THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

Identification and Mobility of Israeli Patenting Inventors

Manuel Trajtenberg[1], Gil Shiff[2]

Discussion Paper No. 5-2008

April, 2008

***Acknowledgements**:* This project benefited greatly from the work of a group of very talented and dedicated research assistants, primarily Michael Katz, Alon Eizenberg, and Ran Eilat, and the singular contribution of Ran Melamed. Useful comments were provided by participants in various seminars. We acknowledge the financial support at various stages and for different facets of the project by the Sapir Center, the Israel Science Foundation, and the Neaman Institute at the Technion**.**

The paper can be downloaded from http://econ.tau.ac.il/sapir

[1] Manuel Trajtenberg, The Eitan Berglas School of Economics, Tel Aviv University, Ramat Aviv.

[2] Gil Shiff, The National Economic Council, Prime Minister's Office, Jerusalem.

Abstract

Relying on the methodology developed in Trajtenberg, Shiff and Melamed (2006), we identify a rather comprehensive list of over 6,000 Israeli inventors that have patented in the US. These inventors represent the backbone of innovation in Israel, and the driving force of its flagship High Tech sector. We examine up close detailed information on these inventors, including their "fertility" and "importance" in terms of the number of US patents registered in their name as well as qualitative indicators of those patents, in comparison to the universe of patenting inventors (about 1.8 million). We then focus on their mobility, both across assignees (employers) and geographical locations, within and outside Israel. One of the interesting questions in this respect is the determinants and consequences of mobility: who tends to move, and what happens to the quality of innovations following a move? We find that Israeli inventors are much more mobile than others, and that there is an association between quality and mobility, but we cannot determine at this stage causal links. Using ancillary data on first names, we find that aside from the 6,000 inventors based in Israel, there are another 2,000 that engage in innovation outside Israel, primarily in the US. This is one further manifestation of the brain drain that has been identified also in academia.

JEL: O30, C81, C88
Key words: Patents, inventors, mobility.

I. Introduction

This paper is a first demonstration of the economic research opportunities lay within the creation detailed data on inventors, which was created by Trajtenberg, Shiff and Melamed (2006) (from now on: TSM). TSM outlined a methodology (and corresponding computer algorithms) for matching names and building a comprehensive inventors' data. Thus, the main target of this paper is to demonstrate some of the research opportunities opened with the creation of this data. We use this data to investigate the profile and characteristic of the Israeli inventors, comparing them to the 'average' inventor and to estimate the mobility phenomenon and the variables determining the patents 'quality' and inventors' mobility between assignees and geographical locations.

As mentioned, this paper will focus on the Israeli inventors' data. There are three main advantages of focusing on the Israeli inventors. First, the Israeli inventors' data were created both manually and by using computerized matching process (the *"CMP"*). The manually created file can be related as an almost fully "accurate" dataset and will be used to give fully reliable results and for verifying the automated results quality. Second, our thorough knowledge with the data enables us to closely verify and investigate some of the results. Third, as will be later elaborated, the Israeli inventor has some interesting characteristic, especially when investigating mobility, which may give focused insights on the subject.

The structure of the rest of the paper is as follows. Section 2 give the background for the paper and briefly reviews the empirical literature of the patents data and the inventors' data in particular. Section 3 summarizes the matching process, which was in used in this paper and was fully described in the TSM paper. Section 4 present descriptive statistics of the Israeli patents and section 5 presents some statistics of the Israeli inventors. Section 6 investigates on the brain drain phenomena of the Israeli inventors and section 7 displays ecnometrical models for investigating the mobility of inventors, its sources and its influence. Conclusions close this paper.

II. Background

Mobility

In this section, we'll discuss mobility as object of study in economics, and the mobility of inventors in particular. As mentioned, this paper will display a first-cut of econometric results studying the phenomena of inventors' mobility. Until now, the research of mobility was largely neglected, and as will be demonstrated bellow, with the creation of full scale database containing over million and half different inventors, there are exciting research opportunities revealed.

In general, every economic phenomenon takes place 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 growing importance of the space dimension can be view in several main aspects. First, the development of global trade of virtually anything, led to overwhelming volumes of movement of goods and services (i.e., the phenomenon of globalization, outsourcing and off-shoring). Second, the constant need of reallocation of resources due the changing demands and technologies leads to constant mobility of production factors (e.g., mobility across firms and regions, migration of employees, FDI). Last, the emergence of the "Knowledge Economy" led to growing reliance on dissemination of knowledge, information and ideas across space.

While focusing on the "space" dimension, we claim that mobility plays an important role in the process of growth, and vice-versa. The process of growth is caused by and causing a need constant reallocation of resources, which is caused by and causing a mobility of various factors. Thus, the mobility of factors is crucial for generating growth, and can be viewed as one of key factors of growth. Main examples of major mobility processes include the dramatic shifts from agro to industry and to services, and the recent shift within services to ICT and health care sectors.

Because of the importance of the mobility process in economic growth, we need to gain a thorough and systematic understanding of the factors facilitating and hindering mobility, and of the benefits associated with reallocation, increased specialization, as well as its costs (e.g., disruption). More specifically, when focusing of inventors (or in general researches and scientists) and when using empirical observations, we can observe frequent movement of inventors across firms, regions, countries and research fields. The main two questions arising from this phenomenon are:

(i) Why do inventors move? What economic rationale underlies their mobility?

(ii) What are the consequences of moving (for the individual inventor, for the firm and for the economy)?

When addressing the first question, .i.e. why inventors move, and assuming the movement is voluntary, we can expect that a rationale inventor will perform a move only if she expects a utility gain. Thus, an inventor is likely to move only if the expected value of the move minus the movement's costs exceeds the expected value of staying put. The main challenge is therefore to provide an actual empirical content to excepted value of move, its costs and the excepted value of staying. The tentative assumption is that if inventor had more fertile ideas, she will tend to move more, so as to find a better match, but we need to confirm this assumption using empirical results. Following this question, we should examine where and why inventors move to (e.g., from large to small (start up) firms? from "garage" to corporations? from Universities to industry?).

The framework for approaching the second question, i.e., the impacts of mobility, follows Weitzman (1998) of cross-pollination. The probability of "inventing", i.e. of creating a new bit of Knowledge (signed as *"K"*) - *ΔK* - depends upon two main factors. First, it depends on the quantity of *K* to which the researcher is exposed. This exposition requires physical proximity to carriers of *"K"*, signed as d_{ij}. Second, it depends on the variety of *K* to which the researcher is exposed (can be seen as exposition to different approaches), signed by α. Thus, the probability of creating new knowledge can be formulated as:

$$Prob(\Delta K)_i = f(\sum_j (1-d_{ij})K_j^{\alpha}),$$
$$\alpha < 1, \quad 0 \le d_{ij} \le 1$$

Both the quantity of knowledge and its variety have a positive influence on the creation of knowledge. Inventors or researchers that move are likely to be exposed to more, and more diverse, bits of K, hence the probability that they will invent increases. Thus, our tentative hypothesis is that mobility is causing R&D "productivity". Furthermore, mobility also entails a positive externality: not only the moving inventor gets increased exposure, but also her new colleagues get exposed to her, benefiting likewise. Therefore, we may conclude that there may be too little mobility, and when formulating an R&D policy we should consider the mobility exceeding benefits.

Patents Based Research

The idea of using patent data in a large scale for economic research goes back to the seminal work of Schmookler (1966), followed by Scherer (1982), and Griliches (1984).[1] One of the major limitations of these and related research programs, extremely valuable as they had been, was that they relied exclusively on simple patent counts as indicators of innovative output. However, it has long been recognized that innovations vary enormously in their technological and economic "importance", "significance" or "value", and moreover, that the distribution of such "values" is extremely skewed. The line of research initiated by Schankerman and Pakes (1986) using patent renewal data clearly revealed these features of the patent data (see also Pakes and Simpson, 1991). Thus, simple patent counts were seriously and inherently limited in the extent to which they could faithfully capture and summarize the underlying heterogeneity (see Griliches, Hall and Pakes, 1987). A further (related) drawback was of course that these projects did not make use of any of the other data items contained in the patents themselves, and could not do so, given the stringent limitations on data availability at the time.

Keenly aware of the need to overcome those limitations and of the intriguing possibilities opened by patent citations (as revealed for example in Trajtenberg, 1990),

[1] This section is not meant to be a full literature survey but rather give the background for this paper and highlight wide-scale research projects that used inventors' data. For a survey of research using patent data, see Griliches (1990).

Rebecca Henderson, Adam Jaffe and Manuel Trajtenberg undertook work aimed at demonstrating the potential usefulness of citations for a variety of purposes, primarily as indicators of spillovers (Jaffe, Trajtenberg and Henderson, 1993), and as ingredients in the construction of measures for key features of innovations such as “importance”, “originality” and “generality” (Trajtenberg, Jaffe and Henderson, 1997). They used for these projects relatively small samples of patent data that were acquired and constructed with a single, specific purpose in mind. However, as the data requirements grew it became clear that it was extremely inefficient, if not impossible, to carry out a large-scale research agenda on such a piece-wise basis.

Joined by Bronwyn Hall, Jaffe and Trajtenberg undertook to construct a comprehensive patent data file comprising detailed information on each patent as well as a series of indicators based on citations, that could not only account for (at least some of) the heterogeneity of patents, but also allow us to link patents over time and space. The result was the so-called “NBER Patent and Citations Data”, which has been opened for general use since 2001 (see http://www.nber.org/patents/). The data comprise detailed information on almost 3 million US patents granted between January 1963 and December 1999, all patent citations made between 1975 and 1999 (over 16 million), and a reasonably broad match of patents to Compustat (the data set of all firms traded in the US stock market). A book followed soon after (Jaffe and Trajtenberg, 2002), containing many of the authors’ previous articles on patents, as well as a CD with the complete data. The availability of these data has greatly stimulated research in this and related areas, and there are by now scores of papers and ongoing projects using it.

However, an important piece of information appearing in patents has not been used often in research so far, still less on a major scale, and that is the identity of the inventors themselves. If we could unequivocally identify each inventor (e.g. if each had an ID number), then we could follow the patenting history of each of them, trace their mobility, investigate their characteristic etc. TSM tried to tackle this problem by developing a comprehensive automated matching algorithm, with the purpose of determining whether two patent holders are the same person and creating a full scale inventors dataset. As will be later demonstrated in this paper, the research opportunities opened up by harnessing the inventors' data are undoubtedly far reaching and exciting.

Before the TSM methodology, which will summarized in the next section, there have been very few attempts to do so on a large scale (see Table II.1 below), with good reason: a major stumbling block is that we cannot identify from the data as is *"who is who"* among the inventors, due to two fundamental problems. First, the name of the same inventor may be spelled slightly differently across some of his/her patents, it may come with or without the middle name and/or the initial, with or without surname modifies, etc. Thus, a name such as Tra***j***tenberg may be spelled in one patent with a "***j***", in another with a "***ch***" (i.e. Tra***ch***tenberg), and likewise for "Manuel" and "Emanuel". Secondly, suppose that the inventor name in one patent is exactly the same as the inventor name in another patent – do the two correspond necessarily to the same person? We don't know, and cannot infer it just from the name: this is the "John Smith" problem, that is, different inventors having exactly the same name may appear in various patents, and we need to be able to tell them apart.

Absent a way of dealing systematically with these issues the data on inventors is essentially useless, since whatever the shortcut strategy that one may adopt (e.g. match any two patents with exactly the same name, ignore all spelling variations, etc.), it would be riddled with error, and moreover, it would be impossible to assess the true extent and nature of those errors. Tackling these problems properly (and in finite time) is extremely difficult, for two reasons: first, the sheer size of the data, which consist of over 4 million "records";[2] second, almost half of the inventors are located outside the USA, and foreign names, particularly East-Asian ones, present idiosyncratic problems of their own which require careful treatment. It is therefore clear that any attempt to address the *"who is who"* problem must rely on automated, computerized algorithms, and that there are significant economies of scale in doing so.

[2] Each record is a unique combination of a patent and an inventor. Recalling that the NBER data contains over 2 million patents, and that each has on average 2 inventors, the multiplication gives the number of records in the Inventors file.

Aided by a very talented and dedicated team of research assistants,[3] Trajtenberg undertook back in 2002 to develop a "computerized matching procedure" (CMP) that would tackle head on the *"who is who"* problem, and render a list of unique inventors. Joined later by Shiff and Melamed, and after 4 years of intensive efforts, the project has reached fruition: a paper published in 2006 presented a well-performing and reasonably accurate CMP, which produced a list of unique inventors, attached to it detailed data on the inventors' patenting histories, and probed the use of the data by conducting preliminary studies of inventors' mobility.

Over the past 4-5 years there have been a significant number of research projects attempting to take advantage of inventors' data, most of them using relatively small samples, and thus being able to do the matching with the aid of ad hoc, manual methods. There have also been a few attempts to use large scale inventors' data, having to develop for that purpose some sort of computerized procedure. Table II.1 summarizes this emerging literature.[4] These projects have greatly increased our understanding of the potentialities of the inventors' data, shedding light in so doing on interesting aspects of inventors' mobility and related issues. Thus, they should be regarded as important stepping stones towards the development of a more comprehensive and accurate matching methodology, as the one presented at TSM.

Table II-1

Papers Using Patent Inventors Data

#	**Authors**	**Data Source**	**Focus of research**	**Matching algorithm**	**# of inventors**

[3] They included Michael Katz, who did most of the Benchmark Israeli Inventors Set (see Section VI), Alon Eizenberg, who developed the "Mark I" CMP, and Ran Eilat, who developed parts of the final version of the CMP.

[4] Over the past years Trajtenberg presented in numerous seminars the main thrust of the methodology, as well as first-cut results on inventors' mobility. Although he did not communicate the initial phases of the project via (quotable) working papers, the power-point presentations used in these seminars were made widely available and contribute to disseminate the methodological approach.

	Large-scale patent data				
1.	Singh (2003)	NBER Patent file, USPTO, 1975-2002	Mobility of inventors, diffusion and social networks	Same 1st & last names, middle initial, patent sub-category (2 digit)	1.7 million
2.	Kim, Lee & Marschke (2005)	USPTO, NBER Patent File, etc. 1975-2002	Mobility from Universities to Industry	Similar to Trajtenberg (2004), w/t scoring	2.3 million (thru 2002)
3.	Jones (2005)	NBER Patent file, 1963-1999	Changing "burden of knowledge" of inventors; team work	Identical 1st & last names, and middle initials	1.4 million
4.	Fleming, Marx & Strumsky (2006)	Extended NBER Patent File, thru 2002	Employment changes of US inventors, non-compete agreements	Frequencies of names, + overlaps of co-inventors	2 million (thru 2002)
Smaller samples					
5.	Stolpe (2001)	1,398 US patents, 1975-95	Mobility of inventors and spillovers in LCD technology	Acknowledges problem of lack of algorithm.	2,116
6.	Rosenkopf & Almeida (2003)	Patents of 74 semiconductor firms, 1990-95	Firm alliances and the mobility of inventors	NA	NA
7.	Song, Almeida and Wu (2003)	Patents of semiconductor firms, 1975-99	Learning by hiring, move of inventors from US to non-US firms	Exact names matched, plus manual/heuristic checks	180
8.	Crespi, Geuna & Nesta (2005)	PatVal, EPO, 1993-1997	Mobility of academic inventors	Survey	9,000
9.	Hoisl (2006)	Survey German inventors. Pat Val, 1977-2002	Mobility and productivity of inventors	NA	3,049 / several hundred
10	Zucker & Darby (2006)	USPTO, 1976-2004	Careers of star scientists	Names, CVs	1,838
11	Agrawal, Cockburn, & McHale (2003)	USPTO, NBER Patent file, 1990-2002	Social capital effect on knowledge spillovers	Exact name for finding self-citations	59,734 observations on movers
12	Breschi &Lissoni (2003)	Italian inventors, EPO 1978-1999	Localized knowledge spillovers controlled by inventors network	Exact name, technological field	30,170
13	Alcacer & Gittelman (2004)	Sample from USPTO, 2001-2003	The role of inventors and examiners in the generation of patent citations	Exact name, assignee, location	40,797

III. Overview of the Computerized Matching Procedure

This section will summarize the computerized matching procedure, which was fully detailed in the TSM paper.

III.1 The data inputs

The raw data used in this research and in TSM comes from the NBER Patents and Citations Data File (see Hall, Jaffe and Trajtenberg, 2001), and in particular from the PAT63_99 file, the Inventor file and the CITE75_99 file. PAT63_99 contains the main data fields from the front page of utility patents issued by the USPTO between 1963 and the end of 1999, as well as additional variables constructed with the aid of citations. The panel Inventor file consists of all patent-inventor pairs: patents typically have more than one inventor (the mean is 2), and hence each patent generates a number of records equal to the number of inventors appearing in it. The data fields in the Inventors file include the patent number, the name of the inventor (first, last, middle and surname) and her address (street[5], city, state (US only), zip code (only in some US patents and country)

We merged the data of the Inventors file with the PAT63_99 file, thus creating a data set in which each record contains the information about the inventor plus some of the key variables of the patent itself, such as the Assignee and Patent Class. Since as said each patent has on average about 2 inventors, the 2,139,313 patents in PAT63_99 for 1975-1999 generated 4,298,457 records in the new Inventors file;[6] this file constitutes the starting point of the computerized matching work.

Based on this file, the computerized matching procedure (CMP) was built on two stages handling the two fundamental problems posed by the "*who is who*" question: first, the name of the same inventor may be spelled slightly differently across her patents, and second,

[5] This "Street" field is relevant only to unassigned patents, or to those assigned to individuals

[6] The "gross" total was of 4,301,229 records. However, 2,772 records with missing last names or "duplicate records" were eliminated, rendering a net of 4,298,457 records. By duplicate records we mean records that have the same patent number and exactly the same inventor name, and hence are almost certainly mistakes.

even if the inventor name in one patent is exactly the same as the name in another patent we don't know whether or not such name refers to the same person.

III.2 Stage 1: Grouping similar names using Soundex

The first stage of the CMP consists of identifying and grouping together all names/records that are deemed to refer potentially to the same inventor, e.g. Ben Grosman*n*, Ben Grossman and Ben*n* Grossman; such groupings was labeled as ***"p-sets"*** – ***p*** for "potential," that is, potentially the same inventor. Eventually it may turn out that these records refer to different inventors, but the point is that we would never know if the two records are not brought together to begin with and considered for a potential match. The key problem was that the name of a given inventor may be spelled in slightly different ways across the various patents in which the inventor appears. The various spellings may be due to two different problems:

The first problem is technical in nature, and refers to the appearance of all sorts of non-letter characters and symbols in the names. In order to tackle this problem we first standardized all the names by eliminating non-letter characters, symbols and spaces from the names, and rewritten the name in capital letters.

The second problem refers to differences in the actual spelling of names. In order to those spelling variations we needed a set of rules to "standardize" names, such that say the names Grosman*n* and Grossman would be identically coded, and thus (if having the same first name as well) be considered as part of the same ***p-set***. In order to handle this problem we used the **"Soundex"** system. The latter is a coding method adopted by the US Census in the 1930's, in order to tackle the problems posed by variations in the spelling of names. In our context the Soundex method offers a handy tool to group together all records that may potentially refer to the same inventor. This algorithm transforms names into alphanumeric codes by taking the initial letter as is and coding successive, non-identical consonants to numbers (each letter is scored 1 to 6 according to its group sound). We deployed the original code to be more accurate (using 6 digits code rather than 3) and by implementing the same procedure also for the inventor's first name.

The use of names standardization and Soundex then helps us guard against ***"Type I error"***, which occurs if we under-match records, i.e. if we miss records that should be compared to establish whether or not they match, but instead we regard them from the start as different inventors. There are some potential sources of Type I error that one can think of, and that Soundex could not overcome, but it is not possible to pinpoint them in the data and assess their incidence (e.g., mistake at the name's initial which is taken as given by the Soundex, usage of nicknames is some of the inventor's patents, deliberate name change due to martial status or name localization, etc.). In those cases patents of the same inventor might be assigned from the start to different *p-sets* since the Soundex code would be different, and therefore will not be matched. Based just on causal observation our impression is that those remaining Type I errors are very rare overall, and hence that Soundex does a good job at inclusion, i.e. at bringing together names that should be considered as potential matches.

We now turn to ***"Type II"*** errors, that is, those incurred when we end up matching records that belong in fact to different inventors. This will lead, of course, to "too few" inventors, and therefore to spurious mobility, spillovers, etc. This turned out to be the predominant concern throughout, and therefore most of the methodological apparatus that we develop below is meant to tackle it. In principle the second stage of the matching process (i.e. checking every pair of records within a given *p-set* to see if they refer to the same inventor) should take care of Type II errors, but it turns out that the Soundex method itself may induce Type II errors that would have not occurred otherwise: First, we found many cases that Soundex grouped together very different names (e.g., Brook, Bryg and Byres) thus expanding the *p-set* too much. Second, because the Soundex algorithm was originally designed to handle only English last names, its usage for first name and for non-English names (especially East-Asian names) caused as well in many cases over-expansion of the *p-sets*. Given that Stage 2 is not (and cannot be) full proof, the *p-sets* over-expansions might cause Type II errors. Therefore, in order to guard against Type II errors at this initial stage we used a 6-digit numeric code (after the initial) rather than 3 digits as envisioned in the original Soundex, and we've narrowed the *p-sets* definition and stringent the matching criteria for short Soundex-coded first names and for East-Asian inventors.

To recap, Stage 1 consists of transforming the raw file of 4.3 million records into 630,000 mutually exclusive *p-sets*, that is, groupings of records that have sufficiently similar names to be regarded as being potentially the same inventor. In so doing, we first clean-up and standardize the names (last and first names), and apply the 6-digit Soundex coding method to both the first and the last name of each record. Records with the same such alphanumeric code are grouped together into *p-sets*, for consideration in the second stage.

III.3 Stage 2: The matching process

Having grouped the standardized inventors' names in the first stage to *p-sets*, the question now is how to decide whether or not each pair of records within a given *p-set* ("suspects" displaying the same name or equivalent names according to Soundex) refers to the same inventor. There is no way of knowing *"who is who"* within each *p-set*, unless one undertakes to develop a comprehensive, computerized system for comparing look-alike records.

The ensuing procedure involves pair-wise comparisons between any two "suspects", of a series of variables (matching criteria) such as the middle name, the geographic location (e.g. zip codes, cities, etc.), the technological area (i.e. patent class), the assignee, the identity of the co-inventors, etc. If a data item is the same in two suspect records (e.g. if two records display the same address, or are in the same patent class, or share the same partners), then the pair is assigned a certain score. The scores are meant to reflect the strength of each criterion, that is, the extent to which the comparison according to that variable is thought to be informative. If the sum of these scores is above a predetermined threshold, the two records are "matched", that is, they are regarded as being the same inventor. Once that is done for all the pairs in the comparison set we impose transitivity, that is, if record ***A*** is matched to record ***B,*** and ***B*** to ***C,*** then the three are regarded as the same inventor.

The matching criteria

We now lay out the use of matching criteria, and discuss their relative informational strength. As will be presented bellow, for determining the informativeness of some of the criteria we used the "rareness" of the inventors' names, and the size of some of the categories involved (for cities, assignees, and patent classes) as auxiliary tools. Thus for example, if two suspects are located in the same city but the city is large they would receive a ***lower*** score on that account than if the two reside in a small town. The reason is simply that the probability that two records displaying the same inventor name refer to the same individual is deemed higher if the two are located in a small town rather than a large one, and similarly for employers (i.e. assignees) and patent classes. The other parameter affecting the scoring system is the frequency of the names themselves: both family names and first names vary a great deal in terms of their observed frequency in the relevant populations, some being very common, others relatively rare. Thus, if a name is "rare" in terms of the number of times it appears in the Inventors file (e.g. Griliches versus Smith) then the score would be higher. The obvious reason is that two records displaying an identical "rare" name and appearing say in the same city are significantly more likely to refer to the same inventor, than if the name were a common one. The two criteria thus render a scoring ***matrix*** that relies on the size of cities, assignees, and patent class (small or large), and on the relative frequency of the inventor's name (rare or frequent).[7]

The matching criteria used by the CMP are:

1. ***Full Address* -** This criterion is met whenever two records share the same country, city and street address.[8] We consider this to be a very "strong" criterion ("near-certain"), since it is extremely unlikely that two different inventors with the same Soundex-coded name reside in exactly the same address.

[7] Short of using the true frequencies of each name within its population and the actual size of each city and assignee, we computed the frequencies in our patent data, and used these as proxies. Based of these computations we fixed cutoff values determining whether each city, assignee and patent class are small or large, and whether each name is rare or frequent.

[8] For U.S. addresses the Zip code can be used as well.

2. ***Self Citation*** - Consider two patents, 1 and 2, sharing the same Soundex-coded inventor's name; the self-citation criterion is satisfied when patent 2, where Joe Doe name appears as one of the inventors, cites patent 1, where the same Soundex code appears. Since the probability of self-citation is known to be significantly higher *ceteris paribus* than the probability of citing someone else's patent, then the converse must also be true, that is, if we observe a self citation then the two Soundex-equivalent names are likely to refer to the same inventor (i.e., inventor citing another inventor with the same name significantly raises the probability that the two are in fact the same person).

3. ***Shared Partners*** - This criterion refers to the fact that collaborations among inventors are very likely to be persistent: if two patents share the same Soundex-coded name and the same co-inventor(s) Soundex-coded name, then the two quite probably refer to the same inventor.

4. ***Full middle name / middle name initial / surname modifier*** **-** The premise here is that the degree of informativeness of names (regarding the *"who is who"* problem) follows the following order: last (family) name and first name (which are used for determining the *p-sets*), middle name, middle name initial, surname modifier. The ***full middle name*** criterion is satisfied whenever two records share the same Soundex-coded middle name, and that middle name is not just to an initial. In many other records we observe just the ***middle name initial*** rather than the full middle name, and hence we may not be able to tell for example, whether John W. Fields and John William Fields refer to the same inventor. The full middle name criterion would not be satisfied for such two records, but the middle name's initial is off course informative in and of itself, and should increase the likelihood of a match. We make the score associated with this criterion depend also on the frequency of the last and first names involved. Lastly, the ***surname modifier*** criterion is satisfied whenever two records share the same non-missing surname modifier value (e.g., "Jr.").

5. ***Assignee*** - The "assignee" is the organization to which the patent is assigned at issue (or reassigned later on). The assignee may be the firm/corporation in which the inventor works (these are the majority of cases), a Government agency, a University or other such organizations. Missing values for assignee indicate that the patent was unassigned or assigned to an individual. Clearly, if two patents exhibiting the same Soundex-coded name exhibit also the same assignee, it is more likely that the two refer to the same inventor than if the assignees were different. The score for this criterion depends on the assignee "size" and the "rareness" of the inventor's name: a rare name in a small assignee carries more informational weight than a common name in a large assignee.

6. ***City*** - This criterion is satisfied whenever two records sharing the same Soundex-coded name share also the same (non-missing) city (for U.S. inventors the ZIP variable serves the same function).[9] As with assignee, we distinguish between large and small cities, and further differentiate the score by the frequency of names. It should be noted that during the preliminary process, city names had to be "cleaned up" and standardized using an automated process.

7. ***Patent class* -** This criterion pertains to the affinity between records in technology space, as indicated by the patent classification system: inventors are likely to work in the same or similar technological fields over time, and hence are likely to obtain patents classified in the same patent class. To put it differently, two records exhibiting the same Soundex-coded name are more likely to refer to the same inventor if the patent class in both is the same. As with the previous two criteria, belonging to smaller patent classes is deemed to be more informative than belonging to larger ones.

The matching threshold and scores

Clearly, any numerical scheme of scores and thresholds would be inherently arbitrary, since we would be assigning a ***cardinal*** measure to what is essentially only an ***ordinal*** relationship. Nevertheless, we decided that imputing (cardinal) scores was the most efficient method for determining a match. Following a lengthy and cumbersome process of extensive experimentation with alternative scoring schemes and corresponding thresholds, we settled for the one presented below, which seems to perform fairly well. However, we

[9] "Same city" means the same city name in the same country, and in the same state if in the US. Obviously, the city criterion is relevant only if the stronger full address criterion was not used (the full address includes the city name).

should keep in mind that this is by no means a full-proof scheme, and that there is as said an unavoidable measure of arbitrariness in the use of any such procedure.

There is no inherent meaning to the numerical values of the scores, but only in conjunction with the thresholds. For example, a score of 100 for a given criterion *vis a vis* a threshold of 120 just means that this criterion by itself is not enough to ensure a match, but is quite "close" to it, so that in conjunction with just another "weak" criterion it would suffice. Rather than having a unique threshold we specify three different threshold levels, differing according to the extent to which the last and first names are informative in and of themselves:[10] whether or not the names are ***exactly*** the same (as opposed to being the same Soundex-coded), and what is their length in terms of Soundex characters. Thus, the threshold level is lower the more similar the names are to begin with, and the more non-zero Soundex characters they comprise – clearly, longer Soundex-codes are more informative, a fact that is particularly relevant for East-Asian names. Table V.1 presents the criteria used to set the thresholds and their respective numerical values.

Table II.1 - Thresholds

Informativeness of names and determination of thresholds	**Threshold values**
• ***Exactly*** same first name (or Soundex-coded first name has at least 5 non-zero digits) *and* exactly same last name (or Soundex-coded last name has at least 5 non-zero digits)	100
• ***Exactly*** same last name (but not exactly same first name) or • Soundex-coded last name has at least 2 non-zero digits (but less than 5)	120
• All other cases	180

The scoring scheme

We categorize the various matching criteria into four "groups" according to their relative strength in conveying information for the matching decision, and assign to each group a numerical score, which should be interpreted in terms of the specified threshold levels. Thus for example, if two records having the same Soundex-coded name have the same full address then we are as sure as one can be that the two refer to the same inventor,

[10] The (equal) alternative would have been to treat these characteristics as matching criteria, add their scores to the criteria listed above, and compare the total to a unique threshold.

and hence the score on that account will be the highest (and in fact in most cases it will be sufficient for a match). On the other hand, sharing the same patent class is a rather weak indicator, and hence the score on that account will be low and size-dependent.

As mentioned before, the scoring of the criteria related to city, assignee and patent class depends both upon the frequency of names and upon size (computed as the number of patents of each category), as shown in Table V.2:

Table II.2 - Size and Frequency Dependent Scores

	Cutoff levels		**Score**	
	"Rare" name *(freq > 17)*	**"Common" name** *(freq ≤16)*	**Below cutoff**	**Above Cutoff**
City	2,500	1,382 *(median)*	100	80
Assignee	2,500	500	100	80
Patent Class	30,000	18,861 *(median)*	80	50

Table V.3 shows the complete scoring scheme:

Table II.3 - Scoring Scheme
(threshold levels: 100, 120, 180)

Group	**Criterion**	**Score**
1	Exact same address, Self citation, Shared partners (co-inventors)	120
2	Full middle name, Initial of middle name for "rare" names[11], "Small" assignee / rare names, "Small" city (or Zip) / rare names	100
3	"Small" patent class / rare names, "Large" assignee / frequent names, "Large" city / frequent names	80
4	"Large" patent class / frequent names Surname modifier Initial of middle name for frequent names	50

[11] To recall, the cutoff level for names is 16, i.e. if the frequency of a name in the data is less than 16 it regarded as "rare", and the converse for names that appear 16 or more times.

Thus, any of the criteria in Group 1 is sufficient to ensure a match if the last name of the two records compared is ***exactly*** the same, or if the Soundex-coded last name has at least 2 non-zero characters, since in such cases the threshold is 120 and so is the score that Group 1 criteria get. On the other hand, if the names are not very informative to begin with and hence the threshold is 180, then no single criterion is enough, and in fact for weaker criteria it would take at least two of Group 4 and one of Group 3 to ensure a match.

To recap, the matching procedure entails comparing every pair of records within a given *p-set* according to the various matching criteria, so that whenever a criterion holds the pair receives the corresponding score according to the table above. Finally, we compute the total score and compare it to the appropriate threshold, which in turn depends upon the characteristics of the name. If the total score exceeds the specified threshold we regard the two as the same inventor, and assign her a uniquely defined ID.

Transitivity

Stage 2 of the matching procedure entails making *n(n-1)/2* pair-wise comparisons within each *p-set*, where *n* is the number of Soundex-coded names in the *p-set*. Each such comparison renders a discrete decision of whether to match or not, but then we may be confronted with the following conundrum: supposed that there are 3 Soundex-coded names in the *p-set*, *A, B,* and *C*, and that the comparisons indicate that *A* and *B* match, *B* and *C* match, but *A* and *C* do not – whom should we regard as being the same inventor?

Logic dictates that we should impose transitivity, that is, if *A* and *B* refer to the same inventor, and so do *B* and *C,* then *A* should match *C* as well, and thus the three of them should be regarded as one and the same inventor. This is not a trivial decision and certainly not an innocent one, particularly if the *p-set* is large; however, it seems that transitivity is the only plausible course of action in such situations, which would render a logically consistent procedure.

IV. The Israeli Inventors File

The Israeli Inventors File is a comprehensive set of unique Israeli inventors (i.e. inventors appearing in US patents that listed their addresses in Israel at least once). As opposed to the Computerized Matching Procedure (CMP) described above, we initially constructed the Israeli inventors file manually: given that there were relatively few of them (about 6,000 inventors), and in view of our intimate familiarity with the country and its High Tech sector (which is the source of the vast majority of Israeli patented innovations), we could hope to be able to pin them down with a high degree of accuracy in finite time.

Constructing the Israeli Inventors File served three distinct purposes: First, it was a necessary first step of "learning-by-doing" towards the development of the CMP. Second, the resulting file was used as a benchmark to asses the performance of the CMP and to fine-tune it by "calibrating" the computerized results to the benchmark. Third, the file serves us here as original data to investigate the profile and characteristics of Israeli inventors and their mobility.

III.1 The construction of the Israeli Inventors File

We started by gathering all the patents in which at least one of the inventors had an address in Israel (there were 13,565 such records); we then took the names of those inventors, and extracted ***all*** the patents bearing also their names (obviously with addresses in other countries as well), which brought the total to 18,807 records. These can be regarded as the set of all patents associated with Israeli inventors (we refer to it as the "all inclusive set"). The goal was then to create out of this collection of records a list of ***unique*** Israeli inventors.[12]

[12] Note that not all the records end up as part of the final set: if for example we start with inventor ***A*** having a patent located in Israel, and we extract a patent with inventor ***A'*** (i.e. with a name similar or even identical to ***A***) but with an address in another country, then if the comparison of the two records rules out that the two refer to the same inventor, the record belonging to say ***A'*** just gets discarded from the set.

We proceeded by developing a first-cut computerized matching procedure following similar (but much coarser) principles as those outlined above, deployed it on the all-inclusive Israeli set, and examined carefully the ensuing list one by one (in alphabetical order). Suppose that 3 records were "matched" by this method: we observed then 3 rows of data, each with the data fields of each of the 3 patents presumed to belong to the same inventor, including the corresponding name in each case, address, assignee, etc. We then applied specific knowledge of names, spelling, assignees, locations, etc. as much as a healthy dose of discretion and common sense in order to decide whether or not the match was justified. In case of remaining doubts we looked for further clues in the patents themselves, and in a few hundred recalcitrant cases we sought additional external information, including phone calls to dozens of individuals and firms.

This tedious, time consuming procedure was made even harder by the fact that in some cases the initial alphabetical sorting of names did not necessarily bring together (that is, in close proximity) all the names that needed to be considered for a match: Yakoby and Jacoby for example would not appear next to each other on the spreadsheet, and hence if they referred to the same inventor we could easily miss them. Awareness of this problem brought us to develop heuristic rules to seek additional matches, particularly for some letters/initials (such as J and Y).

The end result was a list of **6,023** unique Israeli inventors and all their patents, totaling **15,310** records, which we can safely regard as being as comprehensive and accurate a set as possible. "Accuracy" here means that there should be very few Type II errors left, that is, as far as we know we have not matched together inventors that are in fact different individuals. As to Type I errors, we may have missed records when forming the all-inclusive set, and as said there may still be cases such as "Yacoby and Jacoby" which we did not identify. We shall refer to this final set of Israeli patents as the "Benchmark Israeli Inventors Set," or ***BIIS*** for short.

III.2 Using the BIIS to fine-tune the computerized matching procedure (CMP)

As already mentioned, contrasting the results of the CMP to the BIIS was one of the key methods used to try to improve the matching algorithm. The difficulty lay in the fact that there is no clear way of doing the comparison, let alone of quantifying it. In other words, any specific version of the CMP would render a list of unique Israeli inventors (and their corresponding patents), which obviously would not be identical to the BIIS – how could we then assess the "goodness of fit" between the two (if the latter is regarded as "data")? Spotty comparisons of differences between them are surely informative but can go only so far, and furthermore they cannot be too helpful if one considers multidimensional small changes in the matching parameters. We thus developed three alternative "goodness of fit indices", ***GOFIs***, and used them to fine tune the CMP *vis a vis* the BIIS: we adopted changes in the matching parameters that resulted in an improvement in these indices, worsening would lead to rejection of the changes, and mixed results would prompt us for further checks and close up examinations of the differences.

As a first stage, we "match" each unique inventor arrived at by the CMP (refer to it as ***"C"***) to its counterpart in the BIIS (call it ***"B"***). Accordingly, let $\boldsymbol{C_{ij}}$ be the set of all patents of inventor $\boldsymbol{j}$ named on patent (record) $\boldsymbol{i}$, as identified by the CMP, and $\boldsymbol{B_{ij}}$ the corresponding set found in BIIS. The indices are then defined as follows:

$$(1) \qquad GOFI_1 \equiv Mean\left[\frac{|B_{ij} \cap C_{ij}|}{|B_{ij} \cup C_{ij}|}\right], \quad i = 1,...,N_{IL}$$

where $|B_{ij} \cap C_{ij}|$ is the number of patents assigned to inventor $\boldsymbol{j}$ named in patent $\boldsymbol{i}$ both by the CMP and by BIIS, $|B_{ij} \cup C_{ij}|$ is the number of patents assigned to that inventor by the union of the two, and $\boldsymbol{N_{IL}}$ is the total number of patents/records associated with Israeli inventors. The idea is simply that we compute for each record of each inventor the share of the intersection of both sets out of the union of the sets: the max value is 1, which will be achieved only when both sets are exactly the same, and decreases as the two are less similar.

$$(2)a \quad GOFI_2 \equiv Mean\left[\frac{|B_{ij} \cap C_{ij}|}{|B_{ij}|}\right], \quad (2)b \; GOFI_2 \equiv Mean\left[\frac{|B_{ij} \cap C_{ij}|}{|C_{ij}|}\right]$$

The basic intuition is similar to that of *GOFI*$_1$, except that this index uses the number of patents assigned to the inventor by ***either*** method as the denominator, and not their union. In this case the comparison between (2)*a* and (2)*b* can be quite informative, in terms of which procedure is over or under matching relative to the other, and by how much. Thus for example if the CMP is under-matching then (2)*b* will be close to 1 and larger than (2)*a*.

These indices allow us to diagnose the extent to which the CMP comes close to replicating the BIIS, which we regard as the "true" matching. In practice we proceeded as follows: first, we constructed the BIIS in parallel to developing the first-cut CMP; second, we tested, improved and refined the CMP in a variety of ways; lastly, we compared the (already much improved) CMP to the BIIS using the GOFI indices, and further fine-tuned the CMP.

Table III.1
Comparing the CMP to the BIIS

	CMP	**BIIS**
Number of patents	9,155	
Number of records	15,310[13]	
Number of original names	6,316	
Number of Soundex-coded names (i.e. number of *p-sets)*	5,861	
Final number of unique inventors	6,900	6,025
Average number of patents per inventor	2.22	2.54
GOFI$_1$	0.88	
GOFI$_2$	0.99	0.99

[13] Six "duplicate" records (i.e. records having the same name and same patent number) were deleted in the cleaning procedure.

Table III.1 shows the last round of the latter stage: as we can see, the two methods are quite "close" according to $GOFI_1$, but the difference in the values of $GOFI_2$ reveals that the CMP it still significantly under-matching. Further examining the differences we learnt that the good news is that the incidence of Type II error induced by the CMP is indeed very low: there were only 73 inventors that the CMP over-matched (i.e. they corresponded to 196 inventors as identified by the BIIS). Furthermore, in most cases these were in fact not errors at all, but rather the CMP was right and thus the BIIS was wrong. Given that the emphasis in developing the CMP was in avoiding Type II error, it seems that goal was accomplished. The bad news is the high incidence of Type I error: the CMP under-matched in about 15% of cases, that is, it erroneously split 780 inventors into 1,781. The main reasons for those errors were:

1. ***Little in common*** *(or move without a trace)*: These are cases whereby two records turn out to refer to the same inventor, even though there is little or nothing in common between them other than the name. Formally, that means that the criteria used for matching failed to detect any similarity or linkage between the records. In these cases the matching of records by the BIIS was obviously done according to additional information not found in the patents themselves, and hence this is pretty much the upper bound of the matching ability of the CMP (or any such automated method).

2. ***Spelling mistakes in the names***: Soundex-coded names cannot overcome all possible spelling mistakes, and hence we may not match with the CMP two records that belong to the same inventor simply because they were not in the same *p-set* to begin with. This is a Type I error that could in principle be reduced if the coding improves.

3. ***Errors in the spelling of cities, street addresses and assignees:*** the quality of the match depends to a significant extent on the quality of the data fields used by the matching criteria. If of two records in a given *p-set* one names "Jaffa" as the city of the inventor and the other "Yaffa", we probably will not match them even though we should.

Whereas the frequency of cases corresponding to cause 1 should be seen as an irreducible rate of Type I error, that is not so for causes 2 and 3: further cleaning of the data, and further fine-tuning of the Soundex method may significantly reduce these sources of Type I error as well. Close examination of the distribution of actual causes of Type I error revealed that about ½ of them correspond to cause 1, 1/3 of cases to cause 2 and the remainder of about 1/6 to cause 3. Thus, even if we were able to avoid Soundex-based and other spelling mistakes altogether, the CMP is still expected to result in ***7-8%*** of Type I errors, which thus constitutes a ***lower bound for Type I errors***.

To handle some extent of third problem and due to the manageable size of the file and our close familiarity with the Israeli data, we manually cleaned, fixed and merged the Israeli cities names and merged together duplicate assignees (two assignees IDs which are in fact a single assignee). This process was crucial for getting the most reliable results for the Israeli automated process, and was especially important due to the numerous spelling variations for Hebrew names in English. This process had a dramatic effect particularly on the number of different cities names in the dataset - reducing their number from 1,549 to 741.[14] The result for the assignees number was smaller but nevertheless significant - reducing their number from 1,783 to 1,626. After this process was completed we re-executed the CMP process on the fixed dataset (note that we changed only the data, not the process). Using this new data we reduced the number of different inventors to 6,670, as apposed to the original 6,900. As the more accurate results, this fixed data will be used throughout this paper as the CMP file. Such cleaning process is very difficult to be done on the entire file, but once again its affect is predicted to be much less significant for countries using Latin alphabet.

Table IV-1
Comparing the CMP to the BIIS

	Fixed CMP	**BIIS**
Number of patents	9,155	

[14] Fox example, cities such as Zichron Yaacov has initially 13 (!) different spelling variations.

Number of records	15,310	
Number of original names	6,316	
Number of Soundex-coded names (i.e. number of *p-sets)*	5,861	
Final number of unique inventors	6,670	6,025
Average number of patents per inventor	2.30	2.54

This cleaning process, which reduced the number of inventors by 3.5%, had significant affect for handling the under-matching errors. After this process the CMP erroneously split 629 to 1,400 inventors, while before it split 780 inventors into 1,781. Based on our previous estimations and by examining differences between the files, we conclude that even tough the process did not include standardization of all textual fields (e.g., the street names) and did not handle all the possible mistakes, the process eliminated almost all of the under matching caused by the variables spelling mistakes. Note that the process did not handle any spelling mistakes in the names and the "little in common" problem.

The effect on over matching mistakes of this process was negligible - there are 74 inventors that the CMP over-matched, and corresponded to 199 inventors (in the original dataset 73 inventors corresponded to 196).

Table IV-2 compares the GOFIs using the original and fixed datasets. The two indexes show improvement in the similarity to the benchmark, and can be seen as further evidence for the improvement of the database.

Table IV-2
GOFIs for Comparing Original and Fixed CMP to the BIIS

	Original CMP		**Fixed CMP**	
	CMP	**BIIS**	**CMP**	**BIIS**
$\boldsymbol{GOFI_1}$	0.88		0.90	
$\boldsymbol{GOFI_2}$	0.99	0.89	0.99	0.92

V. The Israeli Inventor's Patents

As a first step towards analyzing the Israeli Inventors and their profile, we first examine their patents characteristics. Note that because of the tight work relationship between Israel and the US, especially in the IP intensive industries, it's reasonable to assume the Israeli patents applied in the US, reflects the vast majority of the overall Israeli patents.

First, in order to put things in context, according to the USPTO website (which include data till 2006, compared to only 1999 in our dataset), there are 14,469 Israeli originated patents, which accounts for 0.42% of the patents between 1977 and 2006. This average number reflects an impressive growth from only 0.14% in 1977 to 0.67% in 2006. This is not a unique phenomenon to Israel, but it magnitude is outstanding. The overall USPTO data shows a similar, but much more moderate, trend of an increasing foreign patents weight. The foreign patents, which accounted for 35.5% of the patents in 1977, accounted for 47.9% in 2006.[15] In our data set there are only 9,155 Israeli patents, which are defined as patents of inventors, which applied for a patent from Israel at least once in their patenting career. Those patents reflect 15,310 'records' and 6,205 inventors using the BIIS or 6,670 using CMP.

For analyzing the field of research of the Israeli inventors, we examine the categories of the Israeli patents, as presented in table IV-3.

Table V-1: Distribution of Patents across Categories

	Israeli Patents			Entire USPTO File		
	All Patents	**Applied Before 1995**	**Applied After 1995**	**All Patents**	**Applied Before 1995**	**Applied After 1995**
1- Chemical	16.71%	19.01%	10.51%	20.06%	21.02%	16.15%
2- Computers & Communications	18.14%	14.80%	27.16%	11.55%	10.09%	17.54%
3- Drugs & Medical	18.80%	17.24%	23.00%	8.57%	7.86%	11.52%
4- Electrical & Electronic	15.88%	15.90%	15.84%	17.15%	16.94%	18.00%
5- Mechanical	14.11%	15.58%	10.15%	21.86%	22.69%	18.42%
6- Other	16.35%	17.47%	13.34%	20.80%	21.40%	18.37%
Total Num. of Patents	**9,155**	**6,681** (72.98%)	**2,474** (27.02%)	**2,139,314**	**1,719,983** (80.40%)	**419,331** (19.60%)

[15] http://www.uspto.gov/go/stats/cst_allh.htm

As expected, the Israeli patents main focus is Computers & Communications and on Drugs & Medical. Those research fields are the foundations for the Israeli successful ICT and Pharmaceutical industries, which have a relatively high importance and contribution to the overall Israeli economy. Furthermore, we see growth of those two fields after 1995 especially for the Computers & Communications category (the data includes 1995). This phenomenon, which exists also in a lower scale for the entire dataset, reflects the Hi-Tech boom during those years.

This data can be viewed as another evidence for the Israeli 'dual economy'.[16] On the one hand the Israeli ICT industry, which had great contribution to Israel's growth during the 1990s', is considered to be innovative with a developed start-ups companies industry. On the other hand, the other research categories have significantly less weight in the Israeli inventions, with extreme gaps in the Mechanical and 'Others' categories. That is while the hi-tech industries invests high amount of efforts and resources in R&D, the more traditional industries show low R&D investments rate, compared to other similar industries in other countries.

As another step towards analyzing the Israeli patent characteristic, we examine which assignees apply the patents. Table IV-4 presents the Israeli patents top 10 assignees.[17]

Table V-2: Israeli Patents Top Assignees

Assignee Name	**Assignee's Israeli Patents Num.**[18]	**Assignee's Total Patents Num.**	**% Israeli Patents**	**Academic Assignee**
Yeda R&D - Weizmann Institute of Science	363	369	98.37%	√
Motorola, INC.	165	12,528	1.32%	X
Intel Corporation	164	3,314	4.95%	X
Yissum Research Development - Hebrew University	163	166	98.19%	√
International Business Machines (IBM)	140	22,500	0.62%	X
Elscint Ltd.	138	159	86.79%	X

[16] For further details on the Israeli 'dual economy' see Lach, Shiff & Trajtenberg (2008).

[17] Note that this list was constructed after the cleaning process, which was earlier discussed.

[18] Israeli patents are defined here as patents applied from Israel and not all the Israeli inventors' patents.

Ramot University Authority for Applied Research – Tel-Aviv University	103	104	99.04%	√
Technion R&D Foundation	103	104	99.04%	√
Ormat Industries, Ltd.	90	93	96.77%	X
State of Israel, Ministry of Defense, Rafael	89	89	100.00%	X
Other assignees (1,451 other assignees)	4,074			
Individuals (no assignee assigned)	1,992			
Total	***7,584***			

This table is another indicator for the Israeli patents orientation and the R&D structure of the Israeli economy. Four out of the ten leading Israeli assignees, including the leading one, are technology transfer offices for academic institutes. Other three assignees are large multinational companies (Motorola, Intel and IBM), which have significant R&D centers in Israel, but most of their activity is outside of Israel as can be seen in the percentage of patents applied from Israel. The other companies represent other Israeli significant industries such as medical devices (Elscint), clean-tech geothermal energy (Ormat) and security industries (Rafael). This list represents most of the variety of the Israeli R&D focus.

The Israeli patents were defined as patents of Inventors, which applied for a patent at least once from Israel. Thus, not all of those patents were indeed applied from Israel: out of the 15,310 records in the dataset only 13,481 records are assigned to Israel. The other 1,829 records are of Israeli inventors, but were applied in other countries, mainly from the US (1,499 records). Table V-3 further examines the patents geographical distribution within Israel, displaying the top Israeli cities.[19]

Table V-3: Israeli Patents Top Cities

City Name	**Records Num.**	**Share**
Haifa	1,835	13.61%
Rehovot	1,551	11.51%
Jerusalem	1,512	11.22%
Tel Aviv - Yafo	1,398	10.37%

[19] Similarly to the assignees list, this list was constructed after the cleaning process, which was earlier discussed.

Ramat Gan	466	3.46%
Rishon Lezion	392	2.91%
Herzliya	380	2.82%
Petah Tikva	356	2.64%
Beer Sheva	332	2.46%
Ra'anana	290	2.15%
Other Cities (414 cities)	4,969	36.86%
Total (Israeli Located Records)	***13,481***	***100.00%***

From the table it's clear that the order of the cities is not completely correlated with their population size. For example, Jerusalem's population is 7 times larger than Rehovot's population (Jerusalem is the most populated city in Israel), but has less registered less patents than Rehovot. Except of their population size and their socio-economic profile it seems that two other main factors are determining the number of patents for each city. First, the location of a significant academic center has a great contribution: the four major patent producing cities are by far the four cities, which reflect the main academic research centers of Israel: Haifa (Technion), Rehovot (Weizmann Institute of Science), Jerusalem (Hebrew University) and Tel-Aviv (Tel-Aviv University). Second, the location of hi-tech R&D centers, and specifically of multinational companies, seems to have an important contribution. For example, Herzliya hosts R&D centers of Motorola, Sun Microsystems, Ra'anana hosts SAP, HP etc., so even tough they are not the ranked in the top 10 most populated cities and do not host significant academic research center, they have generated a significant amount of patents.

VI. Analysis of Israeli Inventors

Equipped with the set of "true" Israeli inventors and their patents contained in the BIIS file, we undertake here to examine the profile of these inventors and their mobility. In so doing we shall compare them when appropriate to the whole population of unique patent inventors (to be referred as PUI), which to recall comprise 1.6 million inventors. As we shall see below, Israeli inventors have some interesting features, and specifically tend to move more often, thus are excellent base for investigating mobility. Judging from the frequencies of their first names (checking the 200 top first names, which stand for 4,317 inventors) only 3% are almost certainly female and about additional 4% carry gender-neutral names.[20] Thus, the upper estimated limit of the female inventors is 7%. None of the examined names are distinctive for the Israeli-Arabs.[21]

The "fecundity" of Israeli inventors

As already mentioned, the 6,025 Israeli inventors in the BIIS file are named in 15,310 patents, yielding an average of about 2.5 patents per inventor. This figure is a bit bellow the average in the PUI file (2.6). Using the CMP file, the average number of patents per inventor is 2.3, due to the under matching problem which was previously discussed.

Figure 1 presents the distribution of patents per Israeli inventor, which is as expected highly skewed.

[20] This figure should be taken with a grain of salt, since there are many gender-neutral names in modern Hebrew.

[21] There are Israeli-inventors, which carry common Arabic names, such as Mohamad, but those names do not belong to the top 200 names.

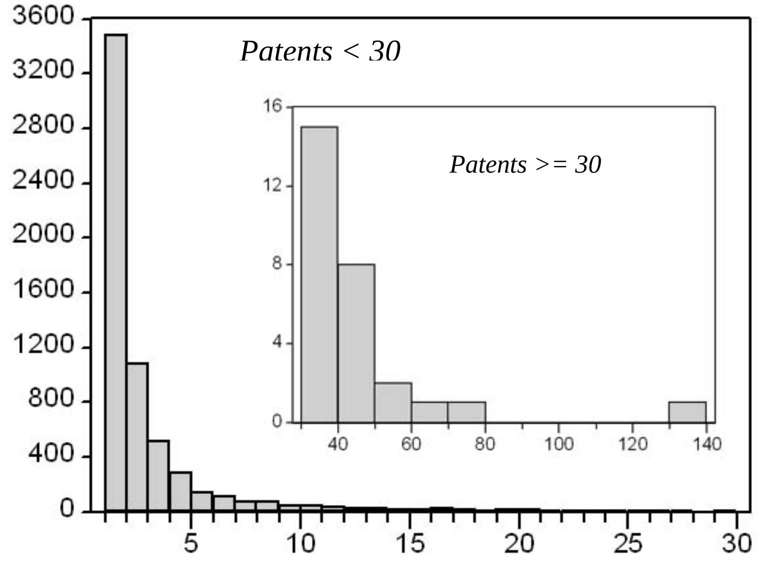

Figure 1: Distribution of number of patents per inventor

In Table V-1 we contrast the distribution of patents per inventor for Israelis using BIIS and CMP vs. the PUI.

Table VI-1: Shares of patents per inventor

	Israeli Inventors -BIIS		**Israeli Inventors -CMP**		**All Inventors**	
Patents	**Inventors**	**Share**	**Inventors**	**Share**	**Inventors**	**Share**
1	3,479	57.74%	4,150	62.22%	983,859	60.27%
2+	1,883	31.25%	1,899	28.47%	497,780	30.49%
5+	443	7.19%	424	6.36%	80,835	4.95%
10+	225	3.73%	192	2.88%	67,537	4.14%
50+	5	0.08%	5	0.07%	2,521	0.15%
total	***6,025***	***100.00***	***6,670***	***100.00***	***1,632,532***	***100.00%***
Average	***2.54***		***2.30***		***2.63***	

From the table we learn that the average Israeli inventor has 5%-15% (using BIIS or CMP) less patents than the overall average inventor. From further examination of this difference we conclude that most of the differences are due to the right tail of the distribution: while only 57.7% of Israeli inventors hold just one patent as opposed to 60.3% for the PUI, 0.08% of the Israeli has over 50 patents versus 0.15% for the PUI.[22] That is,

[22] Our record holder for the number of patents is Benzion Landa (the founder of Indigo, acquired by HP in 2002) with 133 patents. The runner up Drori Mordechai, with no assignees, has "only" 73 patents.

there are relatively few Israeli inventors with very high number of patents (e.g., over 50). This finding can be partially explained by the TSM finding that many of those inventors in the PUI file are east-Asian inventors, which are in fact a result of over-matching mistakes (especially due to Soundex problems with Eastern-Asian names).

The "qaulity" of Israeli inventors

As described in previous works (see for example Jaffe and Trajtenberg, 2002), patent citations constitute a good source of information that can be used to assess various aspects of the "quality" of patents. By extension, one can use indicators based on patent citations to ascertain the "quality" of inventors, meaning of course the mean "quality" of their patents. The first indicator we examine is the number of citations that each patent receives over time, which has been shown to be a good proxy for its importance or impact.

With this mind we analyze the mean number of citations an Israeli inventor has received. Table V-2 presents the distribution of mean citations per Israeli inventor using the BIIS and CMP vs. the PUI file.

Table VI-2: Mean citations received

Mean Citations	Israeli Inventors –BIIS		Israeli Inventors –CMP		All Inventors	
	Inventors	Share	Inventors	Share	Inventors	Share
0	1,660	27.55%	1885	28.26%	346,748	21.24%
0-1	941	15.62%	1022	15.32%	239,571	14.67%
1-3	1,258	20.88%	1345	20.16%	351,664	21.54%
3-5	811	13.46%	882	13.22%	235,176	14.41%
5-10	832	13.81%	929	13.93%	280,644	17.19%
10-20	396	6.57%	451	6.76%	134,856	8.26%
20+	127	2.11%	156	2.34%	43,873	2.69%
total	***6,025***	***100.00%***	***6,670***	***100.00%***	***1,632,532***	***100.00%***
Average	***3.79***		***3.92***		***4.51***	

This table shows that an Israeli inventor gets on average 3.79 citations per patent (or 3.92 in the CMP file), while the overall average is 4.51 citations. This means that the Israeli inventors' patents are less cited and ex-ante might be considered with lower quality.

It's a know fact that the number of citations received is strongly correlated with the patent grant year – the older the patent it has more chances to be cited by other newer patents. So in order to explain some extent of this finding we use our previous analysis of the patents distribution over time. As mentioned, the Israeli patents are on average newer than the overall database, thus have a structured bias in this variable. The correct this bias we first estimate the influence of the application year on the citations received.

Table VI-3: Citations received as Function of Application year	
C	609.13 (1.0453)
Application year	-0.304 (0.0005)
R squared	0.072

The regression results presented here, based on OLS regression of the entire 4.3 records, imply that every additional year yield on average 0.3 citations (e.g., patent applied in 1991 will have -0.3 less citation compared to a patent from 1990). The average Israeli patent was applied in 1990 while in the complete file was applied in 1988. Therefore, this difference explains 0.63 citations per patent and virtually the entire difference between the Israeli inventors and the other inventors (the exact difference is 0.72 citations using BIIS or 0.59 citations using CMP).

The "generality" index is another possible indicator of a patent's quality. This Herfindahl-based index, which values between zero and one where the higher indicates the patents as more general, indicates the variety of fields which cite the patent. (see Trajtenberg, Jaffe & Henderson 1997). High generality score suggests that the patent presumably had a widespread impact, in that it influenced subsequent innovations in a variety of fields.

The average generality measure for Israeli inventor is 0.305 (0.309 using CMP), while the overall inventors average is 0.316. This might indicate that the Israeli inventors generate patents that are more specific for their research fields and not as applicable for other fields compared to the average world-wide inventor. Some extent of the difference can be explained by the fact that this variable is positively correlated with the number of citations received: highly cited patents will tend to have higher generality scores (Hall, Jaffe and Trajtenberg, 2001). Therefore, this indicator might be biased for the less cited Israeli inventors. To estimate the bias, we estimate regression of the generality index as function of the citation received (on the entire file).

Table VI-4: Generality as Function of Citations Received
(t-statistics)

C	0.252 (1,329.6)
Citation Received	0.11 (657.3)
R squared	0.12

We conclude that every citation received increase the generality index by 0.11 on average. Thus, because the Israeli patents has on average 0.72 less citation, the bias of the generality index is 0.08, explaining most of the gap (0.59 less citations and 0.065 explanation, using the CMP)

As a last indicator for the patents quality, we examine the patent "originality" index. The originality index is similar to the "generality" index, but using the citations made and not the citations received. Thus, if a patent cites previous patents that belong to a narrow set of technologies the originality score will be low, whereas citing patents in a wide range of fields would render a high score. The average "originality" index for Israeli inventors is 0.371 (0.370 using CMP), compared to 0.353 of the all inventors dataset. This indicates that the Israeli inventors are more multi discipline and combine more research fields in their inventions. Similar to the "generality" index, this index is positively correlated with the number of cited patents - an Israeli patent cites on average 8.62 patents, while the average patent cites only 7.87 patents. This gap may explain the difference in the "originality" index. Once, again to estimate the bias we estimate function of the originality index by citations made.

Table VI-5: Originalty as Function of Citations Made
(t-statistics)

C	0.293 (1,329.6)

Citation Received	0.008 (657.3)
R squared	0.08

We conclude that the number of citation made bias the originality index by 0.006, which explains a third of the average difference in the originality index between Israeli inventors and other invetors.

Mobility

One of the unique advantages of our data on inventors is that it allows us to follow the career of each inventor along time and across space. In particular, it allows studying patterns of mobility, both geographically and across assignees (that is, across employers who own the right to the patents - corporations, universities, or government agencies).

Geographic Mobility

We will study the geographic mobility of Israeli inventors in two levels. First, we'll examine mobility within the country, and more specifically between districts. Second, we'll examine mobility between Israel and other countries, which will allow us to study the brain-drain phenomena of Israeli inventors.

For examining low-level geographical moves of the Israeli inventors, we first examine moves between cities. We find that 834 movers performed 1,675 moves between cities (the corresponding numbers are 505 inventors and 1,104 moves using CMP). In order to examine the geographical movements' flows in a national macro-view, we examine mobility between districts and not between specific cities. In order to perform this, we first merged the Israeli cities names from our file with the Israeli Central Bureau of Statistics cities classification by districts data (note that Israel is not separated into states, but has different official districts). Even though we cleaned the cities names and standardize them as possible, there are still some cities, which could not be matched with the official cities list, and will be marked as 'Unknown' district.

Table VI-6: Israeli Districts

District Name	**Records Num.**	**Share**
Center Region	4,332	32.09%
Tel-Aviv Area	3,385	25.08%
Haifa Area	2,392	17.72%
Jerusalem Area	1,606	11.90%
North Region	764	5.66%
South Region	690	5.11%
Gaza Strip, Judea & Samaria	119	0.88%
Unknown District	210	1.56%
Total (Israeli Located Records)[23]	***13,498***	

Looking at this table we learn that 57% of the patents were originated in the Israel's center region and Tel-Aviv metropolis, 18% and 12% from the other two main cities of Haifa and Jerusalem. Only 11% of the patents were originated from the periphery districts of Israel that is the outer northern and southern districts. We not turn to investigate the inventors flow between those districts.

Table VI-7: Israeli Inventors Districts Moves

To\ From	To Abroad	Unknown	Jerusalem Area	North	Haifa Area	Center	Tel-Aviv Area	South	Judea & Samaria	Total From
From Abroad	**237**	7	55	7	70	94	64	19	7	**560**
Unknown	1	**6**	5	14	8	15	10	3	4	**66**
Jerusalem Area	52	3	**30**	3	8	8	5	3	3	**115**
North	9	11	3	**22**	5	2	2	1	0	**55**
Haifa Area	65	8	14	7	**37**	13	11	3	0	**158**
Center	68	12	6	4	19	**166**	62	7	1	**345**
Tel-Aviv Area	89	11	4	3	12	88	**96**	1	1	**305**
South	16	2	4	0	5	10	4	**14**	0	**55**
Judea & Samaria	5	4	3	1	0	2	0	0	**1**	**16**
Total To	**542**	**64**	**124**	**61**	**164**	**398**	**254**	**51**	**17**	**1,675**
Net (To-From)	***-18***	***-2***	***9***	***6***	***6***	***53***	***-51***	***-4***	***1***	

[23] This list contains 13,498 records even tough only 13,481 records are assigned to Israel. From examining the 17 other records we found out that those records are indeed Israel record, but were assigned with the wrong country code – instead of IL Israeli code, they are mostly assigned to similar codes such as IT, IR, IS and NL.

From the table we learn that about quarter to third of the movement action is inside the districts, thus local. Most of the movements between districts are traced to the Tel-Aviv area and country's central region, which contribute most of the patents. The interesting figures here are the net flow in and out of those districts. This table indicates a significant in flow to the central district and an out flow from the Tel-Aviv area. Even though there's no clear borderline between those two districts, we can see those finding as evidence for a known phenomenon is Israel: many R&D centers were opened in the last decade outside the Israeli classic "Silicon Wadi" of Tel-Aviv and Herzliya area. Thus, the inventions activity today in Israel is less centralized and was spread to suburbs such as Raanana, Netanya, etc.

From an Israeli national view, the prime interest in terms of geographical mobility resides of course in their mobility in and out of the country, given the strong outward orientation of its highly successful High Tech sector. Thus, we have hereby an almost unique opportunity to look into a particular form of "brain mobility" with individual-level data.

Of the 6,025 Israeli inventors 410 inventors (6.8%) moved at some point in or out of the country (this number drops to 3% using CMP). This may seem like a small number, especially when using the CMP dataset, but it is much higher than the corresponding figure for the PUI, which stands at less than 1% (0.5% to be exact). This confirms the impression that Israeli inventors are much more mobile than their counterparts abroad.

Table VI-8: Israli Inventors Country Moves

# of Moves	# of Inventors	Share of Inventors	Share of Movers
0	5,615	93.20%	-
1	282	4.68%	68.78%
2	81	1.34%	19.76%
3	25	0.41%	6.10%
4-5	13	0.22%	3.17%
6+	9	0.15%	2.20%
Total	***6,205***	100.00%	100.00%

This table shows the distribution of the overall 663 country movements' occurrences in the Israeli file. The vast majority of the inventors did not move, but we should bear in mind that 3,479 inventors out of the 5,615 non-movers have only one patent, thus by definition cannot show any movements. Therefore, the more accurate focal point is the 2,546 inventors which have more than one patent. Most of the movers (69%) moved only once, and 20% of the movers moved twice. The most frantic inventors are two inventors with 13 moves and one with 14. Examining their patents it seems like two of them invented in parallel in two locations of multinational companies (with locations both Israeli the US), while the third researched in two academic institutions (Tel-Aviv University and Cornell). The next table displays the flow between Israel and the other countries.

Table VI-9: Israeli Inventors Country Moves

Destination\ Source Country	From Israel	To Israel	Net Movement to Israel
Canada	12	9	-3
Great Britain	7	5	-2
Italy	5	4	-1
US	268	290	22
USSR	0	5	5
Other	23	20	-3
Total	***315***	***333***	***18***

The figures show that mobility is clearly a two-way street: 333 inventors moved at some point to Israel, whereas 315 moved at some point abroad. Thus, on net there seems to be neither significant gains nor losses in this respect (if anything there is a tiny gain of 18 inventors on net). Almost all of the movements of the Israeli inventors are from\to the US. This phenomenon reflects the fact that the Israeli High Tech sector, which is the source of many of the Israeli patents, is overwhelmingly export-oriented, particularly to the US. Note that in addition to those movements there are 15 moves of Israeli inventors, which does not involve Israel (e.g., an Israeli Inventors that moved between the US and Canada).

These figures are not conclusive for analyzing the brain-drain phenomenon, since there are Israeli inventors that never patented in Israel, and therefore our data do not include them. We explored this issue by searching for common and distinct Israeli first names among the entire inventors dataset. Table VI-8 shows the top ten Israeli common and distinctive names, out of a total of 30 such names. A common and distinct name was defined as a name, which is carried by at least 25 different Israeli inventors and that more than 50% of the inventors caring this name are Israeli.[24] Using our close familiarity with Israeli names, we examine the table and conclude that those names are indeed known Israeli names and from our knowledge are very rare in other countries.

Table VI-10: Israeli Top 10 Common and Distinct First Names

First Name	# of Inventors	# of Israeli Inventors	% that ever applied in Israel	% of inventors that never applied in Israel
MOSHE	267	169	63.30%	36.70%
ZVI	124	73	58.87%	41.13%
SHLOMO	128	72	56.25%	43.75%
SHMUEL	100	70	70.00%	30.00%
HAIM	96	60	62.50%	37.50%
URI	97	56	57.73%	42.27%
MEIR	72	55	76.39%	23.61%
YEHUDA	88	53	60.23%	39.77%
ILAN	100	53	53.00%	47.00%
MORDECHAI	75	48	64.00%	36.00%
Displayed Names Average	***1,147***	***709***	***61.81%***	***38.19%***
All Names Average	***2,245***	***1,383***	***61.60%***	***38.40%***

The table shows the percentage of those inventors that patented abroad but not in Israel, that is, those are presumed to be Israelis that left the country and pursued their career as inventors entirely abroad. The remarkable feature of the table is that for most of the names the percentage is tightly distributed around 38%. We conclude that there is a sizeable number of Israeli inventors abroad (mainly in the US) that pursued their entire careers there. If we take the figure of 38% as representative, then a rough estimate of their total number puts them at about 2,000 (i.e. 1/3 of the 6,000 Israeli inventors identified).

Ben-David (2008) studied the brain-drain phenomena of the Israeli academy researches, and may be used as a benchmark to check our results. He reports that the number

[24] Many of the most common Israeli names are not distinct to Israel (David, Michael, Joseph, Dan etc.).

of the Israeli Scholars in U.S. Universities is 25% out of the academic scholars in Israel, which is close to our finding. Furthermore, Ben-David reports that 32.8% of the total number of senior faculty researches in the computer-science field in Israel can be traced in the top American departments. As mentioned, Israeli inventions are biased towards the computers field, thus this number is a strong reinforcement for our findings.

Mobility across firms (assignees)

As a last step of studying the Israeli inventors' mobility, we take a look at mobility across assignees. We find 1,264 different moving inventors, which performed 2,787 instances of moves between assignees (2,276 moves for 1,054 inventors when using CMP). Table VI-9 presents the distribution of moves between Israeli inventors.

Table VI-11: Israeli Inventors Assignee Moves Distribution

# of Moves	# of Inventors	Share of Inventors	Share of Movers
0	4,761	79.02%	-
1	696	11.55%	55.06%
2	280	4.65%	22.15%
3	111	1.84%	8.78%
4-5	101	1.68%	7.99%
6-10	54	0.90%	4.27%
+10	22	0.37%	1.74%
Total	***6,205***	***100.00%***	***100.00%***

Note that 4,761 inventors have never moved across assignees. However, subtracting those with only one patent (3,479 inventors), which by definition cannot show any movements, we get that almost 50% of the inventors with more than one patent moved between assignees (1,264 inventors out of 2,546). This number is extremely high and requires further examination. The most frantic inventors are four inventors with 24 assignee moves and one inventor with 33 assignee moves (between various optical lens and cameras companies). Table VI-10 specifies the flow of inventors across different types of assignees.

Table III.VI-12: Israeli Inventors Assignee Moves Distribution

		To				
		Corporate	**Individual**	**Government**	**Academy**	**Total**
From	**Corporate**	997	390	31	176	1,594
	Individual	446	-	41	112	599
	Government	54	41	28	23	146
	Academy	230	104	31	83	448
	Total	1,727	535	131	394	***2,787***
	Net (To-From)	133	-64	-15	-54	

The table shows that about third of the movement activity is between different corporate assignees. From the net flows between the different types of assignees, we learn about a significant flow of inventors from academic and governmental institutes to corporate. Those figures reveal a brain-drain phenomenon out of public institutes to private businesses. In addition, there is a substantial two-way flow between inventors which apply as individual (with no assignee) and inventors which apply under corporate assignee, with a net flow towards working for corporate. This means that many inventors start their career as individuals ('garage' inventors) and later join or start their own corporate.

VII. An Econometric Analysis of Israeli Inventors Mobility

In this section we perform a more thorough analysis of the Israeli inventors' mobility. More specifically, we will focus on two main questions:

1. How does mobility affect the quality of a patent?
2. Which inventors tend to move more?

In order to perform these estimations we have created a set of new variables. Since we are looking from the inventor's viewpoint, almost all of them involve data concerning the inventor, and not just a specific patent. These are the new variables:

1. **Patent Sequence** – The sequential number of the patent in the inventor's record.
2. **Partners** – The amount of inventors that applied for the patent, excluding the inventor whose records we are examining. This variable is actually the number of registered inventors of the patent minus one.
3. **Moved Assignee** ('*Move Assig*') – A Boolean variable, which equals one if the inventor had moved assignee while applying for the current patent, i.e., the previous patent was for a specific assignee and the current to another one.
4. **Moved Geography** ('*Move Geo*') – Similar to the previous "Move Assignee" variable, only applying to geographical move. Geographical move is defined as change of at least the registered city.
5. **First Year** – Indicated the Inventor's application year of her first patent.

These new variables mostly relate to the inventor's current patent (excluding the 'First Year' variable). The next group of new variables relates to the inventor's previous patents (in case she has any). When using these variables in regressions we must exclude all inventors' first invention, thus excluding all inventors with only one patent.

6. **Partners (-1)** – The number of partners in the previous patent.
7. **Mean Past Citations** – The mean number of citations the inventor received up to his previous patent.
8. **Sum of Past Assignees Moves** ('*Sum Assig Moves (-1)*') – Accumulation of the inventor's number of assignee moves up to his previous patent.

9. **Sum of Past Geographical Moves** (*'Sum Geo Moves(-1)'*) – Similar to "Sum of Past Assignee Moves", only applied for geographical moves.

Using OLS (with White Heteroskedasticity Consistent Standard Errors), we estimated the influence of several parameters on the amount of citations a patent received. The regression includes dummy variables for patent categories 1 to 5 (category 6 serves as the benchmark).

Indicators of patent "quality" as function of mobility

As a first step for studying the importance of mobility, we will estimate the affect of mobility on indicator of the patent "quality". First, we examine regressions with *creceive* as the dependant variable, i.e. Citations to this patent as a function of control variables, previous history of inventor, and whether she moved in the current patent, compared to the previous one. The full regression equation is:

$$\begin{aligned}&Citations = \beta_1 Appyear + \beta_2 Pat\_Seq + \beta_3 Partners + \beta_4 Mean\_Citations(-1) + \\ &\beta_5 Move\_Assig + \beta_6 Move\_Geo + \beta_7 Sum\_Assig\_Moves(-1) + \beta_8 Sum\_Geo\_Moves(-1) + \\ &\beta_9 FirstYear + D_1 Cat1 + D_2 Cat2 + D_3 Cat3 + D_4 Cat4 + D_5 Cat5 + C\end{aligned}$$

The results of this regression are presented in Table VII-1:

Table VII-1: Dependent Variable: Citations

OLS (White SE), t-statistic scores in parenthesis

Variable	BIIS	CMP
Observations	9,285	8,640
Application Year	-0.44 (25.74)	-0.44 (24.60)
Pat Sequence	0.02 (1.74)	0.02 (2.10)
Partners	-0.01 (-0.49)	-0.02 (-0.75)
Mean Past Citations	0.26 (10.97)	0.28 (11.67)
Assignee Move	0.33 (1.88)	0.38 (2.20)
Geographic Move	0.50 (2.44)	0.62 (2.41)
Sum Assig. Moves (-1)	-0.02 (-0.69)	-0.02 (-0.85)
Sum Geo. Moves (-1)	-0.003 (0.11)	-0.02 (-0.52)
First Year	0.04	0.06

	(3.76)	(4.67)
C	785.28 (24.15)	753.48 (22.82)
R^2	0.22	0.24

We can see that the coefficients in both regressions are quite similar, which is another indicator for the similarity of the data, thus for the 'quality' of the CMP data. It appears that the only variables which are not robust are Partners, and the sum of previous moves (both of assignees and geographical moves). From these results, we may conclude that the number of partners does not affect the amount of citations received of the invention.

Regarding the sum of moves, these variables are highly correlated to the dummy variables Move Assig and Move Geo, which are robust. Therefore there appears to be a multicolinearity problem. Moreover, there is also multicolinearity problem between assignee moves and geographical moves (when moving from one city to the other the inventor has high probability of changing also assignee). Though it may not be clear which effect of mobility is dominant, it is evident that the inventor's mobility has a significant effect on the amount of citations the current patent receives, hence, on the patent's 'quality'.

Next, we will estimate further regressions for finding the impact of mobility on other indicators of patent “importance” as the dependant variables. The "importance" variables were: Generality, Originality and Claims, while the other explaining variables are similar.

We already discussed and explained the concept of the 'generality' and 'orginality' indicators in a previous chapter: 'Generality' is defined as 1 – Herfindahl on patent classes of citations received, and 'Originality' as 1 – Herfindahl on patent classes of citations made.

The number of claims appears on the front page of patent application and specifies the components of the invention. Therefore it may be indicative to the patent's scope (see Hall, Jaffe & Trajtenberg 2001).

Table VII-2: Impact of inventors' mobility on other patent qualtative indicators. OLS (t- statistic in parenthesis, WHITE SE)

Variable	**Generality**		**Originality**		**Claims**	
Database	**BIIS**	**CMP**	**BIIS**	**CMP**	**BIIS**	**CMP**
Observations	6,087	5,662	8,912	8,294	8,051	7,491
Application Year	-0.011	-0.013	0.006	0.006	0.213	0.202

	(12.48)	(12.69)	(7.92)	(7.03)	(5.95)	(5.05)
Patent Sequence	-0.002	-0.001	-0.001	-0.001	0.053	0.052
	(3.13)	(1.61)	(2.81)	(2.36)	(2.23)	(2.07)
Partners	0.003	0.002	0.013	0.014	0.399	0.433
	(1.86)	(1.11)	(10.54)	(10.44)	(3.21)	(3.38)
Mean Past Citations	0.005	0.006	0.002	0.003	0.086	0.086
	(10.12)	(11.47)	(5.64)	(7.43)	(4.55)	(4.30)
Move Assignees	0.018	0.020	0.007	0.004	0.400	0.434
	(2.27)	(2.39)	(1.03)	(0.54)	(1.24)	(1.29)
Move Geographical	0.016	0.012	-0.006	-0.003	1.412	2.244
	(1.63)	(1.06)	(-0.72)	(-0.34)	(3.10)	(3.89)
Sum Assig. Moves (-1)	0.0004	-0.0005	0.004	0.004	0.204	0.242
	(0.29)	(-0.30)	(2.89)	(2.64)	(3.05)	(3.31)
Sum Geo. Moves (-1)	0.008	0.006	0.004	0.004	0.010	-0.002
	(3.67)	(2.46)	(2.52)	(2.69)	(0.11)	(-0.02)
First Year	0.002	0.003	0.0004	0.001	0.066	0.069
	(2.03)	(3.36)	(0.65)	(1.37)	(1.93)	(1.87)
C	19.76	19.36	-11.64	-12.62	-544.52	-526.80
	(14.16)	(13.70)	(-10.47)	(-11.16)	(-9.55)	(-9.02)
R^2	0.10	0.11	0.05	0.06	0.04	0.04

First, we conclude that the results imply that these three qualitative indicators act in a similar way in both databases (CMP and BIIS). In both cases the same variables are robust (or not), and the coefficients' values are very close to each other.

From this table we conclude some main findings. First, earlier patents of inventors (implied by the patent sequence) tend to be more “original" and "general". Second, we find highly significant lagged mean dependent variables (i.e., mean of past citations), which can be views as sort of “fixed effect". Thirds, we find highly positive impact of number of partners on the "generality" of the patents, meaning more partners give a more multidiscipline view. Regarding the impact of movements, we find that moves across assignees and/or location have a positive impact on the “value” of patent taken at the new place (except for not robust negative effect of location move on the 'Originality'). We find differences between the impact of assignee move and location move - assignee movement has stronger impact on the generality, while location movement has stronger impact on the number of claims.

Correlates of Mobility

We now turn to examine the decision to “move or not”, of each inventor at each point in time (i.e., with each additional patent). We examine the probability of an inventor to move as a function of her past history and performance, i.e. the “quality” of her previous patents, and of all relevant control variables. These estimations were performed using Binary Logit method, where the dependant variables were Move Assig and Move Geo. Formally:

$$Binary\,Logit(Move\_Assig/Geo)=f\left(\begin{array}{l}Appyear, Fyear, Pat\,Seq,\\ Partners(-1), Sum\_Assig\_Moves(-1),\\ Sum\_Geo\_Moves(-1), Citations(-1), Generlity(-1),\\ Originality(-1), Claims(-1), Dummies(Cat1-6), C\end{array}\right)$$

The outputs are presented the following table:

Table VII-3: Dependant Variables: Move Assig/Geo

Binary Logit (z-statistic in parenthesis)

Dependant Variable	**Move Assig**		**Move Geo**	
Database	**BIIS**	**CMP**	**BIIS**	**CMP**
Observations	6,717	6,216	9,762	9,761
Application Year	0.04 (9.57)	0.04 (9.76)	0.06 (13.89)	0.05 (10.18)
First Year	-0.03 (8.48)	-0.03 (8.25)	-0.06 (16.06)	-0.05 (11.81)
Patent Sequence	-0.04 (12.85)	-0.04 (12.15)	-0.03 (10.01)	-0.02 (8.01)
Partners (-1)	-0.03 (3.18)	-0.02 (2.58)	-0.02 (2.27)	-0.002 (0.17)
Sum Assig. Moves (-1)	0.11 (13.28)	0.12 (13.47)	-0.05 (6.05)	-0.04 (4.50)
Sum Geo. Moves (-1)	0.01 (0.99)	0.01 (0.85)	0.20 (17.27)	0.22 (17.85)
Citations (-1)	0.003 (1.88)	0.003 (1.61)	0.005 (2.96)	0.006 (2.79)
Generality (-1)	0.16 (2.46)	0.17 (2.52)	0.19 (1.95)	0.15 (2.05)
Originality (-1)	-0.003 (0.06)	0.02 (0.25)	-0.11 (1.83)	-0.09 (1.25)
Claims (-1)	0.0003 (0.27)	0.001 (1.33)	0.002 (1.67)	0.004 (3.09)
C	-16.61 (2.54)	-22.59 (3.25)	1.31 (0.21)	2.16 (0.31)
Probability (LR stat)	0.00	0.00	0.00	0.00

Once again, the results are similar in both databases (CMP and BIIS). We not turn to study how does the patenting “history” affects the probability of moving. In most cases it seems that the results affirm assumptions that one may consider to be quite intuitive. First, as the spreading of the 'globalization' effect, it appears that as the patent is applied at a later date, the probability to move (both assignee and geographically) increases. However, as the inventor is less a "veteran" (a later first year of invention) and the inventors is early on in her patenting career (the patent sequence) the probability to move decreases. Inventors with more partners in the previous patent, tends to diminish the probability to move. It is quite interesting to see that the cumulative amount of assignee moves (up to the previous patent) increases the probability to move to another assignee, but decreases to probability to move geographically (some kind of fixed effect). The cumulative amount of geographical moves increases the probability to move geographically (again), however, it is not robust regarding assignee move. Last, the inventors tend to move if, prior to the move, they had patents that are more "general" and were more highly cited and had more claims (most of those results are robust of both movements). The inventors are less likely to move was more 'original' (not robust).

To summarize this section, we conclude that inventors that have already produced "better" patents tend to move more often. Conversely, moving seems to impact favorably the "quality" of subsequent patents. We can interpret those results in two ways. First, inventors have better information on the expected impact of their patents than their employers, hence more likely to move if having patents with greater generality and citations (which are hard to observe *ex ante*). Second, employers successfully preempt moving of inventors with patents that are "better" in observable ways (claims and originality).

VIII. Summary

This paper is a first demonstration of the research opputonitues opened with the creation of the Inventors file.

Future researches may try to develop the mobility model in various ways: Deal with endogeneity, apply Arrelano-Bond model, bringing in data on firms and markets. Study the impact of inventors' mobility on firms' innovative performance. Use together both data on mobility of inventors and on citations to trace spillovers. Study mobility of inventors between regions and firms, as function of regional and firm-related variables.

References

Alcacer, Juan and Gittelman, Michelle, “How do I know what you know? The role of inventors and examiners in the generation of patent citations.” Cross Disciplinary Strategy Seminar, Stern School of Business, NYU, Spring 2004.

Algawar, Ajay K., Cockburn, Iain M., McHale, John, “Gone But Not Forgotten: Labor Flows, Knowledge Spillovers and Enduring Social Capital.” National Bureau of Economic Research Working Paper No. 9950, September 2003.

Ben-David

Breschi, Stefano and Lissoni, Francesco, "Mobility and Social Network: Localized Knowledge Spillovers Revisited", CESPRI Working Paper No. 142, March 2003.

Crespi, Gustavo A., Geuna, Aldo and Nesta, Lionel J. J., "Labor Mobility of Academic Inventors. Career Decision and Knowledge Transfer", EUI Working Paper RSCAS No. 2006/06, June 2006.

Fleming, Lee and Marx, Matt, “Non-competes and inventor mobility: the Michigan Experiment.” Harvard Business School Working Paper,2006.

Griliches, Zvi, (ed.) R&D, Patents, and Productivity, NBER Conference Proceedings. University of Chicago Press, 1984.

Griliches, Z., Hall, B.H. and A. Pakes, "The Value of Patents as Indicators of Inventive Activity," in P. Dasgupta and P. Stoneman, eds., Economic Policy and Technological Performance. Cambridge: Cambridge University Press, pp. 97-124, 1987.

Griliches, Zvi, “Patent Statistics as Economic Indicators,” Journal of Economic Literature 92: 630-653, 1990.

Hoisl, Karin, "Tracing Mobile Inventors – The Causality between Inventor Mobility and Inventor Productivity", Munich Business Research Working Paper Series No. 2006-9, May 2006.

Jaffe, A., Trajtenberg, M. and R. Henderson, “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” Quarterly Journal of Economics, pp. 577-598, August 1993.

Jaffe A. and Trajtenberg, M, Patents, Citations and Innovations: A Window to Knowledge Economy. Cambridge Mass: MIT Press, 2002.

Jones, Benjamin F., “The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?” National Bureau of Economic Research Working Paper No. 11360, May 2005.

Kim Jinyoung, Lee Sangjoon John, Marschke Gerald, “The Influence of University Research on Industrial Innovation.” National Bureau of Economic Research Working Paper No. 11447, June 2005.

Lach, Shiff, Trajtenberg

Pakes, Ariel and Simpson, Margaret, “The Analysis of Patent Renewal Data.” Brookings Papers on Economic Activity, Microeconomic Annual, pp. 331- 401, 1991.

Rosenkopf, Lori and Almeida, Paul, “Overcoming Local Search Through Alliances and Mobility.” Management Science, Vol. 49, No. 6, pp. 751-766, June 2003.

Schankerman, M. and A. Pakes, “Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period,” Economic Journal, Vol. 96, No. 384, pp. 1052-1077, December 1986.

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

Schmookler, J. Invention and Economic Growth. Cambridge: Harvard University Press, 1966.

Singh, Jasjit, "Inventor Mobility and Social Networks as Drivers of Knowledge Diffusion", Harvard University Working Paper Series, October 2003.

Song, Jaeyong, Almeida, Paul and Wu, Geraldine, “Learning-by-Hiring: When is Mobility More Likely to Facilitate Inferfirm Knowledge Transfer?.” Management Science, Vol. 49, No. 4, pp. 351-365, April 2003.

Stolpe, Michael, “Mobility of Research Workers and Knowledge Diffusion as Evidence in Patent Data – The Case of Liquid Crystal Display Technology.” Kiel Working Paper No. 1038, April 2001.

Trajtenberg, M. "A Penny for Your Quotes: Patent Citations and the Value of Innovations," The Rand Journal of Economics, 21(1), 172-187, Spring 1990.

Trajtenberg, M. “The Names Game: Using Inventors Patent Data in Economic Research”. http://www.tau.ac.il/~manuel/, Power-point presentation, 2004.

Trajtenberg, M., Jaffe, A. and R. Henderson, "University versus Corporate Patents: A Window on the Basicness of Invention," Economics of Innovation and New Technology, 5 (1), pp. 19-50, 1997.

Trajtenberg, Shiff and Melamed "The "Names Game": Harnessing Inventors' Patent Data for Economic Research", National Bureau of Economic Research Working Paper No. 12479, August 2006.

Weitzman M "Recombinant Growth" QJE 1998

Zucker, Lynne G., and Darby, Michael R., "Movement of Star Scientists and Engineers and High-Tech Firm Entry", National Bureau of Economic Research Working Paper No. 12172, April 2006.