Knowledge spillovers and the geography of duplicated inventions: an analysis from patent citations

Stefano Baruffaldi
École Polytechnique Fédérale de Lausanne
stefano.baruffaldi@epfl.ch

Julio Raffo
World Intellectual Property Organization
julio.raffo@wipo.int

AQR – Workshop

October, 2012

Barcelona, Spain

Motivation

- "far from being odd or curious or remarkable, the pattern of independent multiple discoveries in science is in principle the dominant pattern" (Merton, 1961).
- Duplication and the cumulative nature of knowledge makes harder for future generations of inventors to be innovative and leads to diminishing returns on R&D investments (Jones, 2009; Kortum, 1993; Gómez, 2011;).
- "..a significant and growing number of very expensive lawsuits occur each year because firms have invested millions of dollars on the research, development, and commercialization of technology that is legedly owned by others" (Bessen and Meurer, 2008)

➢ *Does geography affect duplication?*

Duplication in literature

- « It is in the air » (Merton, 1979; Lamb, 1984; Murray & O'Mahony, 2007)
  - Knowledge is cumulative
  - The same invention can be reached by more than one inventor when the preconditions are present.
  - State of the equipment, level of scintific ideas and the readiness to accept or require a certain invention.

- The theory of Communication: (Merton, 1979; Brannigan, 1983)
  - "unnecessary duplication" resulting from imperfections in the channels of communication among scientists.
  - Failure of the notice function of patents. (Bessen and Meurer, 2008)

- Competition (Dasgupta and Maskin, 1983; Aghion et al., 2002; Encaoua et al., 2005)
  - Competition induces inventors to chose overly similar research projects
  - Patent races and inventing around
  - Pre-emptive patenting

Duplication and knowledge spillovers

- **Unawareness:** lack of awareness of replicating others' research efforts leading to independent duplications.

  - *Diffusion of knowledge makes less likely that inventions are involuntarily replicated.*

- **Competition:** the inventor is aware of replicating others' research and voluntarily engages in this effort leading to competitive duplications.

  - *Higher knowledge spillovers increase duplications when there are incentive to compete on the same technological path.*

Hypotheses

H1: Duplication is more likely to occur close in space for recent technologies.

H2: Duplication is more likely to occur far in space for not recent technologies.

H3: Geographic distance affect more duplication of complex technologies than discrete technologies.

Methodology: Patent citations

Assumption: the knowledge disclosure in patent document is less than perfect. (Atal and Bar, 2010; Walsh et al., 2007; Bessen and Meurer, 2008)

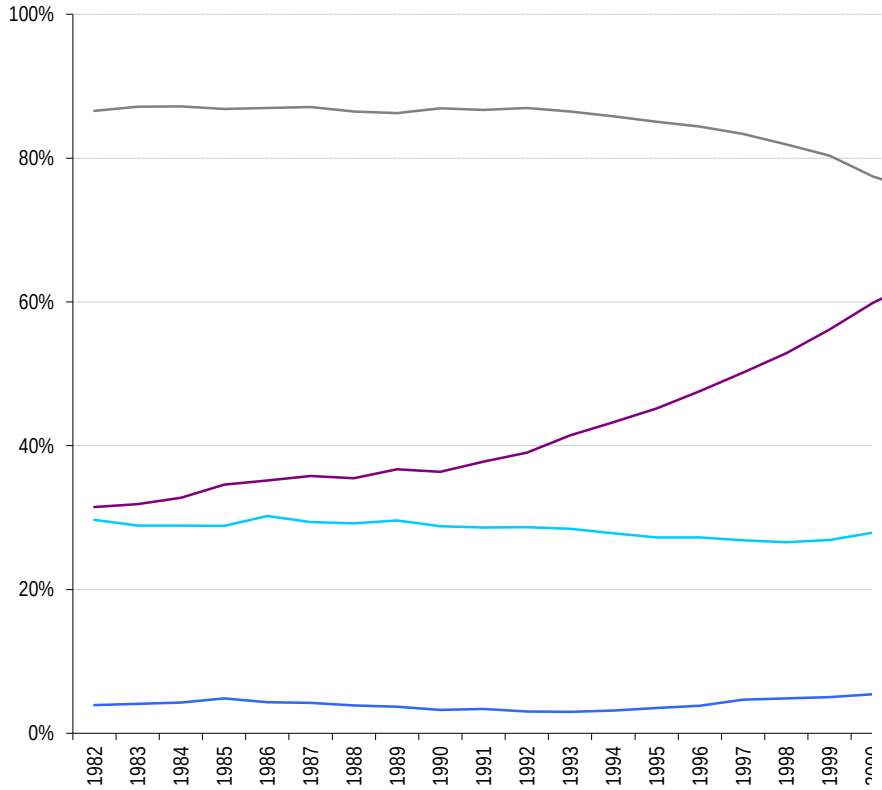
Duplication revealed by patent citations:

Category	Definition	Construct
A	Citations to documents defining the state of the art and not prejudicing novelty	Knowledge base
X, E	Citations to patents that **taken alone** imply **the lack of novelty** of a claimed invention.	Revealed duplication

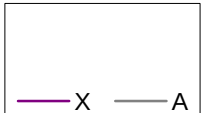
Methodology: Data

- Patent citations data from EPO's Worldwide Patent Statistics Database (PATSTAT, September 2010):
  - 994,193 EPO to EPO citations (605,181 citing patents)
  - 1982-2007
- Localization of inventors from the OECD's REGPAT Database (December 2010).
- Each NUTS 3 region has been geo-localized in order to construct distance measure between citations.
- Citations categories and origin available

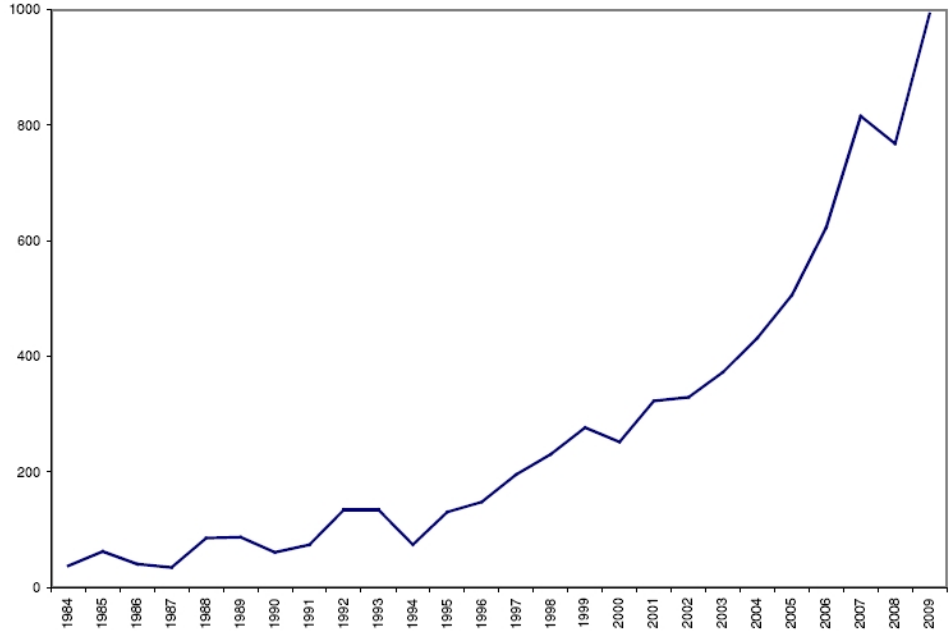
Descriptives: Citations categories over years



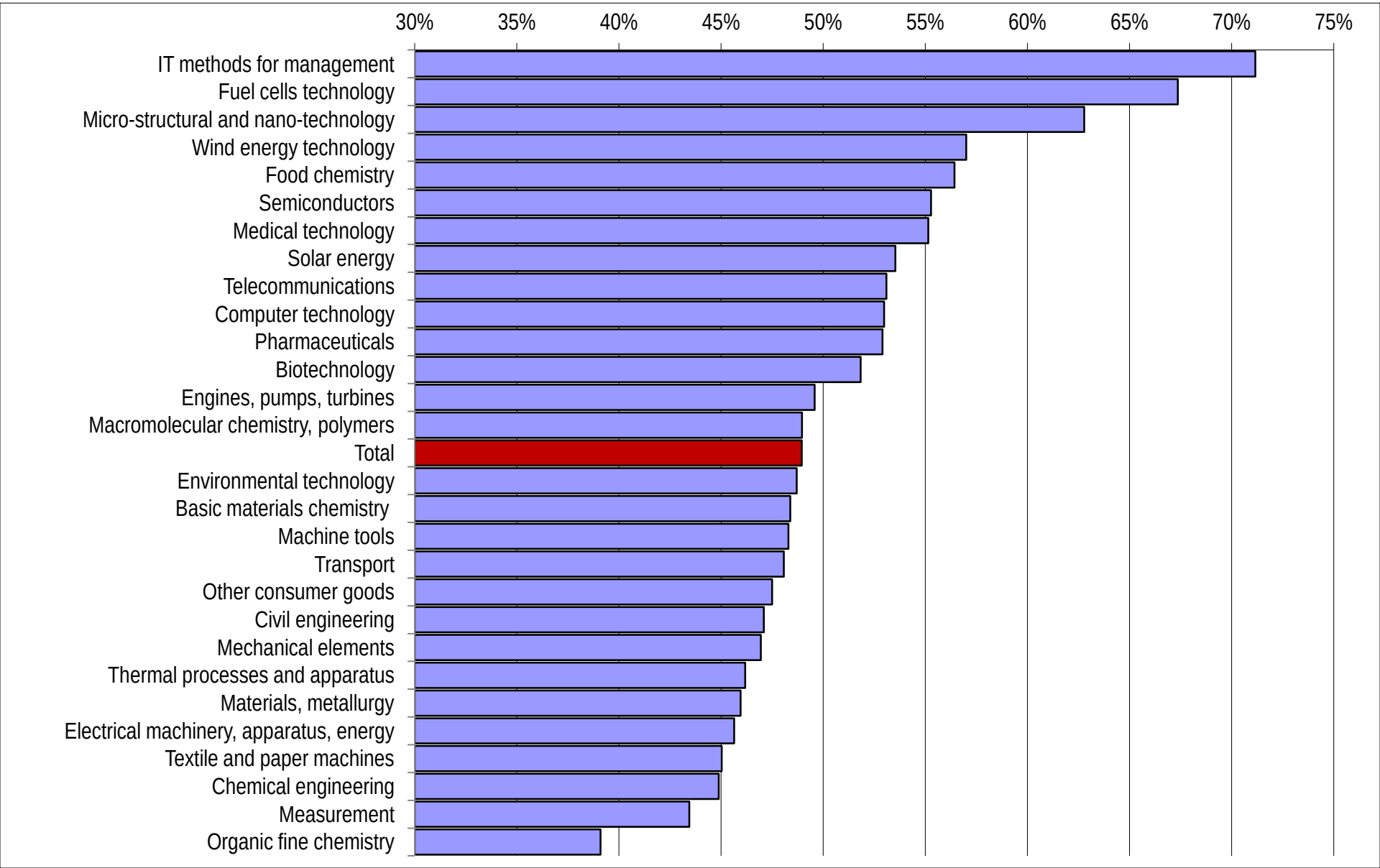
(Share of patents with at least one cit. of the categor



Number of patent lawsuit filings involving software patents



Descriptives:
patents with at least 1 X citation over sectors



Methodology: Model

Linear probability model with fixed effects:

$$P(Y_{ij} = 1 \mid X_{ij}) = \beta_0 + \beta_1 A_{ij} + \beta_2 I_{ij} + \beta_3 T_{ij} + \beta_4 G_{ij} + \beta_5 T_{ij} \times G_{ij} + a_i + \varepsilon_{ij}$$

Y_{ij} = 1 if patent citation linking patent *i* cites and patent *j* is an "X" or an "E", 0 otherwise

X_{ij} = regressors depending on *i* and *j*

A_{ij} = 1 if the applicant of patent *i* is the same inventor of patent *j*

I_{ij} = 1 if the inventor of patent *i* is the same inventor of patent *j*

Controls

T_{ij} = time distance between patent *i* and patent *j* → Time lag control

G_{ij} = geographic distance between patent *i* and patent *j* → H1 (-)

$T_{ij} \times G_{ij}$ = time and geographic distance interaction → H2 (+)

a_i = citing patent fixed effect

Results: full sample

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
same_inventor	0.0575***	0.0535***	0.0523***	0.0475***	0.0441***
	(0.00225)	(0.00227)	(0.00228)	(0.00235)	(0.00241)
same_applicant	-0.0117***	-0.0139***	-0.0107***	-0.0121***	-0.0157***
	(0.00196)	(0.00197)	(0.00202)	(0.00222)	(0.00223)
prior_diffy	-0.0113***	-0.0136***	-0.00829***	-0.00895***	-0.00917***
	(0.000181)	(0.000232)	(0.000231)	(0.000203)	(0.000197)
Distmin	5.50e-07***	-2.71e-06***			
	(1.80e-07)	(2.73e-07)			
c.prior_diffy#c.distmin		6.00e-07***			
		(3.77e-08)			
same_ctry_any			0.0241***		
			(0.00239)		
same_ctry_any#c.prior_diffy			-0.00686***		
			(0.000324)		
same_reg2_any				0.0382***	
				(0.00272)	
same_reg2_any#c.prior_diffy				-0.00941***	
				(0.000368)	
same_reg3_any					0.0498***
					(0.00290)
same_reg3_any#c.prior_diffy					-0.0110***
					(0.000401)
Constant	0.332***	0.345***	0.323***	0.324***	0.324***
	(0.00137)	(0.00157)	(0.00153)	(0.00127)	(0.00122)
F	1359.92***	1139.36***	1189.14***	1223.15***	1239.52***
Observations	626,726	626,726	626,726	626,726	626,726
Number of groups	237,714	237,714	237,714	237,714	237,714

Results: sample without self citations

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
prior_diffy	-0.00881***	-0.00985***	-0.00794***	-0.00855***	-0.00869***
	(0.000224)	(0.000308)	(0.000261)	(0.000234)	(0.000229)
distmin	2.29e-07	-9.85e-07***			
	(2.08e-07)	(3.23e-07)			
c.prior_diffy#c.distmin		2.22e-07***			
		(4.53e-08)			
same_ctry_any			0.00970***		
			(0.00297)		
same_ctry_any#c.prior_diffy			-0.00274***		
			(0.000417)		
same_reg2_any				0.00813**	
				(0.00410)	
same_reg2_any#c.prior_diffy				-0.00232***	
				(0.000603)	
same_reg3_any					0.00549
					(0.00517)
same_reg3_any#c.prior_diffy					-0.00186**
					(0.000765)
Constant	0.331***	0.337***	0.329***	0.332***	0.332***
	(0.00158)	(0.00195)	(0.00166)	(0.00139)	(0.00134)
F	775.33***	524.97***	533.18***	522.14***	518.96***
Observations	441,394	441,394	441,394	441,394	441,394
Number of groups	206,748	206,748	206,748	206,748	206,748

Results: Marginal effects of distance along time

$$\frac{\partial P(Y_{ij} = 1 \mid}{\partial G} = \beta_4 + \beta_5 T$$

Distance **decrease** the probability of duplication

Distance **increase** the probability of duplication

Time lag (years)	*Model 2 (Km distance)*	*Model 3 (same country)*	*Model 4 (same region)*	*Model 5 (same province)*
0	**-9.85e-07*** **	0.009704***	0.00813**	0.005487
	(3.23E-07)	(0.002967)	(0.004099)	(0.005174)
1	**-7.63e-07*** **	0.006964***	0.00581	0.003632
	(2.90E-07)	(0.002666)	(0.003661)	(0.004616)
2	**-5.41e-07** **	0.004223*	0.003491	0.001777
	(2.60E-07)	(0.0024)	(0.003276)	(0.004124)
3	-3.19e-07	0.001483	0.001172	-7.8E-05
	(2.36E-07)	(0.002182)	(0.002965)	(0.003725)
4	-9.66e-08	-0.00126	-0.00115	-0.00193
	(2.18E-07)	(0.002027)	(0.002753)	(0.003453)
5	1.26e-07	-0.004	-0.00347	-0.00379
	(2.09E-07)	(0.001951)	(0.002663)	(0.003337)
6	**3.48e-07***	-0.00674**	-0.00579**	-0.00564*
	(2.09E-07)	(0.001963)	(0.002709)	(0.003393)
7	**5.70e-07*** **	-0.00948***	-0.0081***	-0.0075**
	(2.19E-07)	(0.002061)	0.002882	(0.003615)
8	**7.92e-07*** **	-0.01222***	-0.01042***	-0.00935**
	(2.37E-07)	(0.002233)	(0.003163)	(0.003974)
…	….	…	…	…
31	**5.90e-06*** **	-0.07524***	-0.06376***	-0.05202***
	(1.17E-06)	(0.010867)	(0.015811)	(0.020036)

Results: Robustness

	Model 1 (Complex technologies)	*Model 2 (Discrete technologies)*	*Model 3 (Semiconductors)*	*Model 4 (Mechanical elements)*	*Model 5 (Basic materials chemistry)*
prior_diffy	-0.0146***	-0.0122***	-0.0213***	-0.0110***	-0.0116***
	(0.00043)	(0.00048)	(0.00186)	(0.00166)	(0.000853)
distmin	-3.35e-06***	-1.07e-06*	-6.89e-06***	1.04e-06	6.26e-07
	(4.95e-07)	(6.30e-07)	(1.74e-06)	(2.27e-06)	(1.15e-06)
c.prior_diffy# c.distmin	8.03e-07***	3.16e-07***	1.65e-06***	5.56e-07*	-1.01e-07
	(7.11e-08)	(7.80e-08)	(2.68e-07)	(2.91e-07)	(1.40e-07)
same_inventor	0.0619***	0.0493***	0.0605***	0.0143	0.0309***
	(0.00440)	(0.00486)	(0.0179)	(0.0181)	(0.00850)
same_applicant	-0.0205***	-0.0027	-0.0398***	-0.00997	-0.0166**
	(0.00360)	(0.00438)	(0.0145)	(0.0158)	(0.00757)
Constant	0.3498***	0.3460***	0.402***	0.285***	0.346***
	(0.00284)	(0.00376)	(0.0113)	(0.0122)	(0.00639)
F	352.43***	236.11***	34.5***	12.41***	64.84***
Observations	206,039	109,483	16,033	9,292	31,435
Number of groups	81,622	39,144	6,379	3,849	10,417

N° of sectors	21/35	4/35	10/35
% of sample	70%	14%	16%

Conclusions

- Geography affect the rate of duplication
- Bivalent effect of proximity; trade - off
- Imperfect disclosure of knowledge in patents
- Contribution to the literature on the meaning of patent citations

Limits:

- Which proximity?
- Underlying R&D efforts and incentives?

Thank you!