Concordance and complementarity in IP instruments

Marco Grazzi [1] Chiara Piccardo [2] Cecilia Vergari [2]

[1]Università Cattolica del Sacro Cuore, Milano
[2]University of Bologna

Oct, 24th 2018
IPSDM 2018, Alicante

Motivations / Main Focus

- IP activities and firm performance: focus on the (virtual) universe of Italian firms
    - <250 employees firms are not negligible (75% TM, 60% pat)
- To what extent IP instruments (patent and trademark) are complementary
- Diversification in knowledge and production.
    - Firms "know more than they make" Patel and Pavitt (1997) or the opposite (Dosi et al., 2017)?
- Overlapping/coherence in bundles of patents and trademark.
- How we do it ($\approx$ *what you ll have to bear with*):
    1. BvD data on firms, patents and trademarks (AIDA, Italian firms)
    2. Patent to trademark (through economic sector) crosswalk, Lybbert and Zolas (2014) and Goldschlag et al. (2016)
    3. The concordance measure we propose

Previous literature on IP bundles

- Patents and trademarks are both meaningful proxies for firms' innovation activity
  - Schautschick and Greenhalgh (2016); Flikkema et al. (2014); Helmers and Rogers (2010); Mendonça et al. (2004)
- Literature on the relationship between IPRs and firms' performances tend to focus on top R&D investors, large and medium firms, quoted companies, or specific industries
  - Castaldi and Dosso (2018); Llerena and Millot (2013); Greenhalgh and Rogers (2006); Graham and Somaya (2004)
- Only few contributions are conducted on small firms
  - Rogers et al. (2007)

Previous literature on IP bundles

- Market values, venture capital funding, profits and survival are used as proxies for firms' performances
  - Dosso and Vezzani (2017); Zhou et al. (2016); Llerena and Millot (2013); Greenhalgh and Rogers (2012); Sandner and Block (2011); Buddelmeyer et al. (2010); Helmers and Rogers (2010); Wagner and Cockburn (2010); Jensen et al. (2008); Rogers et al. (2007)
- Fewer studies analyze the impact of IPRs on firms' growth
  - Guzman and Stern (2015); Castaldi and Dosso (2018)

Firm Level Data

- Analisi Informatizzata delle Aziende (AIDA data, 2005-2014)
    - The dataset virtually covers the universe of Italian limited liability manufacturing firms independently of their size
- AMADEUS dataset (2005-2014)
    - The dataset includes information on the stock of patents and registered trademarks
- We employ the BvD provided patent-to-firm match
    - Lotti and Marin (2013) propose an improvement of the match for largest Italian firms

IP and firms distributions: whole sample

Year	Firms	Firms with tm(%)∗	Firms with pat (%)∗	Firms with tm and pat (%)∗	Firms with $conc_{i,t} > 0$ (%)∗∗	Num of tm	Num of pat
2006	116507	4732	8616	1798	1623	13190	77312
		(4.062)	(7.359)	(1.543)	(90.267)		
2007	123007	5340	8884	2018	1822	15582	82797
		(4.341)	(7.222)	(1.641)	(90.287)		
[...]							
2012	135299	8616	9419	2891	2606	30384	101552
		(6.368)	(6.962)	(2.137)	(90.142)		
2013	132731	9070	9062	2927	2644	32612	98872
		(6.833)	(6.827)	(2.205)	(90.331)		
2014	129253	8863	8608	2830	2556	31740	93229
		(6.857)	(6.660)	(2.190)	(90.318)		

Note. We only consider firms operating in manufacturing sectors (we exclude firms operating in the following 2-digit ATECO 2007 code: 12 and 33).

∗ *In brackets, percentage of total firms.*

∗ ∗ *In brackets, as a percentage of firms with trademarks and patents.*

IP and firms distributions in 2014: micro, small, med, large

Size	Firms	Firms with tm(%)$*$	Firms with pat (%)$*$	Firms with tm and pat (%)$*$	Firms with $conc_{i,t} > 0$ (%)$*$	Num of tm (%)$**$	Num of pat (%)$**$
0-9	73641	1668	1673	223	187	3076	5910
	(57.996)	(18.944)	(19.556)	(7.911)	(7.345)	(9.736)	(6.366)
10-49	44148	3805	3805	933	805	9356	20141
	(34.769)	(43.214)	(44.477)	(33.097)	(31.618)	(29.612)	(21.695)
50-249	8016	2653	2432	1208	1114	11246	27599
	(6.313)	(30.131)	(28.428)	(42.852)	(43.755)	(35.594)	(29.728)
250^{+}	1172	679	645	455	440	7917	39187
	(0.923)	(7.712)	(7.539)	(16.140)	(17.282)	(25.058)	(42.211)

Note. We only consider firms operating in manufacturing sectors (we exclude firms operating in the following 2-digit ATECO 2007 code: 12 and 33). The number of firms in this Table differ from the previous Table because 2276 firms do not have information on their size, measured in terms of workers, in 2014.

$*$ *In brackets, percentage of total firms, total firms with trademarks, patents, trademarks and patents, concordant trademarks and patents, respectively.*

$**$ *In brackets, percentage of total number of trademarks and patents, respectively.*

Proxies for Innovative activities

- Stock (and yearly flow) of (both granted and not granted) patents applied to United States Patent and Trademark Office (USPTO), European Patent Office (EPO), or Italian Patent and Trademark Office (IPTO)
  - We only account for patents applied in the last 20 years
- Stock (and yearly flow) of registered trademarks filed at the United States Patent and Trademark Office (USPTO) or at the European Union Intellectual property office (EUIPO)
  - We consider trademarks applied before or in the year of interest, that expire after the referred year

Intuition behind the crosswalk

"Algorithmic approach to construct concordances between the IPC system [...] and industry classification systems that organize economic data, such as the [...] International Standard Industrial Classification (ISIC), the North America Industry Classification System (NAICS) and the Harmonized System (HS). 'Algorithmic Links with Probabilities' (ALP) approach mines patent data using keywords extracted from industry descriptions and processes the resulting matches using a probabilistic framework."

(Getting patents and economic data to speak to each other, Lybbert and Zolas, 2014)

SKIP Building a concordance measure for IP bundles

- In our data (Amadeus) a patent is associated to more than one IPC, whereas a trademark to just one NICE
- We convert each 3-digit IPC code and 2-digit NICE code to 2-digit ISIC codes.
- For each firm and year, we identify the sets of 3 digit IPC and 2 digit NICE codes associated to the stock of patents and trademarks, respectively.

Building a firm-level concordance measure for IP bundles

1. At the firm level, from the stock of patents and trademarks we derive the related sets of IPC and NICE codes corresponding to such stock.
2. At the IPC-NICE level, we compute the overlapping coefficient for each L-K pair of (3 digit) IPC and (2 digit) NICE codes,

$$overlap^{i,t}_{L-K} = \sum_{j=1}^{N_{LK}} min\{p_L(isic_j), p_K(isic_j)\}, \tag{1}$$

   $p_L(isic)$ is the likelihood of the linkage between IPC and each ISIC code and $p_K(isic)$ the likelihood of the match between NICE and each ISIC code

3. At the firm level, we compute the degree of concordance in the IP bundles, summing up and normalizing by $N^i_L * N^i_K$

$$conc_{i,t} = (\sum_{L=1}^{N^i_L} \sum_{K=1}^{N^i_K} overlap^{i,t}_{L-K})/(N^i_L * N^i_K) \tag{2}$$

IPRs technological concordance: an example

ID	PAT	IPC
1	1	A61
1	1	H02
1	2	C09
1	3	A61
1	4	A61

ID	TM	NICE
1	1	2
1	2	2
1	3	2
1	4	2
1	5	2
1	6	2
1	7	2
1	2	2
1	8	2
1	9	35

Concordance in IP bundles: an example

IPC	ISIC	P_{IPC}(ISIC)
A61	10	0.0361579
A61	20	0.0505005
A61	21	0.8616829
A61	32	0.0302412
A61	36	0.0214176
C09	20	0.9659607
C09	23	0.0340393
H02	24	0.0637551
H02	26	0.0888389
H02	27	0.8212442
H02	28	0.0261618

NICE	ISIC	P_{NICE}(ISIC)
2	2	0.0491944
2	20	0.1203539
2	25	0.8065651
2	41	0.0238867
35	46	0.1524157
35	63	0.0744019
35	69	0.0240318
35	70	0.0710176
35	73	0.4864157
35	78	0.0490504
35	82	0.0307544
35	90	0.0424586
35	94	0.0694538

$$conc_{i,t} = (0.0505005+0+0.1203539+0+0+0)/(3*2) = 0.028475733 \quad (3)$$

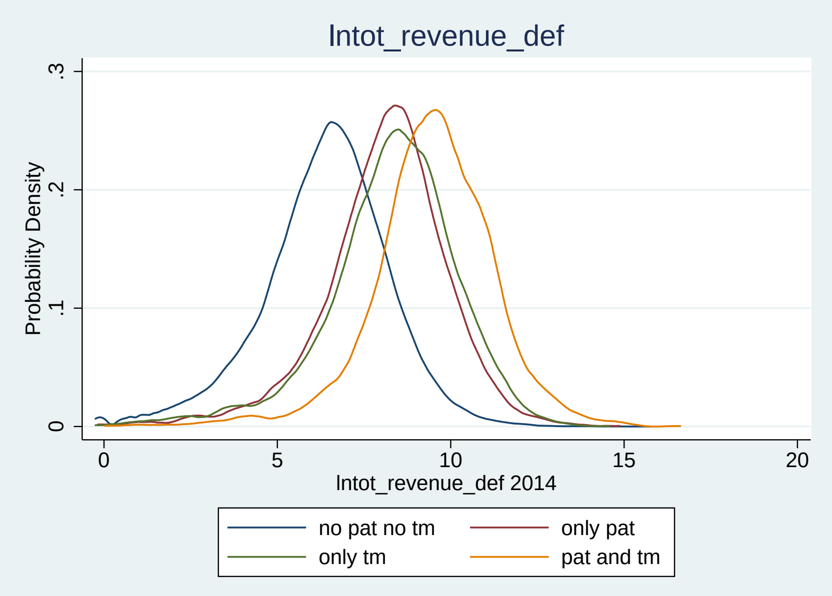
The concordance also enables to assess the internal dynamics of IP bundles: of all the patent-TM matches, in 55% to 50% cases the patent pre-dates the TM

External validity on a different dataset

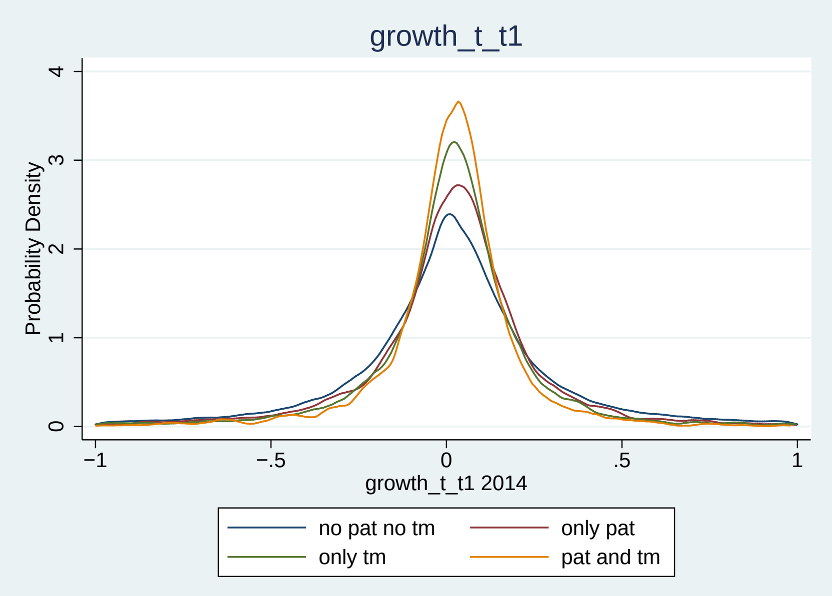
Product Rank	(1)	(2)
1	0.155	0.501
2	0.081	0.376
3	0.042	0.327
4	0.041	0.297
5	0.027	0.267
6	0.023	0.239
7	0.029	0.243
8	0.024	0.205
9	0.023	0.235
10	0.015	0.188

(1) Of all patents that link to products, 15% are linked to prod. ranked 1^{st}.
(2) On average, 50% of firms have their highest rank product "backed" by a patent.

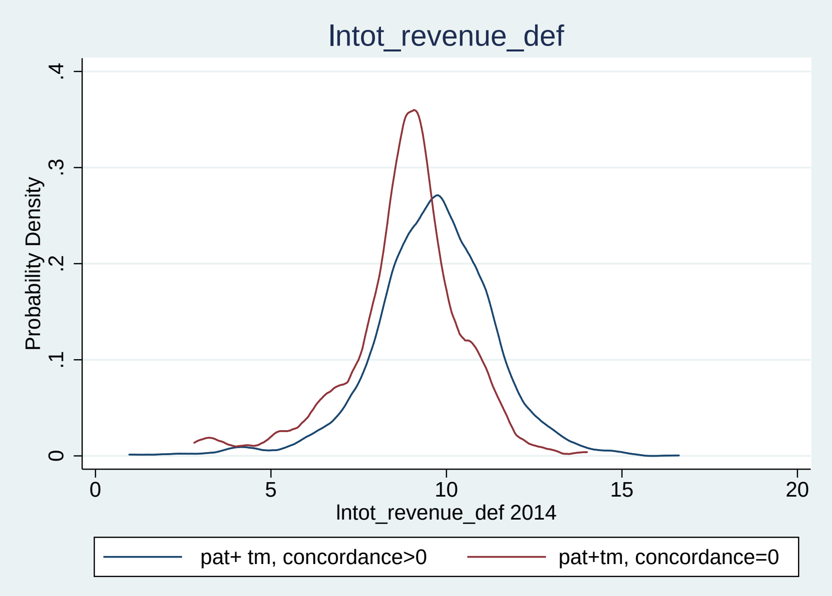
Total Sales: *non-innovative* vs *innovative* firms



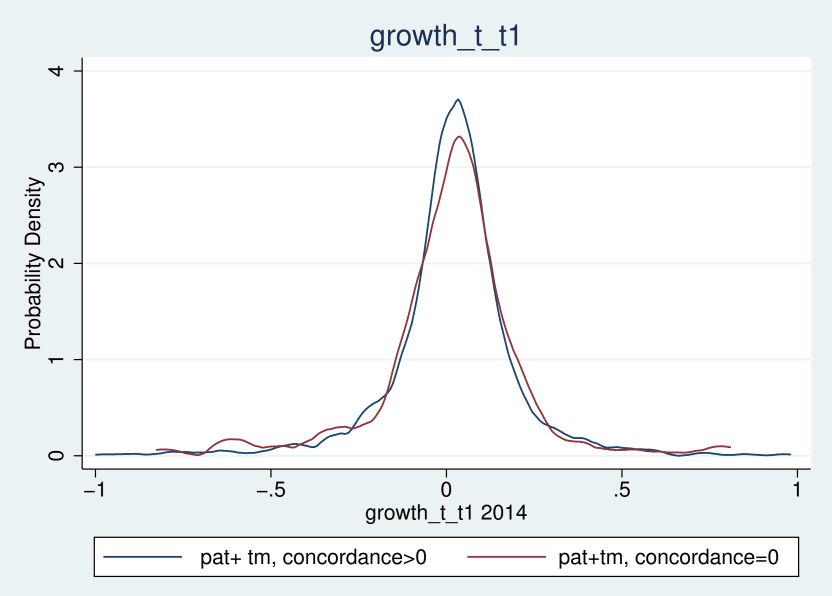
Growth: *non-innovative* vs *innovative* firms



Total Sales: firms with and without concordant IPRs



Growth: firms with and without concordant IPRs



Pooled OLS Model specifications

- Dependent var: firms' (log) total revenues and yearly growth rate
- Explanatory var(s): 3 binary indicators equal to 1 if firms own, respectively, only patents, only trademarks, both; firms' size; productivity (baseline specification); the degree of technological concordance between IPRs (extended specification)
- Controls: 2-digit industry dummy variables, geographical area (North, Centre and South of Italy) and year dummies
- Why OLS: (1) for many firms IPR does not change over time, would be absorbed by FE; (2) estimating fully saturated dummy variables models with OLS is fully general (Angrist and Pischke, 2008).

$$Y_{i,t} = c + \beta_1 tm_{i,t-1} + \beta_2 pat_{i,t-1} + \beta_3 both_{i,t-1} + \\ + \beta_4 ln(workers_{i,t-1}) + \beta_4 ln(LP_{i,t-1}) + X'_{i,t}\alpha + u_{i,t} \quad (4)$$

Pooled OLS models: whole sample

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(totrev_i,t)	ln(totrev_i,t)	ln(totrev_i,t)	growth_i,t	growth_i,t	growth_i,t
ln(workers_i,t-1)	0.898***	0.965***	0.965***	-0.0183***	-0.00521*	-0.00523*
	(0.000920)	(0.00437)	(0.00437)	(0.000616)	(0.00310)	(0.00310)
ln(LP_i,t-1)	0.742***	0.782***	0.782***	-0.0538***	-0.0110	-0.0110
	(0.00166)	(0.0138)	(0.0138)	(0.00123)	(0.00960)	(0.00958)
tm_i,t-1	0.306***			0.0525***		
	(0.00434)			(0.00253)		
pat_i,t-1	0.169***			0.0358***		
	(0.00367)			(0.00234)		
both_i,t-1	0.322***			0.0759***		
	(0.00480)			(0.00303)		
conc_i,t-1			-0.0210			-0.0153
			(0.0275)			(0.0185)
_cons	2.772***	2.599***	2.605***	0.286***	0.123**	0.127**
	(0.00873)	(0.0797)	(0.0789)	(0.00584)	(0.0572)	(0.0552)
N	686252	17929	17929	684323	17917	17917
r2	0.781	0.877	0.877	0.0248	0.0444	0.0445
F	55579.6	2069.7	2008.3	357.6	28.30	27.45

*Note. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*
In each specification we include ATECO 2 digit sectors, geographical areas and years dummy variables.

Pooled OLS models: high technology sectors

	(1) ln(totrev_i,t)	(2) ln(totrev_i,t)	(3) ln(totrev_i,t)	(4) growth_i,t	(5) growth_i,t	(6) growth_i,t
ln(workers_i,t-1)	0.908***	0.946***	0.945***	0.000197	-0.00619	-0.00646
	(0.00348)	(0.0105)	(0.0106)	(0.00233)	(0.00791)	(0.00795)
ln(LP_i,t-1)	0.705***	0.731***	0.729***	-0.0429***	-0.00895	-0.00963
	(0.00704)	(0.0457)	(0.0459)	(0.00498)	(0.0173)	(0.0176)
tm_i,t-1	0.264***			0.0374**		
	(0.0210)			(0.0146)		
pat_i,t-1	0.0971***			0.0212**		
	(0.0155)			(0.0103)		
both_i,t-1	0.361***			0.0524***		
	(0.0177)			(0.0112)		
conc_i,t-1			0.117			0.0419
			(0.0823)			(0.0568)
_cons	2.484***	2.433***	2.426***	0.178***	0.155	0.153
	(0.0338)	(0.238)	(0.237)	(0.0230)	(0.136)	(0.134)
N	29798	1596	1596	29704	1595	1595
r2	0.831	0.904	0.904	0.0120	0.0119	0.0124
F	9456.4	923.9	860.6	18.85	1.805	1.699

*Note. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*
In each specification we include geographical areas and years dummy variables.

Pooled OLS models: medium-high technology sectors

	(1) ln(totrev_i,t)	(2) ln(totrev_i,t)	(3) ln(totrev_i,t)	(4) growth_i,t	(5) growth_i,t	(6) growth_i,t
ln(workers_i,t-1)	0.914***	0.974***	0.974***	-0.0113***	-0.00485	-0.00484
	(0.00181)	(0.00666)	(0.00666)	(0.00129)	(0.00444)	(0.00444)
ln(LP_i,t-1)	0.710***	0.830***	0.830***	-0.0571***	-0.00539	-0.00538
	(0.00354)	(0.0182)	(0.0182)	(0.00273)	(0.0133)	(0.0133)
tm_i,t-1	0.269***			0.0526***		
	(0.00853)			(0.00537)		
pat_i,t-1	0.115***			0.0316***		
	(0.00620)			(0.00424)		
both_i,t-1	0.275***			0.0696***		
	(0.00790)			(0.00522)		
conc_i,t-1			-0.0720*			-0.0315
			(0.0403)			(0.0238)
_cons	2.484***	1.936***	1.951***	0.260***	0.0506	0.0569
	(0.0173)	(0.0952)	(0.0948)	(0.0130)	(0.0728)	(0.0720)
N	154962	7731	7731	154495	7726	7726
r2	0.803	0.867	0.867	0.0273	0.0628	0.0630
F	35563.2	2025.6	1870.5	232.6	46.19	42.36

*Note. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*
In each specification we include geographical areas and years dummy variables.

Pooled OLS models: medium-low technology sectors

	(1) ln(totrev_i,t)	(2) ln(totrev_i,t)	(3) ln(totrev_i,t)	(4) growth_i,t	(5) growth_i,t	(6) growth_i,t
ln(workers_i,t-1)	0.919***	1.004***	1.004***	-0.0230***	-0.00562	-0.00558
	(0.00165)	(0.00829)	(0.00828)	(0.00111)	(0.00607)	(0.00605)
ln(LP_i,t-1)	0.743***	0.859***	0.859***	-0.0621***	-0.00218	-0.00207
	(0.00301)	(0.0240)	(0.0240)	(0.00227)	(0.0211)	(0.0210)
tm_i,t-1	0.311***			0.0606***		
	(0.00940)			(0.00528)		
pat_i,t-1	0.210***			0.0427***		
	(0.00581)			(0.00344)		
both_i,t-1	0.322***			0.0713***		
	(0.00898)			(0.00518)		
conc_i,t-1			-0.0297			-0.0139
			(0.0612)			(0.0334)
_cons	2.334***	2.048***	2.052***	0.293***	0.0705	0.0721
	(0.0143)	(0.120)	(0.121)	(0.0104)	(0.0965)	(0.0972)
N	231639	4485	4485	231072	4481	4481
r2	0.772	0.872	0.872	0.0356	0.0531	0.0531
F	38535.1	1519.7	1406.5	442.4	24.47	22.47

*Note. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*
In each specification we include geographical areas and years dummy variables.

Pooled OLS models: low technology sectors

	(1) ln(totrev_i,t)	(2) ln(totrev_i,t)	(3) ln(totrev_i,t)	(4) growth_i,t	(5) growth_i,t	(6) growth_i,t
ln(workers_i,t-1)	0.893***	0.957***	0.958***	-0.0196***	-0.00166	-0.00158
	(0.00154)	(0.00912)	(0.00904)	(0.000952)	(0.00644)	(0.00633)
ln(LP_i,t-1)	0.780***	0.749***	0.750***	-0.0451***	-0.0135	-0.0134
	(0.00261)	(0.0303)	(0.0301)	(0.00183)	(0.0218)	(0.0215)
tm_i,t-1	0.406***			0.0572***		
	(0.00612)			(0.00334)		
pat_i,t-1	0.166***			0.0281***		
	(0.00873)			(0.00494)		
both_i,t-1	0.369***			0.0784***		
	(0.00981)			(0.00600)		
conc_i,t-1			0.0518			0.0120
			(0.0506)			(0.0365)
_cons	2.427***	2.611***	2.602***	0.215***	0.0363	0.0343
	(0.0124)	(0.152)	(0.150)	(0.00815)	(0.106)	(0.102)
N	269853	4117	4117	269052	4115	4115
r2	0.749	0.868	0.868	0.0153	0.0212	0.0213
F	46006.9	1150.8	1089.9	196.5	11.32	10.37

Note. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
In each specification we include geographical areas and years dummy variables.

Conclusions

- IPRs exert a positive impact on firms' performance
- Patents and trademarks exhibit complementarity
- The degree of technological concordance between IPRs on firms' performance is not significant
- Italy, given the small number of firms with both patents and trademarks is not the ideal testbed.

References I

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Buddelmeyer, H., P. H. Jensen, and E. Webster (2010). Innovation and the determinants of company survival. *Oxford Economic Papers 62*(2), 261–285.

Castaldi, C. and M. Dosso (2018). From r&d to market: using trademarks to capture the market capability of top r&d investors. Technical report, Joint Research Centre (Seville site).

Dosi, G., M. Grazzi, and D. Moschella (2017). What do firms know? what do they produce? a new look at the relationship between patenting profiles and patterns of product diversification. *Small Business Economics 48*(2), 413–429.

Dosso, M. and A. Vezzani (2017). Firm market valuation and intellectual property assets. Technical report, Joint Research Centre (Seville site).

Flikkema, M., A.-P. De Man, and C. Castaldi (2014). Are trademark counts a valid indicator of innovation? results of an in-depth study of new benelux trademarks filed by smes. *Industry and Innovation 21*(4), 310–331.

Goldschlag, N., T. J. Lybbert, and N. J. Zolas (2016, March). An 'Algorithmic Links with Probabilities' Crosswalk for USPC and CPC Patent Classifications with an Application Towards Industrial Technology Composition. Working Papers 16-15, Center for Economic Studies, U.S. Census Bureau.

References II

Graham, S. and D. Somaya (2004). The Use of Patents, Copyrights, and Trademarks in Software: Evidence from Litigation. In D. Guellec (Ed.), *Patents, Innovation and Economic Performance*. OECD.

Greenhalgh, C. and M. Rogers (2006). Market Valuation of UK Intellectual Property: Manufacturing, Utility and Financial Services Firms. In *The Management of Intellectual Property*, Chapters, Chapter 7. Edward Elgar Publishing.

Greenhalgh, C. and M. Rogers (2012). Trade marks and performance in services and manufacturing firms: Evidence of schumpeterian competition through innovation. *Australian Economic Review 45*(1), 50–76.

Guzman, J. and S. Stern (2015). Where is silicon valley? *Science 347*(6222), 606–609.

Helmers, C. and M. Rogers (2010). Innovation and the survival of sew firms in the uk. *Review of Industrial Organization 36*(3), 227–248.

Jensen, P. H., E. Webster, and H. Buddelmeyer (2008). Innovation, technological conditions and new firm survival. *Economic Record 84*(267), 434–448.

Llerena, P. and V. Millot (2013). Are Trade Marks and Patents Complementary or Substitute Protections for Innovation. Working Papers of BETA 2013-01, Bureau d'Economie Thèorique et Appliquèe, UDS, Strasbourg.

Lotti, F. and G. Marin (2013, June). Matching of PATSTAT applications to AIDA firms: discussion of the methodology and results. Questioni di Economia e Finanza (Occasional Papers) 166, Bank of Italy, Economic Research and International Relations Area.

References III

Lybbert, T. J. and N. J. Zolas (2014). Getting patents and economic data to speak to each other: an 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. *Research Policy 43*(3), 530–542.

Mendonça, S., T. S. Pereira, and M. M. Godinho (2004). Trademarks as an indicator of innovation and industrial change. *Research Policy 33*(9), 1385–1404.

Patel, P. and K. Pavitt (1997). The technological competencies of the world's largest firms: Complex and path-dependent, but not much variety. *Research Policy 26*(2), 141–156.

Rogers, M., C. Helmers, and C. Greenhalgh (2007). An analysis of the characteristics of small and medium enterprises that use intellectual property. Technical report, London: Report for the UK Intellectual Property Office.

Sandner, P. G. and J. Block (2011). The market value of r&d, patents, and trademarks. *Research Policy 40*(7), 969–985.

Schautschick, P. and C. Greenhalgh (2016). Empirical studies of trade marks–the existing economic literature. *Economics of Innovation and New Technology 25*(4), 358–390.

Wagner, S. and I. Cockburn (2010). Patents and the survival of internet-related ipos. *Research Policy 39*(2), 214–228.

Zhou, H., P. G. Sandner, S. L. Martinelli, and J. H. Block (2016). Patents, trademarks, and their complementarity in venture capital funding. *Technovation 47*, 14–22.

Zolas, N., T. J. Lybbert, and P. Bhattacharyya (2017). An 'algorithmic links with probabilities' concordance for trademarks with an application towards bilateral ip flows. *The World Economy 40*(6), 1184–1213.