The mobility of inventors and the productivity of research

Manuel Trajtenberg

Tel Aviv University, NBER and CEPR

July 2006

Plan of Talk

Themes:

- ***Mobility as object of study in economics; in particular mobility of inventors***
- ***Harnessing patent data on inventors for economic research***
- ***Study of inventors' mobility – first-cut econometric results***

Themes as promising research agenda

Mobility as theme of research in economics

Largely neglected, exciting research opportunities ahead.

Every econ phenomenon takes place in a certain ***"location" in time and space***. Lots of attention to the time dimension (e.g. discounting), less so to ***space***.

- ***Trade*** => Movement of goods and services; globalization, outsourcing/offshoring;
- ***Reallocation of resources*** => Mobility of factors (across firms and regions; migration, FDI);

and,

- ***Emergence of the Knowledge Economy*** => reliance on ***dissemination*** of Knowledge, of info, of ideas

Reallocation, Mobility and Growth

Growth ⇔ constant reallocation of resources

⇔ mobility of factors

e.g. dramatic shifts from agro to industry to services; within: to ICT, to Health Care.

Need systematic understanding of,

- Factors facilitating and hindering mobility;
- Benefits associated with reallocation, increased specialization, as well as costs e.g. disruption.

Mobility of Inventors

(scientists as well)

Empirical observation: we observe frequent movement of inventors across firms, regions, countries (see below).

(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, for the economy?

(*i*) Determinants of Inventors' Mobility

Econ 101 rationale (if voluntary move),

(Expected value of move – costs) > (expected value of staying put)

The point is to provide actual empirical content to *E(move)*, costs, *E(stay)*.

Tentative ***H***: if inventor had more fertile ideas, she will tend to move more, so as to find a better match.

Where move to? From large to small (start up) firms? From "garage" to corporations? From Universities to industry?

(ii) Impact of mobility

Framework follows "Recombinant Growth" by M. Weitzman, ***QJE*** 1998 – cross-pollination…

The probability of "inventing", i.e. of creating a new bit of ***Knowledge***, ***ΔK***, depends upon,

- the ***quantity*** of ***K*** to which the researcher is ***exposed***; requires ***physical proximity*** to carriers of ***K:*** d_{ij} (the "Agora factor")
- the ***variety*** of ***K*** to which the researcher is ***exposed*** (think of it as different approaches) => $\alpha < 1$

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

$$\alpha < 1, \quad 0 \le d_{ij} \le 1$$

Impact of mobility – cont.

Inventors/researchers that move are likely to be exposed to more, and more diverse, bits of K, hence the ***prob*** that they will invent increases; i.e. tentative hypothesis:

Mobility → R&D "productivity"

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.

→ there may be too little mobility

Second theme/research agenda:

Harnessing Patent Data on Inventors for Economic Research

Patents contain also information about identity of inventors (millions of them); unable to use it so far because of the ***"who is who?"*** problem, now made possible, vast research opportunities opened up *(not just for mobility)*

Front page of patent *(partial)*

United States Patent 6,539,988

Pressurized container adapter for charging automotive systems

Inventors:

Cowan; David M. (Brooklyn, NY); **Schapers; Jochen** (New York, NY); ***Trachtenberg*; *Saul*** (New York, NY); **Nikolayev; Nikolay V.** (Flushing, NY)

Assignee: **Interdynamics, Inc.** (Brooklyn, NY)

Filed: **December 28, 2001**

Current U.S. Class:141/67; 137/614.04; 141/351; 251/149.1

Intern'l Class: B65B

Key Issue: *Who is who?*

How do we know that two records with "same" or "similar" names refer to the same inventor?:

1. Is ***Manuel Trajtenberg*** the same inventor as ***Manuel Trajtenberg*** ?
2. Is Manuel Tra***j***tenberg the same inventor as Manuel Tra***ch***tenberg? Same as ***Manuel D. Trajtenberg***?

Magnitude of problem:

- Sheer ***size***: over 4 million "records"
- Have to rely ***only*** on information given in patents.
- About ½ of all patents are ***foreign*** (non-US), problems with e.g. Asian names.

Two-Stage Methodology for Matching Names

Stage 1:

Put together records "suspected" of being the same inventor: those with identical names, as well as sufficiently similar names, e.g. Manuel Tra***j***tenberg and Manuel Tra***ch***tenberg; use the ***SOUNDEX*** coding method.

Stage 2:

Match records/names within the above set deemed to be the same inventor, according to a set of criteria; ***Critical stage!***

2nd stage: methodology for matching names

If two records display the same name, ***how do we know they refer to the same inventor?***

- Compare the two records according to data given in the patents (*address*, *tech field*, *assignee*, *etc*.); give "scores" for each matching criteria.
- Examine other links between the two records (shared "partner", cite each other); give scores if link holds.
- Compute overall score for the pair, if above threshold then make the "match", e.g. decide it is same inventor.

Set threshold & scoring system considering the two types of error: over/under-matching

Criteria of varying strength

- ***Strong criteria*** –sufficient for match Soundex, ***same full address, self-citation, shared partner.***
- ***Medium criteria*** –sufficient if ***identical*** names: ***same middle name, same Zip*** *(US only).*
- ***Weak criteria***: combination needed for a match:

  * ***Size threshold:*** Two individuals located in ***New York*** weaker info than if located in a ***small town***; same for assignee (***IBM*** weaker than small startup).

  * ***Name frequency:*** If ***"rare"*** name, higher prob. that two individuals with same name are the same guy.

Matrix of size thresholds and scores

(in terms of number of patents)

	Thresholds for Name frequency		Score	
	"Rare" < 10	"Common" ≥ 10	Below threshold	Above threshold
City	2,500	1,322 *(median)*	100	80
Assignee	2,500	500	100	80
Patent class	30,000	18,597 *(median)*	80	50

Impose Transitivity

A matched to ***B***

B matched to ***C***,

A matched to ***C***

Even though ***A*** and ***C*** may have little or nothing in common, except of course for (at least) same Soundex-coded name

Diagnostics: ex post average matching score

Diagnostic tools critical: otherwise too large a file to assess the "quality" of the matches done.

Compute average matching score for each "group" of matched inventors:

- for each pair (permutation) compute the actual matching score (e.g. the sum of the points of each common criteria); there are ***m=n (n-1)/2*** permutations.
- Compute the average as: $\frac{\sum_i^m pairwise\ score_i}{m}$

The numbers…

Original patent file:

- ***2,139,313*** patents
- average number of inventors per patent: ***2.01***

- ***4,298,912*** "records" (*patents* x *inventors*)

End result:

Matching rendered 1,565,780 distinct inventors

- Average number of patents per inventor: 2.7

Number of patents per inventor
(or how much "action" can we expect?)

Out of 1,565,780 inventors, the number of inventors with,

- just one patent: 911,943 (58%)
- 2 or more: 653,837 (42%)
- 5 or more: ***203,302*** (13%)
- 10 or more: ***73,072*** (5%)

Mobility of inventors across assignees

Number of assignees	**Number of inventors** *(with patents>1)*
1	437,256
2	158,737
3	38,727
4	11,838
5+	7,279
Total:	653,838
# of movers	**216,581 (33%)***

* But probably overstates moves: need to consolidate assignee codes.

Mobility of inventors across US states

Number of states	**Number of US inventors** *(with patents>1)*
1	292,333
2	39,123
3	4,334
4	556
5+	120
Total:	336,466
# of movers	**44,133 (13%)**

Net international flows

Country	Moves in	Moves out	Net
Canada	1392	1554	-162
Switzerland	702	693	9
Germany	1551	1701	-150
France	665	665	0
GB	2181	2809	-628
Israel	248	219	29
Italy	205	186	19
Japan	1114	1244	-130
Korea	371	270	101
Netherlands	453	527	-74
Taiwan	275	176	99
US	8041	7272	769

Flows of Inventors across types of assignees

To

From	Corporate	Individual	Govermnt.	Total
Corporate	298,472	57,698	5,379	361,549
Individual	59,487	0	1,799	61,286
Govermnt.	7,710	2,024	1,834	11,568
Total	365,669	59,722	9,012	434,403
Net	4,120	-1,564	-2,556	

Empirical analysis: determinants and impact of mobility

(i) Each inventor one observation (descriptive):

- ***Number of moves*** = ***f*** (inventor-level variables, citations received, controls).

Contrast US, Japan, ROW; 600K obs: # of inventors with patents >1

(ii) **Panel** – observation: each patent of each inventor

- ***Citations to this patent*** = ***f*** (controls, previous history of inventor, ***moved or not***)
- ***Probability of moving in this patent*** = ***f*** (controls, previous history, ***quality of previous patents***)

1.3 million obs: records of US inventors with patents >1

Panel of Inventors

Each observation: a patent of an inventor; hence have sequence for each inventor, can study what "causes" moves, and what the moves "cause".

Look at:

1. "Quality" of patents (e.g. citations received) as function of moves
2. Probability of moving (across assignees, geography) as function of past history,

But endogeneity/simultaneity! Need e.g. Arrelano Bond (1991) Dynamic Panel

I. Indicators of patent "quality" as function of mobility

Citations to this patent =

= ***f*** (controls, previous history of inventor, ***moved or not***)

Patents of US inventors only, 1.3 million obs.

Dependent Variable: Citations received **OLS, US inventors, 1.3 M obs. *(White SE)***			
Variable	**Coefficient**	Std. Error	**t-Statistic**
C	**638.2**	2.24	**284.9**
Application year	**-0.32**	0.001	**-285.1**
Patent Sequence	**-0.02**	0.0005	**-46.9**
Patent Sequence2**	**4.7E-05**	1.3E-06	**36.7**
# of Partners	**0.17**	0.004	**38.5**
Mean Past Citations	**0.38**	0.003	**118.8**
MOVED ASSIGNEES	**0.08**	0.019	**4.5**
MOVED GEOGRAPHY	**0.12**	0.019	**6.2**
ΣMoves_Assignees (-1)	**-0.02**	0.003	**-7.3**
ΣMoves_Geography(-1)	**-0.004**	0.002	**-2.1**
Includes controls for 6 tech categories			
R^2 = 0.24			

Impact of mobility on other indicators of patent "importance"

Run similar regressions with other indicators of patent "importance" as dependent variables (all other "*X's*" included as well):

- ***Generality***: (1 – Herfindahl on patent classes of citations received)
- ***Originality***: (1 – Herfindahl on patent classes of citations made)
- ***Number of claims***

Impact of moves on qualitative indicators of patents

(t-values in parenthesis, based on White SE)

	Dependent Variable:			
	Citations	**Generality**	**Originality**	**Claims**
Mean of dep.var.	***5.3***	***0.34***	***0.40***	***14***
MOVED ASSIGNEES	**0.08** *(4.5)*	**0.006** *(7.9)*	**0.006** *(10.0)*	**1.1** *(34.4)*
MOVED GEOGRAPHY	**0.12** *(6.2)*	**0.004** *(5.0)*	**0.001** *(2.1)*	**0.46** *(15.3)*
Σ Moves Assignees (-1)	**-0.02** *(-7.3)*	**0.0003** *(2.1)*	**0.002** *(18.4)*	**0.11** *(15.8)*
Σ Moves Geography(-1)	**-0.004** *(-2.1)*	**2.42E-05** *(0.3)*	**0.0001** *(1.4)*	**-0.01** *(-3.4)*
R2	**0.25**	**0.14**	**0.13**	**0.13**

Results – variables other than moves

- Earlier patents tend to be more "valuable", but sort of quadratic *(well, bottoms at 250…)*
- Highly significant lagged (mean) dependent variables (sort of "fixed effects")
- Highly positive impact of number of ***partners***
- Weighting by scores does not make much of a difference, but fit improves.

(Don't have to worry about multicollinearity…)

Results: Impact of moves

I. Contemporary moves:

- Having just moved across assignees and/or location has a ***positive impact on the "value"*** of patent taken at the new place.
- Moving to a new assignee has a ***stronger*** impact than moving geographically (except as measured by citations).

II. Previous moves:

- No impact of previous ***geographical*** moves
- Past ***assignee moves*** have a small positive impact on originality and claims, a small negative small impact on citations

For those that moved assignees *(257,401 obs.)*
Dep. Variable: citations received
Baseline: move to and from government

Variable	Coefficient	Std. Error	t-Statistic
Move ***to*** corporation	***2.42***	0.08	28.6
Move ***to*** "garage"	1.84	0.09	20.2
Move ***from*** corporation	0.22	0.09	2.4
Move ***from*** "garage"	0.15	0.097	1.5

All other controls included
R^2=0.19

For those that moved assignees…

- Moving ***to*** a corporation results in much better patents than moving to a government agency or ***to*** a "garage";
- Origin does not matter much: coming from a corporation barely better than coming from the government or "garage".
- Very similar results for the other qualitative indicators, except for originality: no significant differences there.

II. Correlates of mobility: To move or not to move…

Examine the decision to “move or not”, of each inventor at each point in time (actually with each additional patent),

- as a function of the “past history / performance” of the inventor, i.e. the “quality” of his/her previous patents,
- and controls.

Dependent Variable: *move to other assignee*

Binary Logit - 1,062,037 obs. *(from 2nd patent on)*

1st set of estimates

Variable	Coefficient	Std. Error	z-Statistic
C	-52.53	0.928	-56.6
Application year	0.087	0.0006	139.7
Year of 1st patent	-0.061	0.0006	-100.8
Patent_Sequence	-0.085	0.0010	-83.7
Partners(-1)	-0.017	0.0016	-10.7
Corporate(-1)	-0.461	0.0073	-62.9
Σ Moves Assignees (-1)	0.272	0.0026	105.9
Includes also Tech category dummies			

move to other assignee - continued
Binary Logit – 2nd set of estimates

Variable	Coefficient	Std. Error	z-Statistic
F_CITATIONS(-1)	**0.003**	0.0002	**10.9**
GENERALITY(-1)	**0.248**	0.0100	**24.7**
ORIGINALITY(-1)	**-0.152**	0.0099	**-15.7**
CLAIMS(-1)	**-0.001**	0.0002	**-4.6**

Move to other geographical location

Binary Logit – includes all other controls

Variable	Coefficient	Std. Error	z-Statistic
F_CITATIONS(-1)	0.003	0.0002	13.4
GENERALITY(-1)	0.142	0.0101	14.2
ORIGINALITY(-1)	-0.111	0.0097	-11.4
CLAIMS(-1)	-0.005	0.0002	-19.2

How your patenting "history" affects the probability of moving?

You are ***more likely to move***, both to another assignee and/or to another location,

- Early on in your patenting career
- If you had fewer partners
- If you do ***NOT*** work for a corporation
- If you have a previous history of moving
  (sort of fixed effects)

Probability of moving - cont. 1

You are ***more*** likely to move if, prior to the move, you have patents that are,

- more "general" *(very robust)*
- more highly cited

You are ***less*** likely to move if you have

- More original patents
- Patents with more claims

Probability of moving - cont. 2

- More “general” patents: useful in a wider range of fields;
- highly cited: more down-the-line applications,

presumably more “movable” inventors

- But why negative sign on claims and on originality?

Probability of moving - cont.3

"Importance" in the sense of more citations and higher ***generality*** is hard to observe/verify in advance, hence inventor probably has better inside information than employer, the latter will not act to retain inventor.

Originality and ***Claims*** known by the time the patent is filed, hence employer will try to preempt move.

But then this should hold just for corporations, not for others!

Move to other assignee *II*

include interactions w/lagged corporate dummy

Variable	Coefficient	Std. Error	z-Statistic
F_Citations(-1)	**0.007**	0.0007	**10.2**
F_Citations(-1)*Corp(-1)	***-0.005***	*0.0008*	***-6.9***
Generality(-1)	**0.382**	0.024	**15.8**
Generality(-1)*CORP(-1)	***-0.166***	*0.026*	***-6.3***
Originality(-1)	**0.138**	0.023	**5.9**
Originality(-1)*CORP(-1)	***-0.354***	*0.025*	***-13.9***
CLAIMS(-1)	**0.004**	0.0006	**6.4**
CLAIMS(-1)*CORP(-1)	***-0.005***	*0.0006*	***-8.7***
Corporate(-1)	**-0.166**	0.015	**-10.8**

Point estimates: coming from a corporation versus coming from a "garage" or government

Variable	Point estimates
F_Citations(-1): non-corp	0.007
Corp	0.002
Generality(-1): non-corp	0.38
Corp	0.22
Originality(-1): non-corp	0.14
Corp	*-0.22*
Claims(-1): non-corp	0.0037
Corp	*-0.0017*

Mobility of Inventors from or to software

Dependent Variable: MOVE_ASS obs: 97211			
Method: ML - Binary Logit; 10 iterations; QML (Huber/White) standard errors;			
Sample: 1 364281 IF PAT_SEQ>1 AND (SOFT_PAT=1 OR SOFT_PAT(-1)=1)			
Variable	**Coefficient**	**Std. Error**	**z-Statistic**
C	-36.19661	4.780065	-7.572410
APPYEAR	0.114341	0.003063	37.33432
FIRSTYEAR	-0.095229	0.002507	-37.98036
PAT_SEQ	-0.078756	0.004168	-18.89362
Partner_Count(-1)	0.011334	0.004146	2.733784
CORP(-1)	-1.267292	0.083694	-15.14192
SOFT_PAT(-1)	-0.541547	0.027280	-19.85123
Includes patent categories			

Mobility of Inventors – logit reg, cont.

MOVE_ASS_CUM(-1)	0.25	0.013	19.47
CRECEIVE(-1)	0.005	0.0007	6.62
GENERAL(-1)	0.24	0.05	5.36
CLAIMS(-1)	-0.006	0.001	-5.73
ORIGINAL(-1)	-0.32	0.04	-7.18
Log (Assignee_size)	**-0.21**	**0.012**	**-17.80**
Log (Assignee_size)*Soft_Pat	**-0.06**	**0.004**	**-14.08**
Log (Assignee_size)(-1)	**-0.09**	**0.012**	**-7.45**
LR statistic (19 df)	14708.94	McFadden R2: 0.21	
Obs with Dep=1	11210		

Patent citations – correlates

Dependent Variable: **CRECEIVE**			
Sample: 1 364281 IF PAT_SEQ>1; obs: 269793			
White Heteroskedasticity-Consistent Standard Errors & Covariance			
Variable	**Coefficient**	**Std. Error**	**t-Statistic**
C	784.2141	8.402587	93.33007
APPYEAR	-0.394468	0.004207	-93.77185
M_CRECEIVE(-1)	0.789497	0.009866	80.02456
PAT_SEQ	-0.018512	0.000633	-29.24048
PART_COUNT	0.134062	0.007927	16.91107
Patent categories included			
SOFT_PAT	0.716159	0.127542	5.615067
MOVE_ASS	0.192870	0.058744	3.283217
Log (Assignee_size)	**0.06**	**0.01**	**6.15**
Log(Assignee_size)*Soft_Pat	**-0.08**	**0.02**	**-5.33**
R-squared	0.320016		

Some “take aways”

- Inventors that have already produced “better” patents tend to move more often;
- Conversely, moving seems to impact favorably the “quality” of subsequent patents.
- 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*).
- Employers successfully preempt moving of inventors with patents that are “better” in observable ways (claims and originality).

Further work

- Deal with endogeneity, apply Arrelano-Bond model, bring in data on firms, markets
- Study impact of inventors' mobility on firms' innovative performance, *both ways!*
- 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.

Annex: Descriptive regressions
Each inventor one observation

Summary variables of their patenting career:

- ***Number of moves*** (e.g. across countries, assignees, cities, etc.),
- ***Means*** for their patents: citations, number of "partners" (co-inventors), % of their patents in tech categories, etc.
- ***Timing***: year of first and last patent, hence
  "***Age"*** = 1999 – year first patent
  Duration = year last patent – year first patent

annex – cont.

Regress ***number of moves*** on:

- Controls (e.g. number of patents, duration)
- % of patents in 6 tech categories; "tech focus" (1 – Herf of patents in tech categories)
- "Age," number of partners
- ***"Importance" of patents: # of forward citations*** *(i.e. citations received)*
- Contrast US, Japan, Rest of the World (ROW)

annex – cont. 2

Purely ***descriptive*** regressions, since endogeneity/ selection:

> ***Movers may be already "special" (e.g. produce more important patents), and/or the moving itself may impact them.***

But at least differences across countries may be informative.

Negative Binomial regressions; inventors with more than one patent (with 1 could not observe move).

Distribution of assignee moves per inventor		
# of moves	# of inventors	% of inventors
0	437,256	66.88
1	125,553	19.20
2	47,823	7.31
3	18,357	2.81
4	9,606	1.47
5	5,166	0.79
6	3,228	0.49
7	1,886	0.29
8	1,339	0.20
9	852	0.13
10 – 19	2,350	0.36
20 - 49	423	0.04
100 - 200	8	0.00
Total	653,837	100.00

Dep. variable: number of moves across assignees **Negative Binomial - obs: 653,837** *(base: ROW)* *1st set of results*			
Variable	**Coefficient**	Std. Error	**z-statistic**
AGE	**-0.016**	0.0006	**-26.6**
AGE * US	**-0.011**	0.0007	**-15.8**
AGE * JAPAN	**0.011**	0.001	**9.2**
DURATION	**0.095**	0.0005	**185.7**
PARTNERS	**0.012**	0.002	**5.9**
PARTNERS*US	**-0.021**	0.003	**-7.4**
PARTNERS*JAPAN	**0.056**	0.004	**14.0**
TECH_FOCUS	**0.775**	0.009	**86.9**
Number_Patents	**0.048**	0.0005	**90.1**
Dummies for tech fields, interacted w/US, Japan			

Moves across assignees – cont.

Variable	Coefficient	Std. Error	z-statistic
F_CITATIONS	**0.008**	**0.0009**	**9.9**
F_CITATIONS *US	***0.0005***	***0.0009***	***0.6***
F_CITATIONS *JP	***-0.011***	***0.0019***	***-6.1***
US	0.122	0.0135	9.1
JAPAN	-1.074	0.023	-46.9
LR INDEX	***0.15***		

Movement of inventors across assignees associated with…

1. "***Younger***" inventors
2. Having more patents in ***Drugs and Medical***
3. Having more ***partners***
4. Being more ***technologically focused*** (i.e. their patents more concentrated in tech categories)
5. Having more "***important***" patents (but the opposite in Japan: only Japanese "losers" move)

Similar results for moves across countries

Main differences between the US, Japan and ROW

1. US inventors tend to move ***more*** across assignees, ***less*** across countries
2. Japanese inventors tend to move ***much less*** than ROW and US inventors
3. Inventors that move across ***countries*** have more important patents, ***not*** so US inventors
4. Japanese inventors that move across assignees have ***less important*** patents, and are older than ROW, US
5. Inventors in Drugs and Medical move a lot, particularly Japanese inventors.