acquaint the reader with potential flaws and mistakes in existing datasets. All these aspects help to generate a dataset ridden of such defects. Nevertheless, the methodology followed in this thesis is unique in the sense of creating an additional measurement to test the hypothesis.

Section 3 – Data Collection

This thesis concentrates on the topic of labor mobility over the industry life-cycle. It asks the related question: Are pioneers more likely to attract intra-industry inventors? These are inventors that work for other firms in the same industry.

With the purpose of answering this question two datasets are employed; one dataset by Lee Fleming et al. and another one by Clay Christensen. Both of these contain information on the patent-, inventor-, and company-level in the disk drive industry.

Lee Fleming's Dataset

The most difficult task in researching the topic of labor mobility is the matter of obtaining data that provides a sufficient and complete set of information. Patent data has proven to be very beneficial for this specific topic. Unfortunately, the acquisition of such a dataset is highly problematic due to a variety of factors. First, gathering individual data on patents is a tedious task, entailing an enormous amount of time and requiring access to certain databases. Once information on patents is collected, it has to be aligned with the corresponding inventor. Regrettably, inventors do not have an individual reference number (yet)[3], this may potentially cause confusion about the unique identity of inventors. Potential threads to correct identification include different ways of spelling (foreign) names, same names for different inventors, and the inclusion or exclusion of middle names to distinguish among different inventors. Furthermore, an invention is often developed by a number of people, and not solitarily. Hence, multiple inventors are often associated with one invention. Additionally, when looking at labor mobility in particular, the assignee of the project also has to be identified. Over time these information may be disturbed by occurrences like company name changes, defaults, or mergers and acquisitions.

To find out the addresses of the assignee should usually be a task possible to achieve; this usually allows for certain research regarding labor mobility. But assignees, especially global players, often own multiple business units located in completely different locations, or an

[3] An introduction of personal reference numbers per individual inventor would make research much more straightforward and simpler. The same is true for assignees.

assignee entrusts an external laboratory with a certain project. Thus, the assignees' addresses do not necessary reflect the working premises of the scientists. Hence, the personal addresses of the inventors give a much better overview of where the scientists live and thus work. Obviously, obtaining information on such a personal level is almost impossible to gather for such multitudes of people.

For these reasons extensive datasets on patents, assignees and inventors are not only rare but also very expensive. I had the benefit of obtaining access to a dataset developed by Lee Fleming, Ronald Lai and Amy Yu, all from Harvard Business School, as well as by Alexander D'Amour and Ye Sun, both from the Harvard Institute of Quantitative Social Science. These authors developed and publicized an extensive dataset of individual inventors from the U.S. utility patent database ranging from 1975 until 2010 by means of a Bayesian supervised learning approach in a disambiguation algorithm[4]. Additionally, the dataset covers patents' co-authorship networks and social network measures.

The authors used several sources in order to develop this comprehensive dataset, namely the NBER database by Hall, Jaffe and Trachtenberg mentioned above, as well as the weekly publications of the U.S. Patent and Trademark Office (USPTO) and the 1998 Micropatent CD product. In addition, secondary data sources were included to achieve better identifying parameters. These sources comprised of the USPTO CASSIS dataset, the National Geospatial-Intelligence Agency country files, the US Board on Geographic Names and the NBER File of Patent Assignees.

The main challenge for the authors was to probabilistically combine patents and inventors, a process called disambiguation. Previously, this had been attempted via manually applied fixed weights for a diversity of variables; however, this brings forth several potential biases such as model-dependence, fixed weights leading to inaccuracy, incompleteness due to imperfect data provision as well as data forfeiture through unitless match score.

However, these problems were circumvented by applying an algorithm-variation of Torvik and Smalheiser which in this case employs a seven-dimension profile, consisting out of first name, middle initial, last name, author location, assignee, technology class and coauthors. Fleming et al. split these seven variables into two groups: name characteristics and patent characteristics. The former containing the first name, middle initial and last name, and the

[4] A complete description of the dataset by the authors themselves may be found under: Lai, Ronald, Alexander D'Amour, Amy Yu, Lee Fleming, 2010, „Disambiguation and co-authorship networks of the U.S. Patent Inventor Database, Harvard Business School, August 26th

latter containing author address, assignee, technology class and coauthors. These two groups are assumed to be independent of each other. This method was motivated by Torvik's work from 2005 and allows producing a sample of matches (or non-matches) by exploiting their independence. After applying a triplet correction step, which includes a clustering procedure by means of a threshold value and several waves of blocking and consolidation, a dataset of matched pairs (inventors and patents) was obtained.

The outcome of Fleming et al.'s work is one of the most complete and reliable datasets available and proves to be highly suitable for the econometrics applied in this paper.

Lee Fleming's Dataset - Manipulation

As mentioned above, the available dataset is sufficient for the purpose of this thesis. Nevertheless, the data has to be changed, cleaned and re-organized to suit its intended purpose. In order to work appropriately with the data, I use Stata 12.0 Special Edition as a tool. This statistical and data analysis program functions well with files of large size and is very convenient for processing datasets including a large number of variables and observations.

The dataset provided by Fleming et al. supplies information on patents registered during the 35-year period. Hence, the original dataset contains almost 6 gigabyte of information. To avoid long waiting times when processing the data, some (string) variables that are irrelevant for this research will be dropped namely 'patent', 'lat', 'lon', 'appdatestr' and 'ayear'. Then, the investors' first and last names are combined in a group labeled 'inventorid', which serves the sole purpose of a more systematic overview.

A major step in downsizing the dataset is by flagging and separating the patents regarding the disk drive industry which are identified by the US patent class number 360. To further clean up the dataset, only patents that belong to an assignee remain in the dataset, thus, patents without a relation to an organizational number (variable 'orgno') are dropped from the set. These are patents assigned to individual inventors and form a small fraction of the overall patenting in the industry. After additional cleaning processes in the dataset such as dropping duplicates or trimming edges to dispose of duplicates created by spaces, a dataset of disk drive patents remains, that includes assignees, inventors, and patents as well as detailed specifications about each of those.

However, even though irrelevant variables have been dropped and a certain standard of order has been preserved, the data still contains flaws and defects. An issue that remains with the data is imperfections in spelling and naming that lead to multiple entries of the same name, as for example: 1Vision Inc and 1Vision Software Inc, or 8 8 Inc and 8x8 Inc. Furthermore, especially large corporations do not patent under their general company name, but under the name of a certain department, for example: Siemens Medical Solutions USA INC and Siemens Medical Laboratories Inc. The consequence of such multiple entries may affect the data analysis since multiple entries will cause one company with manifold patents to appear as several companies with very few patents each. Furthermore, inventors may potentially have moved departments within a firm which does not qualify as a genuine move to another company and thus has to be corrected for[5].

Consequently, an excel sheet is created that solely contains the organizational names and numbers from the dataset. Hereafter, every firm name is inspected and cross-examined with similar entries. In case of multiple organizational numbers, I intuitively decide upon the number retained in the dataset representing the main organizational department. Through this measure the original 4134 company names and organizational numbers are reduced to 3774 individual names and numbers.

Afterwards, this excel sheet with cleaned organizational numbers is merged with the dataset described previously, leading to 126,864 matched observations and 197,235 non-matched observations; a total of 324,099 observations.

Creation of an additional measure

Clay Christensen's dataset contains additional patent- and company-relevant information such as "netincome", "netsales" and years of existence among others. This dataset is going to be deployed in order to create an additional measure. The measure is generated with the purpose of becoming the independent variable of interest explaining the dependent variable

[5] One error has been detected in the dataset. The US patent 5,266,650 appears as filed on November 30th, 1909. However, the actual date of file was November 30th, 1993 (http://patents.justia.com/1993/05266650.html). This has been corrected for.

("all_moves"). The variable incorporates the concept of innovation of a company and is supposed to measure its impact on the labor mobility of employees.

The measurement (Measure One) determines whether a firm has introduced a certain-sized diameter of disk drives to the market and is named "new_drives". This means that this specific diameter may have been introduced to the market previously by somebody else but the firm decided to compete with this specifically sized product. Hence, this variable allows a measurement of the influence of incremental innovation on labor mobility.

The measure is combined with Lee Fleming's dataset on the basis of the companies' names. Unfortunately, Fleming's and Christensen's dataset contain different variables to identify the company ("assignee" and "namefranco2"). With the aim of merging both sets of data, an Excel sheet is created in which both variables are manually cross-examined to match the same companies and control for dissimilar spelling.

Hence, the two datasets and measure one are merged and generate a dataset apposite to run regressions. These regressions are based on the regression model which is developed in "Section 5 – Model".

Section 4 – Descriptive Statistics

This descriptive analysis of Lee Fleming’s dataset gives comprehensive insides about the composition of the data and may lead to intuitive thoughts about the extent of labor mobility over the life-cycle.

The reader must be aware that this dataset is by far not complete. Only a small fraction of the employees that have invented certain products are included in this dataset. Yet, preliminary conclusions can be drawn from this analysis.

The entire dataset is composed of 324,099 observations with 10,105 individual companies which are or were active in the disk drive industry (Appendix: Figure 1- Number of firms). Naturally, these firm appearances are and were subject to constant entry and exit, with firms leaving or entering the market as well as merging with each other.

When observing the annual progress of the industry hitherto a cyclical development of the number of firms can be witnessed (Appendix: Figure 2 - Active companies per year). Beginning in 1957 with IBM it took twelve years until investors realized the potential of this market and competitors were able to enter the market. From 1969 onwards, the numbers of active firms kept increasing and rose steadily until peaking in the year 2001 with only minor exceptions in rare cases (year 1979, year 1990, year 1998, year 1999). With the rupture of the internet bubble in 2001, the market participants’ number declined rapidly from 1720 firms, representing 5.19% of firms in the entire dataset, down to only 24 firms (0.07%) in 2008.

The dataset contains 23,121 individual inventors (Appendix: Figure 3 - Number of unique inventors). Each of these inventors created an average of 10,2 patents while an average of 23,4 patents were assigned from by each company (Appendix: Figure 4 - Patents per inventor/assignee).

After this initial descriptive analysis of Lee Fleming’s dataset is completed, this dataset is merged with Clay Christensen’s dataset and measure one. Thus, patent data and firm-level data are combined in order to achieve a suitable dataset for the hypothesis testing.

Measure one identifies the timing of 22 different sizes of diameter introductions to the market by different companies throughout the industry cycle. While some diameters are only used by very few companies (3.5mm by one firm; 5.32mm by seven firms) other diameters are used by a wide range of companies (5.2mm by 3,134 firms; 14mm by 2,817 firms). In total 344 companies have participated in the market with at least one new drive diameter produced by themselves.

As has been identified by previous literature, the year 1983/1984 marks the beginning of the shakeout period in the disk drive industry (see appendix: Picture 3 - Entrance and exit into and out of the industry between 1976 and 1989). Thus, special attention should be paid to the pre- and post-shakeout year.

When observing the amount of moves individuals have undertaken per year, the year 1984 indeed highlights a special occasion (see appendix: Figure 7 - Moves per year). From 1969 until 1983 the cumulative percentage of moves totals 17,38 %. The year 1984, the year of the beginning of the shakeout, contains fewer moves than the previous year (from 92 moves down to 88); this marks a unique occurrence in the entire time period (except for the year 1992). The following thirteen years (1985 until 1997), which represents half of the available time period in the dataset, contain 80,18 % of the total moves of employees. The sum of moves per year almost fourfold within this relatively short period of time from 109 moves in 1985 to 427 moves in 1997. This suggests that the shakeout has led to a much higher activity of labor movement. However, this descriptive statistic is merely an indication as further (and more complete) data on the inventors is missing. Nevertheless, this preliminary insight encourages the upcoming model building.

Section 5 – Model & Results

The purpose of this study is to investigate whether inventive employees tend to move to companies that pioneer in new product markets. Therefore, the model applied for the regression analysis is a basic multiple regression model. The dependent variable "all_moves" describes the total number of movements per inventor from one company to another. The previously created variable measure one ("new_drives") is the independent variable of interest for the regressions. In order to control for external factors additional variables existing in the dataset will be included in the regression model. External factors such as company fixed effects and patent fixed effects may have a severe influence on employees' choice of switching employers.

$$\boldsymbol{Prob(all\_moves) = \alpha + \beta_0\, new\_drives + \sum_k Company_k + \sum_m Patent_m + \epsilon}$$

The dataset provides a variety of possibilities to control for some of these external factors. Thus, the variables "netsales", "netoutput", "netincome", "captive", "oem" and "maxdia" may potentially be included in the regression model. While "netsales", "netoutput" and "netincome" all control for the size of a company, "captive" and "oem" both control for the type of a company; whether it is an original equipment manufacturer (oem), a supplier of an oem, or perhaps both. "Maxdia" on the other hand controls for the diameter size of the patent of a company.

$$\begin{aligned}\boldsymbol{Prob(all\_moves)}&\\ &\boldsymbol{= \alpha + \beta_0\, new\_drives + \beta_1 netsales + \beta_2 netoutput + \beta_3 netincome}\\ &\boldsymbol{+ \beta_4 captive + \beta_5 oem + \beta_6 maxdia + \epsilon}\end{aligned}$$

Incipiently, the correlations among each other give a good indication of which variables are most suitable for the regression analysis since some variables control for the same concept.

TABLE 1: CORRELATIONS AMONG VARIOUS MEASURES

	all_mo~s	new_dr~s	netsales	netout~t	netinc~e	captive	oem	maxdia
all_moves	1.0000							
new_drives	0.0257	1.0000						
netsales	0.1498	-0.1087	1.0000					
netoutput	-0.0810	-0.1362	-0.2045	1.0000				
netincome	0.0165	0.0246	0.3576	-0.1361	1.0000			
captive	0.0764	0.1074	-0.0094	-0.2307	-0.0457	1.0000		
oem	0.0306	-0.1335	0.1736	0.2045	0.0850	-0.4252	1.0000	
maxdia	0.0112	0.0651	-0.2517	-0.1439	-0.1512	0.3428	-0.6408	1.0000

As can be derived from table 1 there is no high correlation amongst any of the variables (Moore, 2003). However, the variables controlling for firm size (“netsales”, “netincome”, “netoutput”) have a low to moderate correlation with each other. The same applies to the variables controlling for the type of company that correlate negatively with each other (“captive”, “oem”) Furthermore, “maxdia” tends to correlate moderately with most other variables.

When specifying the regression models, the correlations have an impact upon which control variables to include. Five models are employed to see the development of measure one with the control variables. Specification one solely contains the independent variable (and the constant). The second specification includes all variables that control for the company-size. Adding the ‘type-of-company’ measures specifies model three and including all variables constructs model four. However, as was indicated by the correlation table, some of the variables correlate moderately with each other. Thus, the best model comprises of the control variables “netsales”, “netincome”, “oem”, and “maxdia”. The variables “netoutput” and “captive” are ignored because they serve to control for the same concepts yet disturb the regression outcome.

Initially, a standard OLS-regression is run as a preliminary overview of the distribution of the dataset.

TABLE 2: ORDINARY LEAST SQUARES REGRESSION

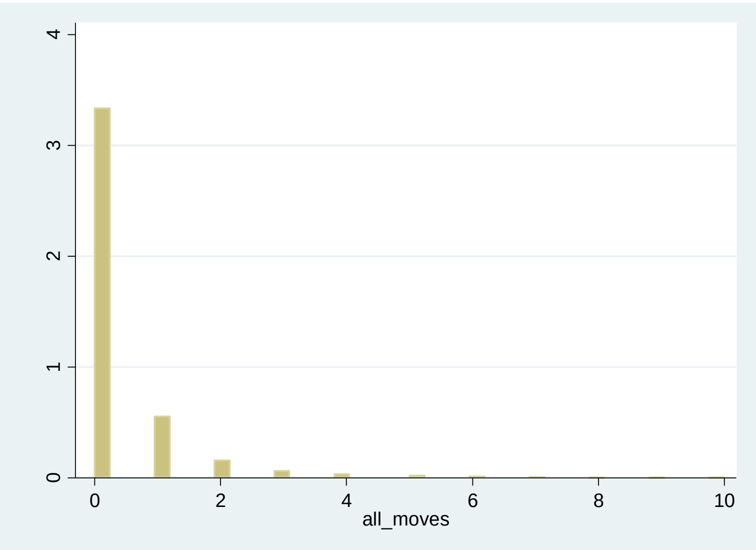
	Model 1	Model 2	Model 3	Model 4	Best Model
New_drives	0.034 [0.0261]	0.057* [0.0294]	0.054* [0.0295]	0.059** [0.0295]	0.062** [0.0270]
Netsales		0.000*** [0.0000]	0.000*** [0.0000]	0.000*** [0.0000]	0.000*** [0.0000]
Netoutput		0.000** [0.0000]	0.000** [0.0000]	0.000* [0.0000]	
Netincome		0.000** [0.0000]	0.000** [0.0000]	0.000** [0.0000]	0.000** [0.0000]
Captive			0.176*** [0.0435]]	0.162*** [0.0437]	
Oem			0.137*** [0.0511]	0.236*** [0.0613]	0.185*** [0-0565]
Maxdia				0.017*** [0.0058]	0.025*** [0.0054]
constant	0.192*** [0.0137]	0.126 *** [0.0262]	-0.020 [0.0516]	-0.208** [0.0822]	-0.210*** [0.0759]

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

The outcome of the OLS regression analysis, illustrated in table 2, shows a constant improvement of the significance of the dependent variable of interest per model specification. While model 1 shows a positive coefficient estimate for “new_drives” it proves not to be significant. Over the course of further specifications this coefficient becomes significant at the 10-percent level in model 2 and model 3. In model 4 it becomes significant at the 5-percent level. The best model specification remains significant at the 5-percent level and provides a coefficient value of 0.062. The other specification variables are even significant at the 1-percent level (with the exception of “netincome”, which is significant at the 5-percent level) and also maintain a positive sign. Only the constant is observed to have a negative sign with a coefficient of -0.210.

These initial regressions suggest that once a company introduces a new drive to the market (it may have been introduced by a competitor before), this increases the probability of moves by more than six percent with a standard error of almost three percent. However, a valid concern exist that the dependent variable is not normally distributed. This can easily be tested by means of a histogram.

GRAPH 1: THE DISTRIBUTION OF THE DEPENDENT VARIABLE “ALL_MOVES”



As can be seen in Graph 1, “all_moves” is strongly skewed to the right and thus not normally distributed. Hence, the OLS regression is actually an inappropriate tool. Since moves is a positive count variable, several options are available such as a tobit regression, poisson regression or a negative binomial regression.

Since count data, which is present in this case, often follows a poisson distribution, this will be tested first in order to determine the correct distribution of the data. One assumption of the poisson distribution is the equality of conditional mean and conditional variance. However, the summary statistic of the "all_moves" variable in table 3 confirms a variance almost three times larger than the mean.

TABLE 3: SUMMARY STATISTIC OF "ALL_MOVES"

all_moves

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	17205
25%	0	0	Sum of Wgt.	17205
50%	0		Mean	.3676257
		Largest	Std. Dev.	.9795564
75%	0	10		
90%	1	10	Variance	.9595308
95%	2	10	Skewness	4.331906
99%	5	10	Kurtosis	27.57721

This indicates signs of overdispersion (or underdispersion) but a goodness-of-fit test will provide certainty. This is conducted after running a poisson regression in the original model.

```
Pearson goodness-of-fit  =  5350.883
Prob > chi2(2625)        =    0.0000
```

The large value of the goodness-of-fit test in combination with a significant p-value supports the previous speculation that the Poisson distribution may not be an appropriate method. Since overdispersion (or underdispersion) is present, a negative binomial regression may suit the distribution better. When applying the negative binomial regressions the likelihood ratio test of each regression confirms that alpha is significantly different from zero and fortifies the pertinence of the negative binomial distribution.

The five model specifications deployed for the regressions remain the same as in the OLS analysis and will continue to remain the same for further investigation.

TABLE 4: NEGATIVE BINOMIAL REGRESSION

	Model 1	Model 2	Model 3	Model 4	Best Model
New_drives	0.155	0.196	0.173	0.173	0.241*
	[0.1196]	[0.1292]	[0.1286	[0.1292]	[0.1255]
Netsales		0.000***	0.000***	0.000***	0.000***
		[0.0000]	[0.0000]	0.0000]	[0.0000]
Netoutput		0.000***	0.000***	0.000***	
		[0.0000]	[0.0000]	[0.0000]	
Netincome		0.000**	0.000*	0.000*	0.000*
		[0.0000]	[0.0000]	[0.0000]	[0.0000]
Captive			0.696***	0.614***	
			[0.1730]	[0.1756]	
Oem			0.583*	0.963***	0.844***
			[0.2398]	[0.2798]	[0.2727]
Maxdia				0.067***	0.109***
				[0.0256]	[0.0238]
constant	-1.648***	-1.92***	-2.606***	-3.342***	-3.545***
	[0.0654]	[0.1191]	[0.2490]	[0.3754]	[0.3594]

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

When applying the negative binomial regression only the best model provides a significant coefficient for the main variable of interest. As can be derived from table 4, “New_drives” is significant at the 10-percent level and the coefficient has a value of 0.241. With a standard error of 0.1255, the coefficient is positive and thus reflects that pioneering firms attract more employees. The other variables are significant at the 1-percent level (except for netincome, which is significant at the 10-percent level) and have a positive sign. Especially “oem” takes a large value with a coefficient of 0.844. However, the constant once again has a large negative coefficient. In comparison to the OLS regression, signs and relative magnitudes of the coefficients have been preserved.

The regression output confirms the original hypothesis of this thesis: Pioneering firms are more likely to attract inventive intra-industry employees. This is confirmed by the previously stated condition that the coefficient of the variable of interest needs to be positive throughout the regressions.

Section 6 – Robustness test

The preceding finding is sensitive to a variety of factors. Thus, a few robustness tests are in order to strengthen the results.

Different diameter sizes

When preparing measure one for the regression analysis, I realized the importance of the variable “maxdia” which was deployed to specify the pioneering firms. As investigated in the descriptive analysis, measure one identifies twenty-two confirmed sizes of disk drive diameters that have been introduced to the market. However, some diameter sizes are rarely manufactured while others are produced extensively. In order to test whether this may have an effect on the outcome of the regression analysis, I combine seldom used sizes of disk drives with the commonly used sizes. I define major sizes as any diameter that exceeds a share of one percent of the overall sample. This includes nine different specifications, namely: 1.92mm, 2.6mm, 3.8mm, 5.2mm, 8mm, 8.4mm, 9.2mm, 10.5mm, and 14mm. When the share of a diameter fails to be larger than one percent the observation is added to the next higher acceptable diameter size. Moreover, the size constriction will be tightened even further in the next step and only diameter shares around five percent will be taken into account. Lastly, only diameters with a share of above ten percent are included in the last specification. Then I repeat the negative binomial regression on the best model.

TABLE 5: ROBUSTNESS TO SIZE DEFINITION

# of diameters	Best Model (“new_sales”)
22 sizes	0.241*
	[0.1255]
9 sizes	0.233*
	[0.1299]
6 sizes	0.349**
	[0.1431]
3 sizes	0.284
	[0.2160]

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 5 demonstrates that a narrower constriction of diameter sizes initially leads to an even better outcome of the coefficients. Especially the specification with six sizes gives significance at the 5-percent level and a positive coefficient with a value of 0.349. However, the narrowest specification proves to be insignificant. This was somewhat expected as it would limit measure one too much and lead to a very biased "new_drives" variable. Still, the other specifications show, that combining rarely manufactured diameter sizes with commonly produced ones, improves the regression output. Nevertheless, the original approach to treat all diameter sizes separately from each other still gives sufficient results for the analysis.

Tobit regression

A very important factor in statistical analysis is the omission of variables in a dataset. This can occur by censoring or truncation. The variable "all_moves" has been purposefully truncated at ten moves, because otherwise, this may have led to biased results. Hence, I employ a Tobit regression in order to test for the potentially truncated data. Even though the Tobit regression has a variety of dissimilar assumptions to the negative binomial regression, it will give a good indication of whether there might be a problem with truncation.

TABLE 6: TOBIT REGRESSION

	Model 1	Model 2	Model 3	Model 4	Best Model
New_drives	.171	.235	0.238	0.253	0.356*
	[0.1813]	[-0.1951]	[0.1955]	[0.1951]	[0.1869]
Netsales		0.000***	0.000***	0.000***	0.000***
		[0.0000]	[0.0000]	[0.0000]	[0.0000]
Netoutput		0.000***	0.000***	0.000**	
		[0.0000]	[0.0000]	[0.0000]	
Netincome		0.000**	0.000*	0.000*	0.000*
		[0.0000]	[0.0000)	[0.0000]	[0.0000]
Captive			1.069***	0.910***	
			[0.2705]	[0.2726]	
Oem			1.106***	1.668***	1.428***
			[0.3736]	[0.4209]	[0.4141]
Maxdia				0.109***	0.171***
				[0.0374]	[0.0358]
constant	-3.695***	-3.960***	-5.105***	-6.233***	-6.623***
	[0.1377]	[0.2708]	[0.4584]	[0.6208]	[0.6103]

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Interestingly, the model produces almost identical outcomes as the negative binomial regression. Although the first four models are all insignificant, the relevant coefficient takes a positive value throughout. Only the best model proves to be significant at the 10-percent level with a coefficient of 0.356 and a standard deviation of 0.1869. The other variables are mostly significant at the 1-percent level and only the constant has a negative value. Since these results are very similar to the negative binomial output, I conclude that potential truncated data has no major influence on the results.

Section 7 – Conclusion

In their quest for ideal data and information, economic researchers have been conducting and expanding their investigations into various industries, periods of time, and geographic locations.

In the past, certain fields of economic research have been impeded by the difficulties of attaining data. However, with the introduction of information technology, this has changed fundamentally. One of the fields of study that has undergone such a transformation is labor mobility and I have argued that the disk drive industry is most suitable for such research. On the one hand, this industry has a very short life-cycle of only 30 years. On the other hand, it passes through a period of a shakeout succeeding several introductions of disruptive technologies.

I was able to create a variable measuring the innovativeness of firms through usage of existing data bases. “New_drives” quantified innovation by identifying the introduction of new disk drive diameters to the market.

Consequently, I was able to test the implications of innovation on labor mobility and in particular whether inventive employees tend to be more attracted by pioneering firms. The findings of this thesis confirm this hypothesis. Hence, this research is a valuable addition to both existing literature on labor mobility and to the analysis of the disk drive industry. The innovation of firms has proven to be partially responsible for the decision of inventive employees to favor one employer over another. Despite the convincing outcome of this research, the result is subject to a variety of limitations, such as the following:

In this thesis I defined labor mobility as an intra-industry transfer of an inventive employee from one company to another. Hence, this definition focuses on geographic mobility. Occupational mobility is mostly ignored as the first employment within the disk drive industry is dropped as an observation from the dataset and no record of pre-retirement exits is kept. Including occupational mobility in the data seems feasible (yet difficult) and may allow for an inter-industry analysis of the underlying factors. Furthermore, coerced labor movements may have a large influence on the results of this research. While this research demonstrates that employees tend to move to more innovative firms, this study utilizes all available observations. It would be very interesting to build a dataset consisting only of voluntary labor movements. While very difficult to accomplish, this would lead to results not distorted by coerced moves.

Further limitations to this model are present in terms of endogeneity. The most likely reason for endogeneity in this thesis may arise due to omitted variables in the regressions. While the model controls for company size, type of company and patent size, there are several more external factors that may have an influence on the attractiveness of companies to potential employees in addition to their pioneering effort.

Some of the factors of endogeneity may be eliminated by the availability of a more complete dataset. This may include a larger number of inventors (and a better specification of individuals such as individual ID) but could also contain further information on the patent level, for example.

In my opinion a major reason for favoring one firm over the other is the perceived quality of a firm: While quality is hard to measure (or even to define), sub-variables could be created as a substitute measure. In particular, variables such as years of existence in the industry (as a measure of steadfastness), location of the company (quality of living for employees), and supplementary benefits to wage categories may be deployed in order to control for further factors.

Such additional research based on this model is greatly welcomed and would enhance the explanatory value of this model. Next to this particular suggestion, numerous research possibilities remain within the datasets and measures deployed in this thesis.

Sources

Bhaskarabhatla, Ajay, 2012, “Competition, Innovation, and Antitrust in the Disk Drive Industry”, January 19th, Erasmus School of Economics, preliminary paper

Castellano, Robert, 2008, “Hard Disk Drive Industry Not Immune To Economic Downturn” URL: http://seekingalpha.com/article/105023-hard-disk-drive-industry-not-immune-to-economic-downturn (Retrieved: 29.05.2012)

Christensen, Clayton, 1993, “The Rigid Disk Drive Industry: A History of Commercial and Technological Turbulence”, URL: http://journals.cambridge.org/download.php?file=%2FBHR%2FBHR67_04%2FS0007680500000088a.pdf&code=4dc0a7b201f4895d533e435a6523a1b7 (retrieved: 21.05.2012)

Christensen, Clayton, 1997, „The innovator’s dilemma”, URL: http://www.businessweek.com/chapter/christensen.htm , Harvard Business School Press (retrieved: 16.05.2012)

Connor, Deni, 2012, “Hard Drive Prices Rise Due To Tai Floods”, InformationWeek, January 9th, URL: http://www.informationweek.com/news/storage/data_protection/232301534 (Retrieved: 29.05.2012)

Doner, Rick, Bryan Ritchie, 2001, “Economic Crisis And Technological Trajectories: Hard Disk Drive Production In Southeast Asia”, Center for International Studies, Massachusetts Institute of Technology, June 20th, URL: http://dspace.mit.edu/bitstream/handle/1721.1/16568/JP-WP-01-06-52125284.pdf?sequence=1 (Retrieved: 29.05.2012)

Farrance, Rex, 2006, “Timeline: 50 Years of Hard Drives”, URL: http://www.pcworld.com/article/127105/timeline_50_years_of_hard_drives.html (Retrieved: 29.05.2012)

Financial Times Lexicon, 2012, URL: http://lexicon.ft.com/Term?term=shakeout (retrieved: 20.06.2012)

Griliches, Zvi, 1984, “R&D, Patents, and Productivity”, NBER Conference proceedings, University of Chicago Press

IBM video, 2007, commercial video, “IBM 305 RAMAC”, URL: http://www.youtube.com/watch?v=zOD1umMX2s8 (retrieved: 22.05.2012)

IDC, 2012, “IDC Forecasts Hard Disk Drive Industry Will Return to Growth After a Difficult 2011”, URL: http://www.idc.com/getdoc.jsp?containerId=prUS23414912 (Retrieved: 29.05.2012)

Jaffe, A.B., B.H. Hall, M. Trajtenberg, 2001, „The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools”, NBER Working Paper 8498

Kim, Jinyoung, 2006, Sangjoon John Lee, Gerald Marschke, “International Knowledge Flows: Evidence from an Inventor-Firm Matched Data Set”, University of Buffalo

Lerner, Josh, 1997, “An Empirical Exploration of a Technology Race”, The RAND Journal of Economics, Vol. 28, No. 2 (Summer, 1997), pp. 228-247

Levy, Steven, 2006, “The hard disk drive that changed the world”, URL: http://www.thedailybeast.com/newsweek/2006/08/06/the-hard-disk-that-changed-the-world.html (retrieved: 22.05.2012)

Moore, David, George McCabe, William Duckworth, Stanley Sclove, 2003, “The Practice of Business Statistics”, New York: W.H. Freeman and Company

Sahlman, William A., Howard H. Stevenson, 1985, “Capital Market Myopia”, Harvard Business School

Scherer, F.M., 1982, “Inter-Industry Technology Flows and productivity Growth”, Review of Economics and Statistics, 64, November

Schmookler, Jacob, 1966, “Invention and Economic Growth”, Cambridge: Harvard University Press

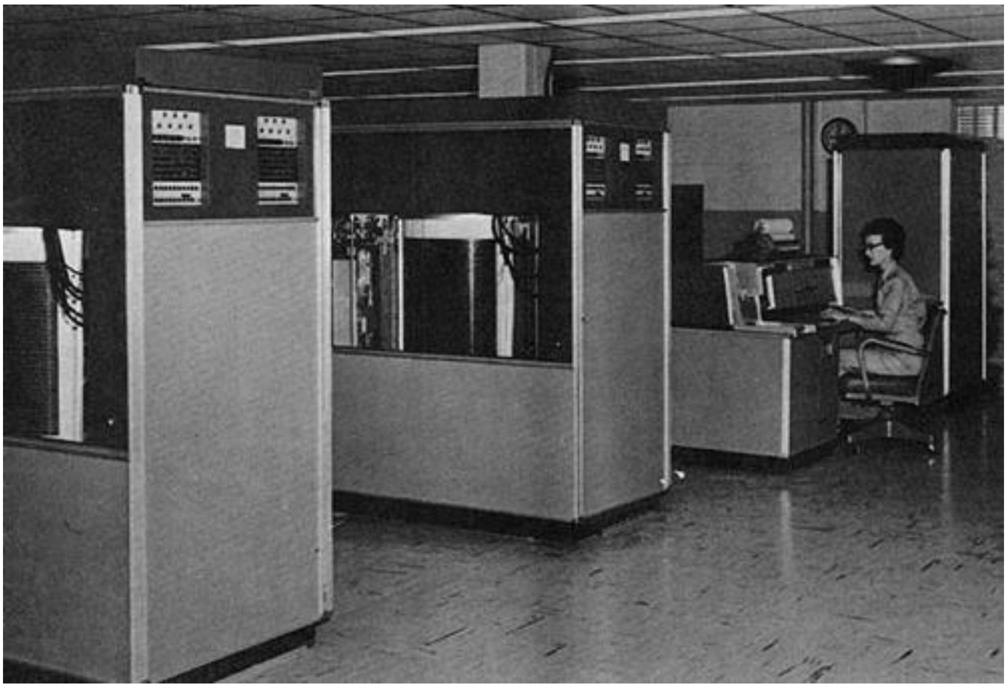
Shilov, Anton, 2011, “Hard Disk Drive Industry Retains Tremendous Expansion Opportunities – IDC”, URL: http://www.xbitlabs.com/news/storage/display/20110610060506_Hard_Disk_Drive_Industry_Retains_Tremendous_Expansion_Opportunities_IDC.html (retrieved: 21.05.2012)

Trajtenberg, Manuel, Gil Shiff, 2008, “Identification and Mobility of Israeli Patenting Inventors” The Pinhas Sapir Center For Development Tel Aviv University

Trajtenberg, Manuel, Gil Shiff, Ran Melamed, 2006, “The ‘Names Game’: Harnessing Inventors’ Patent Data For Economic Research”, National Bureau of Economic Research, Cambridge

Appendix

Picture 1 - The IBM 305 RAMAC



URL: http://www.theregister.co.uk/2011/09/16/ramac_55_year_anniversary/print.html

Picture 2 - Entrance and exit into and out of the industry between 1976 and 1989

Table 2

Number of U.S.-Based Firms Entering, Exiting, and Participating in the Rigid Drive Market, by Type of Firm, 1976–1989

NUMBER OF ENTRANT FIRMS	*Pre-1977*	*1977*	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*	*1989*
Start-Ups	0	0	1	4	7	3	9	4	2	3	3	3	0	1
Related-Technology Firms	4	2	1	0	1	1	0	1	0	0	0	0	0	0
Related-Market Firms	10	1	1	1	1	1	1	1	1	1	0	1	0	0
Forward Integrators	0	0	2	0	0	1	0	2	0	0	0	0	1	0
Vert. Integrated Mfrs. in OE/PC Market[a]	3	0	1	0	1	0	1	0	1	1	1	0	0	0
Vert. Integrated Mfrs. in Captive Production (includes firms just above)	8	1	0	1	0	0	0	1	0	0	0	0	0	0
Subtotal—U.S. Entrant Firms	**22**	**4**	**6**	**6**	**9**	**6**	**10**	**9**	**3**	**4**	**3**	**4**	**1**	**1**
Independent Japanese Entrants	2	0	0	0	1	5	0	2	9	3	2	1	0	0
Vertically Integrated Japanese Entrants	5	0	0	0	0	0	0	0	1	0	0	0	0	0
Independent European Entrants	3	0	1	2	1	0	0	0	0	1	0	0	0	0
Vertically Integrated European Entrants	3	0	2	0	0	0	0	0	0	0	0	0	0	0
Total Entrants—World Industry	**35**	**4**	**9**	**8**	**11**	**11**	**10**	**11**	**13**	**8**	**5**	**5**	**1**	**1**
NUMBER OF ACTIVE FIRMS														
Start-Ups	0	0	3	7	13	15	23	23	24	24	24	21	16	16
Related-Technology Firms	4	6	7	7	9	8	8	8	8	8	5	5	3	2
Related-Market Firms	10	11	11	11	12	10	10	10	9	9	7	4	2	2
Forward Integrators	0	0	2	2	1	1	1	3	3	3	2	1	1	1
Vert. Integrated Mfrs. in OE/PC Market	3	3	4	4	4	3	4	3	3	3	4	3	3	2
Vert. Integrated Mfrs. in Captive Production (includes firms just above)	8	9	9	10	10	10	10	10	9	8	6	6	6	6
Subtotal—Active U.S. Firms	**22**	**26**	**32**	**38**	**45**	**44**	**52**	**54**	**53**	**51**	**43**	**37**	**30**	**27**

Table 2 (continued)

Active Japanese Independents	2	2	2	2	3	8	7	8	16	18	19	15	7	5
Active Japanese Vert. Integrated Mfrs.	5	5	5	5	5	5	5	5	6	6	6	6	6	6
Active European Independents	3	3	3	3	4	6	7	4	4	5	4	4	3	2
Active European Vert. Integrated Mfrs.	3	3	3	3	5	5	5	4	4	4	4	3	3	1
Total Active Firms in World Industry	**35**	**39**	**45**	**51**	**62**	**68**	**76**	**75**	**83**	**84**	**76**	**65**	**49**	**41**
WORLD MARKET SHARES OF PRIMARY GROUPS (%)	*1976*	*1977*	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*	*1989*
U.S. Start-Ups	0	0	0	0.4	1.2	5.9	7.6	23.5	26.2	27.5	31.5	37.8	39.9	56.5
U.S. Related-Technology Firms	10.2	10.6	20.5	19.2	20.8	27.6	28.6	11.8	12.3	8.1	11.3	11.2	7.0	5.2
U.S. Related-Market Firms	49.9	42.8	28.5	28.9	26.0	17.7	13.2	6.6	3.4	3.7	1.1	1.6	1.9	0.5
U.S. Forward Integrators	0	0	0	0	0	0	0	0	0	0.2	0.1	0	1.8	3.2
U.S. Vert. Int. Mfrs.' Share of OE Market	39.9	46.5	35.8	41.1	46.4	37.4	41.1	29.2	24.9	16.7	16.0	15.1	15.9	2.2
Subtotal—U.S. Firms	100	100	84.9	89.6	94.3	85.8	90.5	71.0	66.8	56.2	60.0	65.7	66.5	67.6
Independent Japanese Firms	—[b]	—	0	0	0.3	0	0	1.8	2.5	1.8	2.5	1.9	2.0	3.9
Vert. Int. Japanese Mfrs.' Share of OE Market	—	—	9.8	5.7	2.1	6.3	3.8	19.0	22.1	32.4	30.0	25.8	26.0	23.8
Independent European Firms			5.4	4.6	1.7	4.4	2.5	3.7	5.0	5.6	4.6	2.3	4.3	2.1
Vert. Int. European Mfrs.' Share of OE Market	—	—	0	0	1.5	3.5	3.2	4.5	3.7	3.9	2.9	4.3	4.3	2.1
Total—World Industry	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Size of world OE/PC Markets ($millions):	504	516	890	1,017	1,161	1,729	1,899	2,252	3,330	3,757	4,592	6,123	7,315	8,256

[a] OE/PC = original equipment/plug-compatible.

[b] Data for the size of Japanese and European participants in 1976 and 1977 could not be found. The shares given for the groups of U.S. firms in those years are, therefore, shares of U.S. firms' production only.

Source: Author's analysis of *Disk/Trend Report* data.

Source: Christensen, 1993, table 2

Picture 3 - Experience curve of the industry

10.3.1 Capacity

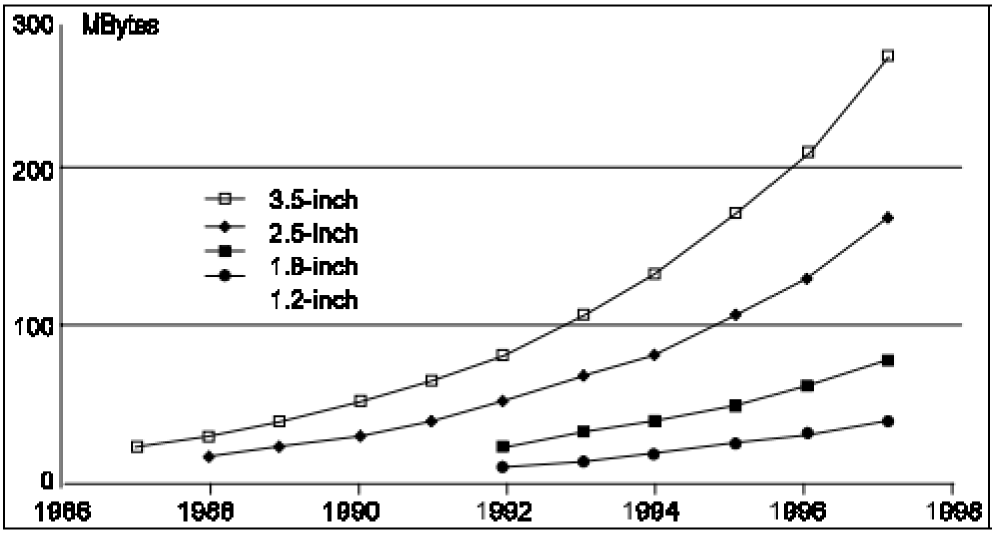
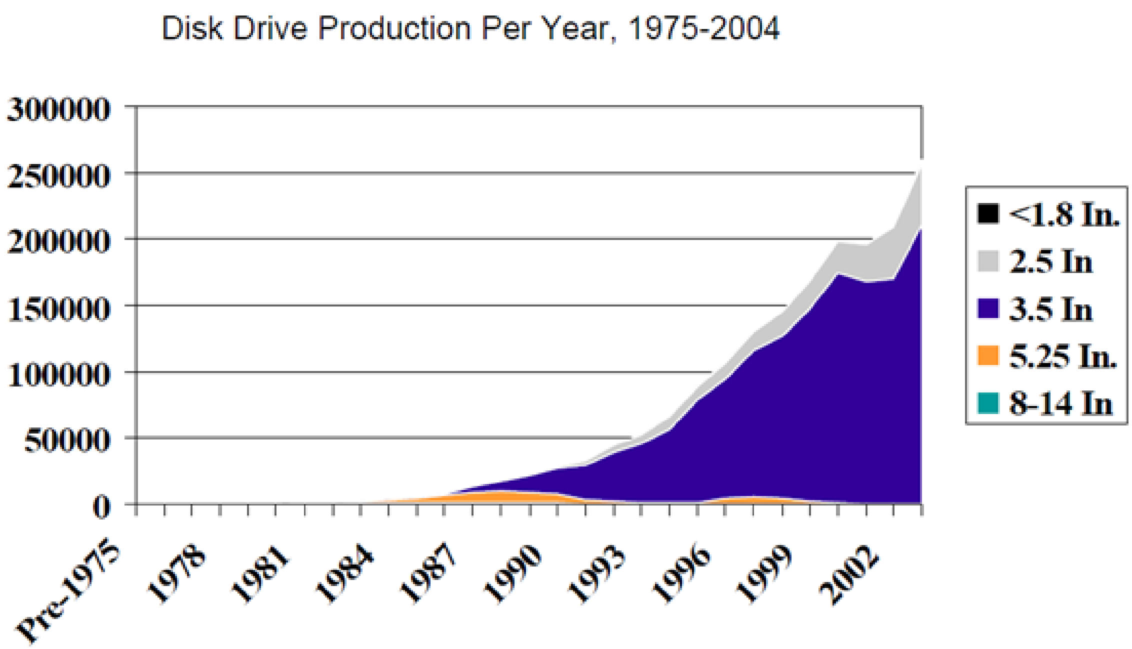


Figure 0-6 Forecast of the growth of disk capacity for a single disk.

URL: http://www.lintech.org/comp-per/10HDDISK.pdf

Picture 4 - Graphical illustration of disk drive production per year



URL: http://www.tomcoughlin.com/Techpapers/DISK%20DRIVE%20HISTORY,%20TC%20Edits,%20050504.pdf

Picture 5 - Different sizes of disk drives

Modern Disk Drives • Shown, clockwise from upper left, are a 20 megabyte full-height 5.25-inch drive used in an early IBM AT personal computer; a 20 MB 5.25-inch drive that was half that height, to allow space for an additional floppy drive in the AT; a 2.5-inch drive (40 MB) used in notebook computers; and a 3.5-inch drive used in early portable and laptop computers. (Photograph, of disk drives in the author's possession, by Ed Maltsky, Boston, Mass.)

Source: Christensen, 1993, page 36

Figure 1 - Number of firms

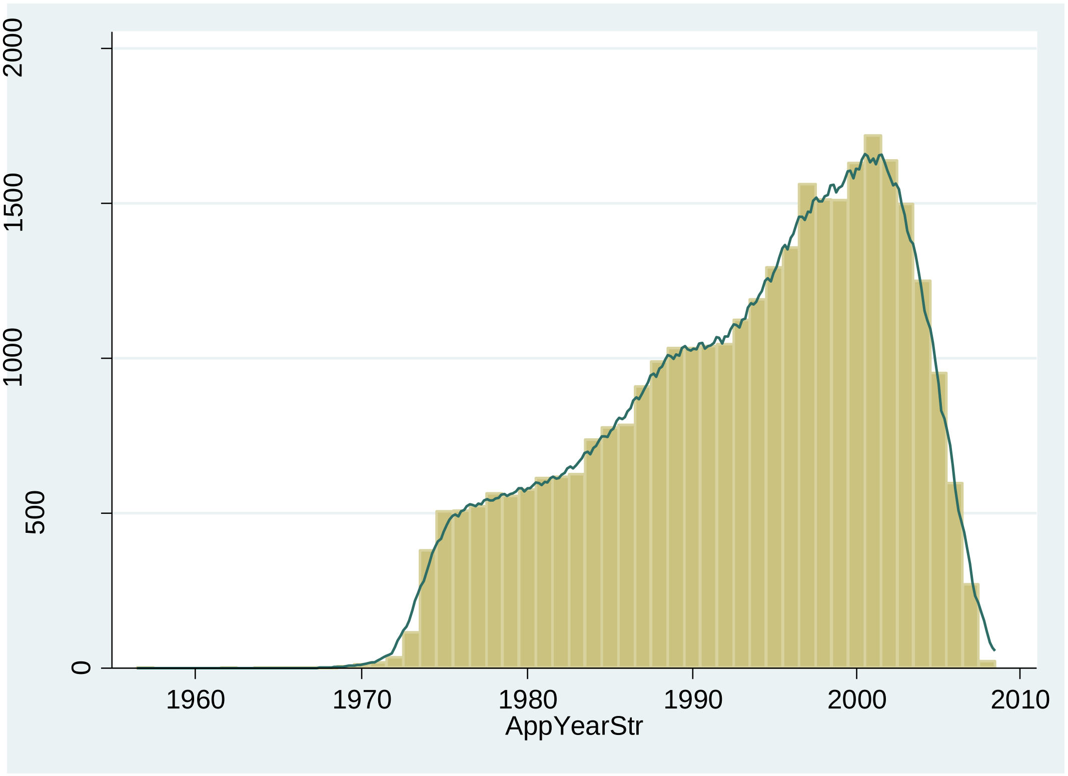
Individual firms defined by organization number

tag(orgno)	Freq.	Percent	Cum.
0	313,994	96.88	96.88
1	10,105	3.12	100.00
Total	324,099	100.00	

”

Figure 2 - Active companies per year

AppYearStr	Freq.	Percent	Cum.
1957	1	0.00	0.00
1962	2	0.01	0.01
1964	2	0.01	0.02
1965	1	0.00	0.02
1966	1	0.00	0.02
1967	1	0.00	0.02
1968	1	0.00	0.03
1969	6	0.02	0.05
1970	12	0.04	0.08
1971	20	0.06	0.14
1972	36	0.11	0.25
1973	117	0.35	0.60
1974	381	1.15	1.75
1975	506	1.53	3.28
1976	509	1.53	4.81
1977	524	1.58	6.39
1978	564	1.70	8.09
1979	556	1.68	9.77
1980	577	1.74	11.51
1981	614	1.85	13.36
1982	618	1.86	15.22
1983	627	1.89	17.11
1984	738	2.23	19.34
1985	777	2.34	21.68
1986	785	2.37	24.05
1987	910	2.74	26.79
1988	990	2.98	29.78
1989	1,034	3.12	32.89
1990	1,033	3.11	36.01
1991	1,040	3.14	39.14
1992	1,047	3.16	42.30
1993	1,124	3.39	45.69
1994	1,190	3.59	49.28
1995	1,294	3.90	53.18
1996	1,358	4.09	57.27
1997	1,562	4.71	61.98
1998	1,513	4.56	66.54
1999	1,511	4.56	71.10
2000	1,631	4.92	76.02
2001	1,720	5.19	81.20
2002	1,640	4.94	86.15
2003	1,498	4.52	90.66
2004	1,250	3.77	94.43
2005	953	2.87	97.30
2006	598	1.80	99.11
2007	272	0.82	99.93
2008	24	0.07	100.00
Total	33,168	100.00	



Graph including Kernel density plot

Figure 3 - Number of unique inventors

tag(inventorid)	Freq.	Percent	Cum.
0	300,978	92.87	92.87
1	23,121	7.13	100.00
Total	324,099	100.00	

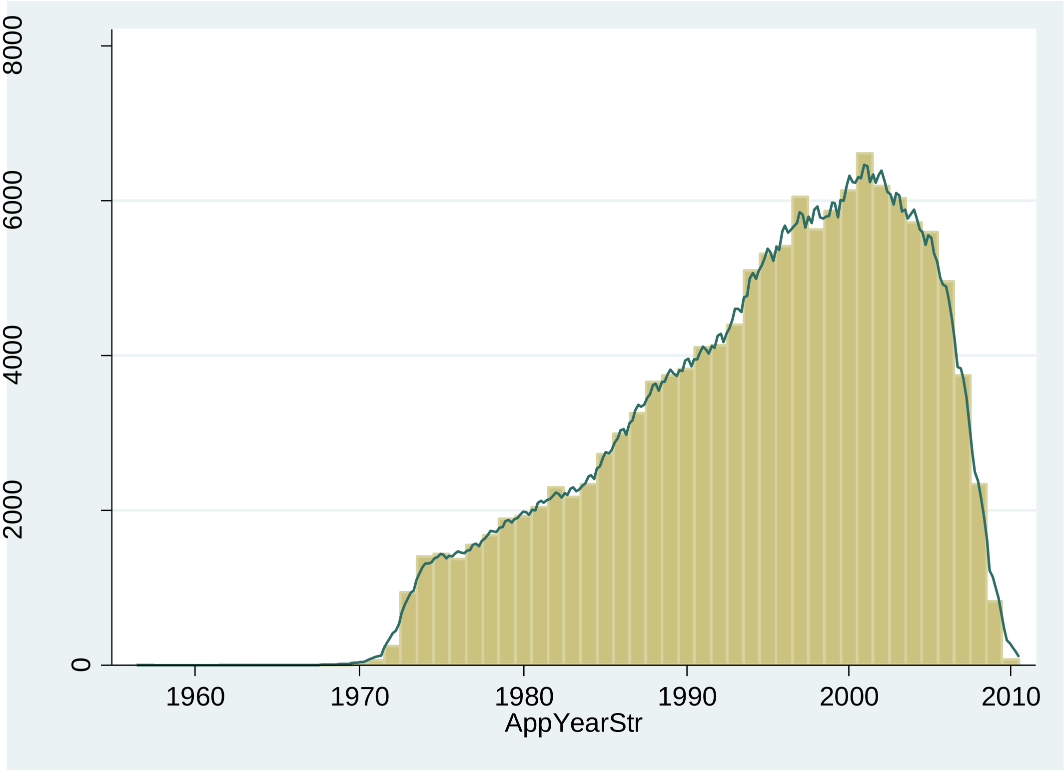
Figure 4 - Patents per inventor/assignee

Patents per inventor: 236,677 patents / 23,121 inventors = 10,2364517105662 patents per inventor

Patent per assignee: 236,677 patents / 10,105 assignees = 23,4217714002969 patents per assignee

Figure 5 - Inventors per year

AppYearStr	Freq.	Percent	Cum.
1957	1	0.00	0.00
1962	3	0.00	0.00
1963	2	0.00	0.00
1964	2	0.00	0.01
1965	3	0.00	0.01
1966	1	0.00	0.01
1967	6	0.00	0.01
1968	6	0.00	0.02
1969	12	0.01	0.03
1970	38	0.03	0.05
1971	69	0.05	0.11
1972	251	0.18	0.29
1973	947	0.70	0.99
1974	1,406	1.03	2.02
1975	1,443	1.06	3.08
1976	1,374	1.01	4.09
1977	1,553	1.14	5.23
1978	1,678	1.23	6.47
1979	1,896	1.39	7.86
1980	1,927	1.42	9.28
1981	2,048	1.51	10.78
1982	2,304	1.69	12.48
1983	2,172	1.60	14.08
1984	2,339	1.72	15.80
1985	2,732	2.01	17.80
1986	2,996	2.20	20.01
1987	3,256	2.39	22.40
1988	3,667	2.70	25.10
1989	3,745	2.75	27.85
1990	3,832	2.82	30.67
1991	4,114	3.03	33.70
1992	4,137	3.04	36.74
1993	4,397	3.23	39.97
1994	5,101	3.75	43.72
1995	5,313	3.91	47.63
1996	5,414	3.98	51.61
1997	6,049	4.45	56.06
1998	5,630	4.14	60.20
1999	5,872	4.32	64.52
2000	6,136	4.51	69.03
2001	6,615	4.86	73.89
2002	6,193	4.55	78.45
2003	6,039	4.44	82.89
2004	5,725	4.21	87.10
2005	5,595	4.11	91.21
2006	4,961	3.65	94.86
2007	3,748	2.76	97.62
2008	2,339	1.72	99.34
2009	825	0.61	99.94
2010	79	0.06	100.00
Total	135,991	100.00	



Graph including Kernel density plot

Figure 6 - Inventors per country

Country	Freq.	Percent	Cum.
AR	3	0.01	0.01
AT	116	0.47	0.48
AU	49	0.20	0.68
BE	40	0.16	0.84
BG	3	0.01	0.85
BR	2	0.01	0.86
BS	1	0.00	0.86
CA	201	0.81	1.67
CH	87	0.35	2.02
CN	152	0.61	2.64
CO	1	0.00	2.64
DE	652	2.63	5.27
DK	10	0.04	5.31
EG	2	0.01	5.32
ES	11	0.04	5.36
FI	9	0.04	5.40
FR	235	0.95	6.35
GB	459	1.85	8.20
GR	1	0.00	8.20
HK	77	0.31	8.51
HU	4	0.02	8.53
IE	39	0.16	8.68
IL	39	0.16	8.84
IN	9	0.04	8.88
IR	1	0.00	8.88
IT	105	0.42	9.30
JP	9,753	39.32	48.62
KR	771	3.11	51.73
MX	4	0.02	51.75
MY	10	0.04	51.79
NL	322	1.30	53.09
NO	52	0.21	53.30
NZ	9	0.04	53.33
OM	1	0.00	53.34
PH	4	0.02	53.35
PK	1	0.00	53.36
PL	1	0.00	53.36
PT	1	0.00	53.36
RU	31	0.12	53.49
SA	1	0.00	53.49
SC	1	0.00	53.50
SE	21	0.08	53.58
SG	370	1.49	55.07
SK	1	0.00	55.08
TH	10	0.04	55.12
TW	192	0.77	55.89
UA	2	0.01	55.90
US	10,933	44.08	99.98
VE	1	0.00	99.98
ZA	5	0.02	100.00
Total	24,805	100.00	

Figure 7 - Moves per year

AppYearStr	Freq.	Percent	Cum.
1969	1	0.03	0.03
1970	1	0.03	0.06
1971	1	0.03	0.08
1972	5	0.14	0.22
1973	32	0.89	1.11
1974	47	1.30	2.41
1975	40	1.11	3.52
1976	31	0.86	4.38
1977	51	1.41	5.79
1978	70	1.94	7.73
1979	49	1.36	9.09
1980	67	1.86	10.95
1981	68	1.88	12.83
1982	72	2.00	14.83
1983	92	2.55	17.38
1984	88	2.44	19.82
1985	109	3.02	22.84
1986	137	3.80	26.64
1987	143	3.96	30.60
1988	161	4.46	35.06
1989	165	4.57	39.63
1990	193	5.35	44.98
1991	197	5.46	50.44
1992	174	4.82	55.27
1993	239	6.62	61.89
1994	272	7.54	69.43
1995	335	9.28	78.71
1996	341	9.45	88.17
1997	427	11.83	100.00
Total	3,608	100.00	