DOUBLE DISCLOSURES AND THE NEGOTIATION OF SCIENTIFIC CREDIT IN RESEARCH TEAMS

L. Cassi, F. Lissoni, F. Montobbio, L. Zirulia

EPIP 2015 Conference - Glasgow, September 2-3

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

université de BORDEAUX

Motivation: Team science and scientific credit

- **The puzzle of cientific credit in team science → collective output vs individual credit**

- **"Sequel"** studies: how does science distribute credit across team members? → social norms transform collective into individual credit
  - resurgence of studies on Matthew effect → e.g.
    - accidental name obliterations *(Simcoe and Waguespack, 2011)*
    - retractions *(Jin et al., 2013)*

- **"Prequel"** studies:
  - norms of credit attribution as determinants of team formation (co-publication patterns) *(Bikard et al., 2015; Gans and Murray, 2013 & 2014)*
  - welfare analysis of co-publication

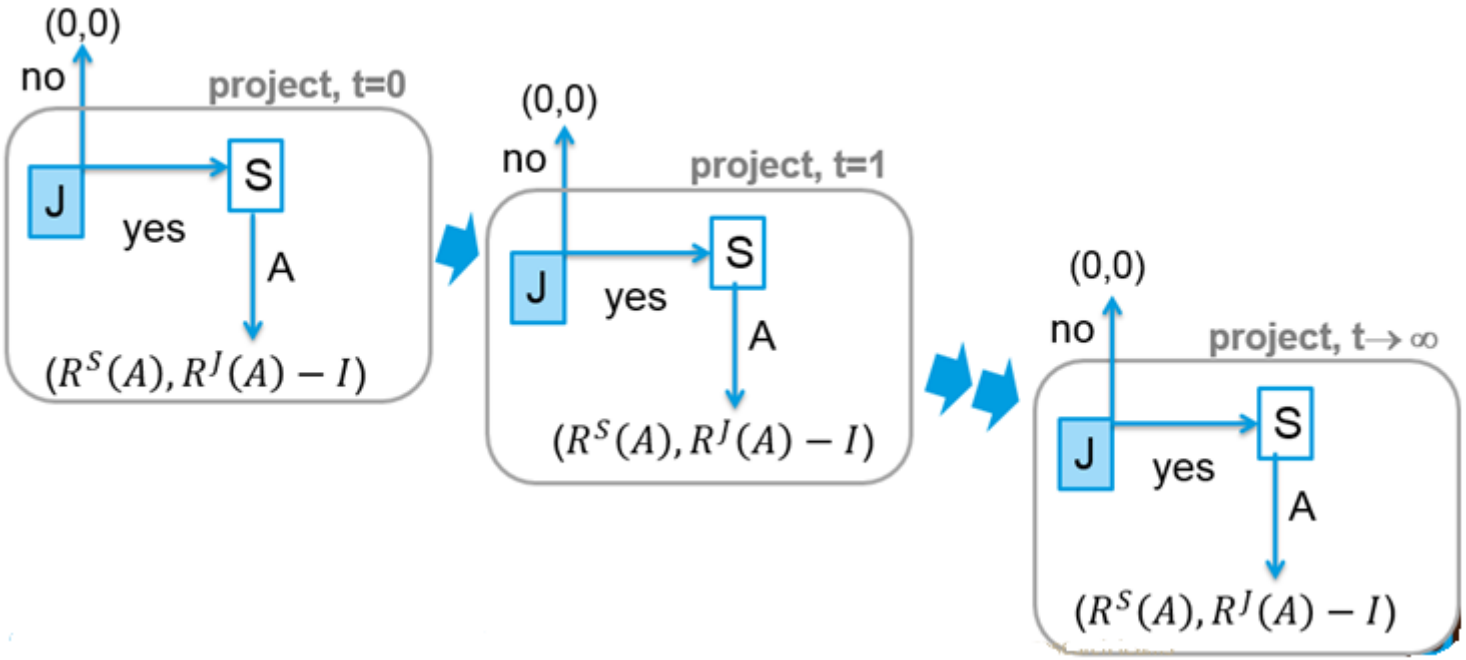
Our contribution

- General idea: to bring more sophisticated (up-to-date?) notions of scientific authorship into economics
- To the "sequel" literature:
  - Team members negotiate the individual credit distribution → exploit social norms of attribution
  - Negotiation outcome depends upon individual preferences & bargaining power
  - Negotiation concerns all of team's activities → "double disclosures" as an exemplary case
- To the "prequel" literature: Team stability as a function of negotiation outcome
- Previous papers: Lissoni et al. (2013), Lissoni & Montobbio (2015)
  → Limitations: Italian data only (pretty old, too) ; "sequel" modelling only

“Double disclosure” and attribution rights

- *Double disclosure*: the same research result is both published and patented → “patent-publication pairs” (PPPs)
- *Attribution rights: authorship & inventorship*
  - Both have an economic value (scientific credit; expert reputation; patent royalties)
  - This value differ across individuals:
    - ✓ junior vs senior scientists: marginal value of (first) authorship is higher at early career stages
    - ✓ male vs female scientists: marginal value of inventorship is lower for women scientists *(Ding et al., 2013; Thursby and Thursby, 2005; Whittington and Smith-Doerr, 2005)*

› Team: senior S and junior J → Sequence of 2-stage/1-period projects



› S makes take-it-or-leave-it offer to J: *A(1st authorship, inventorship) …*
… BUT no commitment → J exits at end-of-period if S reneges on A!

› Participation costs: J's = I ; S's = zero

› Payoffs (value of attribution rights):

- R_1^S (R_1^J) = value of first authorship for *S (J)*
- $R_{N1}^S(R_{N1}^J)$ = value of non-first authorship for *S (J)*
- Value of inventorship: v/n, where $n = 1,2$: nr of inventors on patent

Theoretical model (cont.)

- Intertemporal utility functions with discount factor δ^S and δ^J.
  - $\delta^J = \delta$ with probability 1
  - $\delta^S = \begin{cases} \delta \text{ with probability } \mu \\ 0 \text{ with probability } 1-\mu \end{cases} \rightarrow \begin{cases} \text{forward looking : max long term payoff} \\ \text{myopic: max current payoff } \rightarrow A = (S,S) \end{cases}$
  - δ as objective probability of team's success in each period t → followed by S' probability μ to behave "rationally" and reward J as promised

→ Multiple equilibria

→ Theoretical question: is "exclusion" offer $A = (J,S)$ a "sustainable" equilibrium? IF YES, EXCLUSION DOES NOT LEAD *J* TO QUIT THE TEAM!

From theoretical model to empirical analysis

- Most teams have more than 2 scientists → what about middle authors?
  - Higher or lower probability of exclusion from patents, compared to first authors?
    - *Higher as compensation for non-first authorship*
    - *Lower if lower contribution to research (incl. guest and gift authorship)*
  - If excluded, higher or lower probability of staying in the team, compared to first authors?
    - *Lower, if first authorship was due/expected*
    - *Higher, if lower contribution to research (no legal claim to inventorship)*

EMPIRICAL TEST: 2 REGRESSIONS

1. Exclusion regression:

$$Prob(exclusion)$$

$$\boldsymbol{Prob(no\ inventor|author)_{ij} =}$$

$$\boldsymbol{= f(seniority; gender; byline\ position; controls)}$$

$$i{=}individual \quad ; \quad j{=}patent\text{-}related\ publication$$

2. Renewed co-authorship regression :

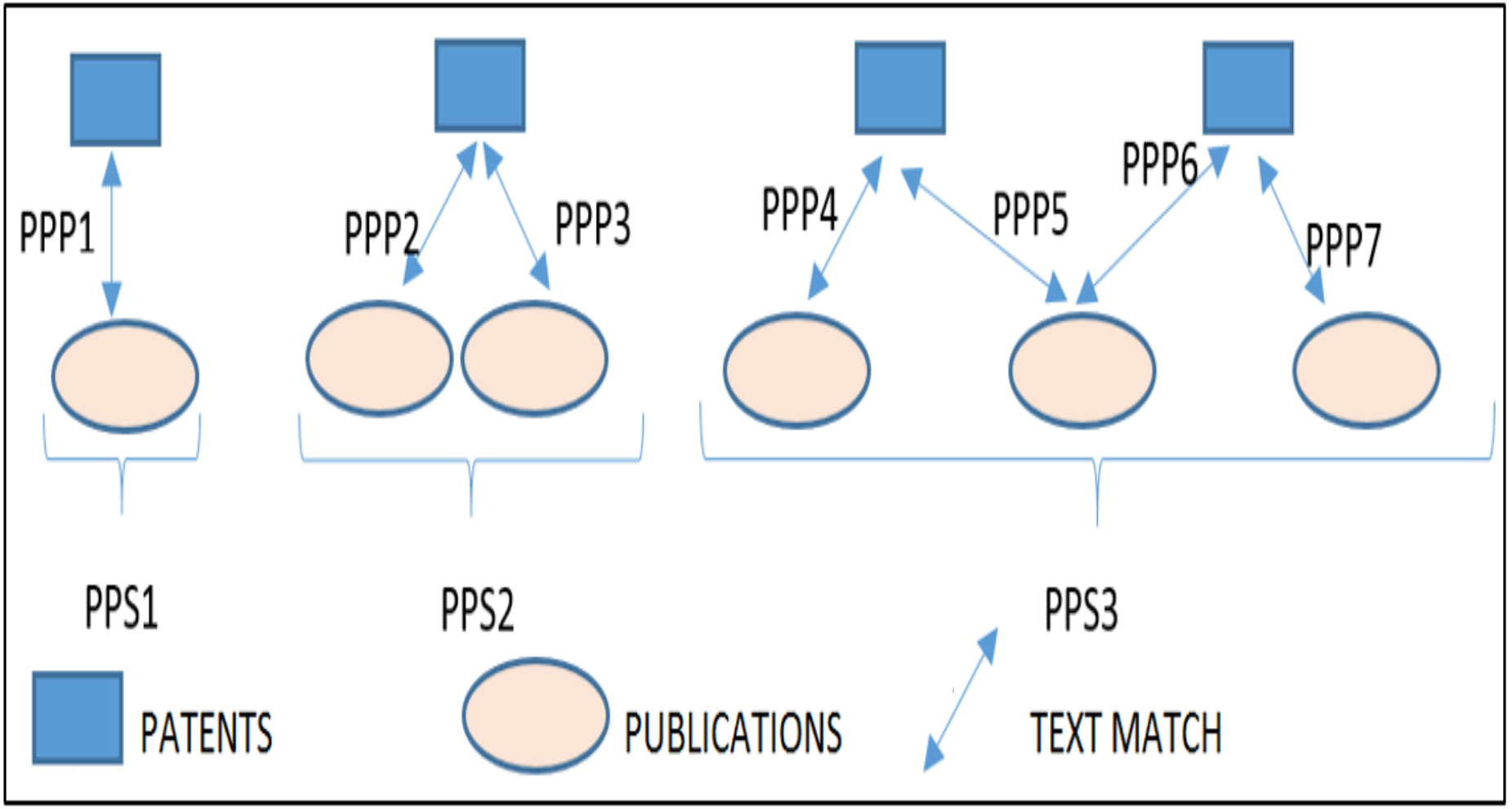
$$\boldsymbol{Prob(new\ co-authorship)_{ij} =}$$

$$\boldsymbol{= f(exclusion; byline\ position; ; controls)}$$

$$i{=}individual \quad ; \quad j{=}patent\text{-}related\ publication$$

Data & methodology:

Figure A.1 Patent-publication pairs (PPPs) and patent-publication sets (PPSs): definition



Data & methodology:

PUBLICATION DATA : Web of Science

- authors with the same name and initials of patent inventors → restrictions by technology/discipline & time range {priority year-2, priority year+2} : ~4.6 million matches

TEXT MATCHING & FILTERING:

- text analysis of titles&abstracts of all matches (weighted bag-of-words method) → {0,1} similarity index)
  - ✓ **3p90 sample**: top 10% of the first percentile (952 PPSs, 1154 patents, 2299 publications)
  - ✓ **2p95 sample**: top 5% of the first percentile (561 PPSs, 668 patents, 1142 publications)

“Exclusion” regression

OBSERVATIONS: author-publication dyads

DEPENDENT VARIABLE *exclusion* : =1 if the author is not an inventor in the PPS

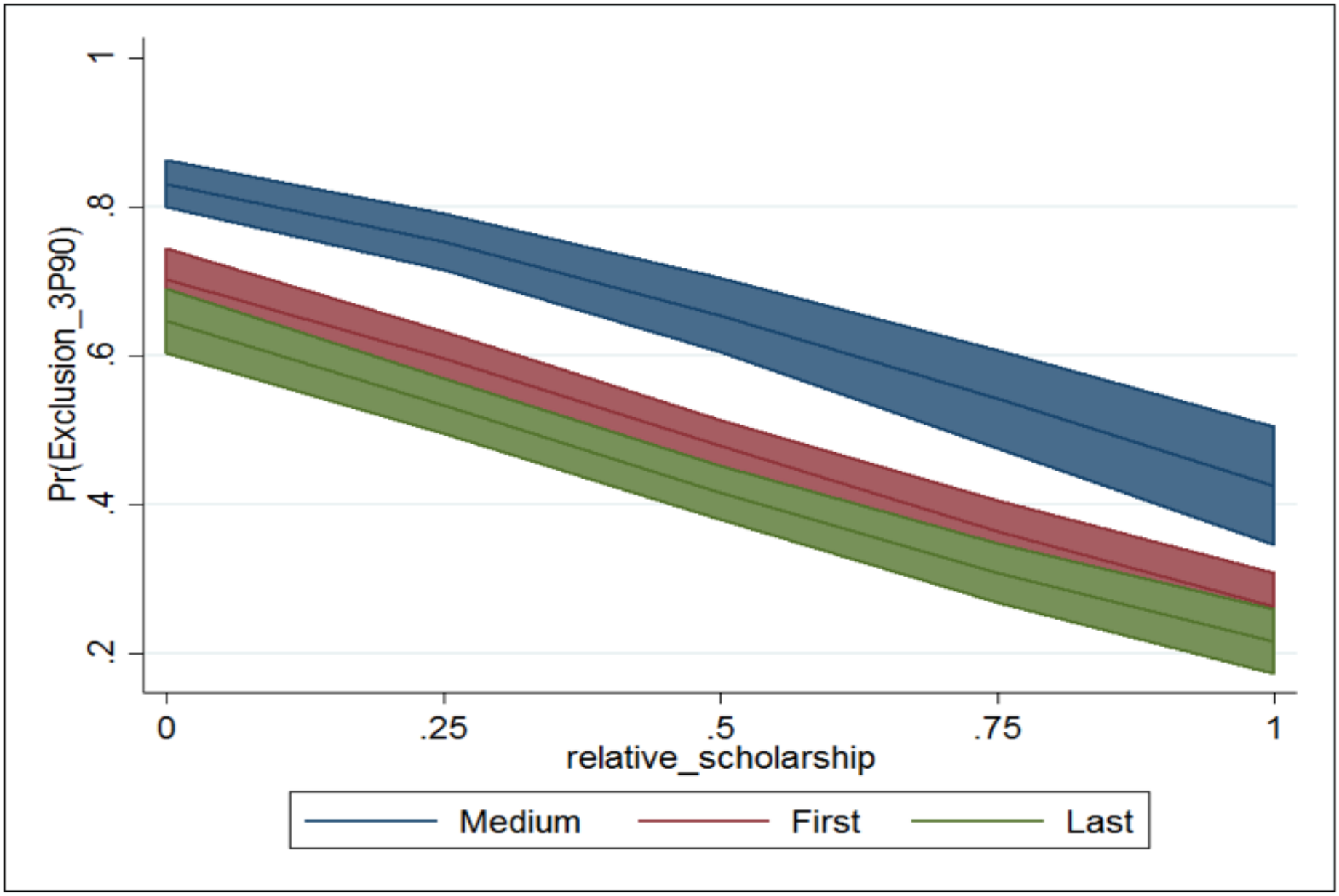
EXPLANATORY VARIABLES OF INTEREST:

- ✓ Author’s position in the byline: *First, Last*; and *Middle* (reference case)
- ✓ Author’s gender: *Female* (unavailable for ~30% observations)
- ✓ Author’s seniority:
  - ➢ $relative\ scholarship_{ij} = \frac{stock_{ij} - \min(stock_j)}{\max(stock_j) - \min(stock_j)} = [0,1]$

    where $stock_{ij}$ = nr publications by author *i* at time of publication *j*

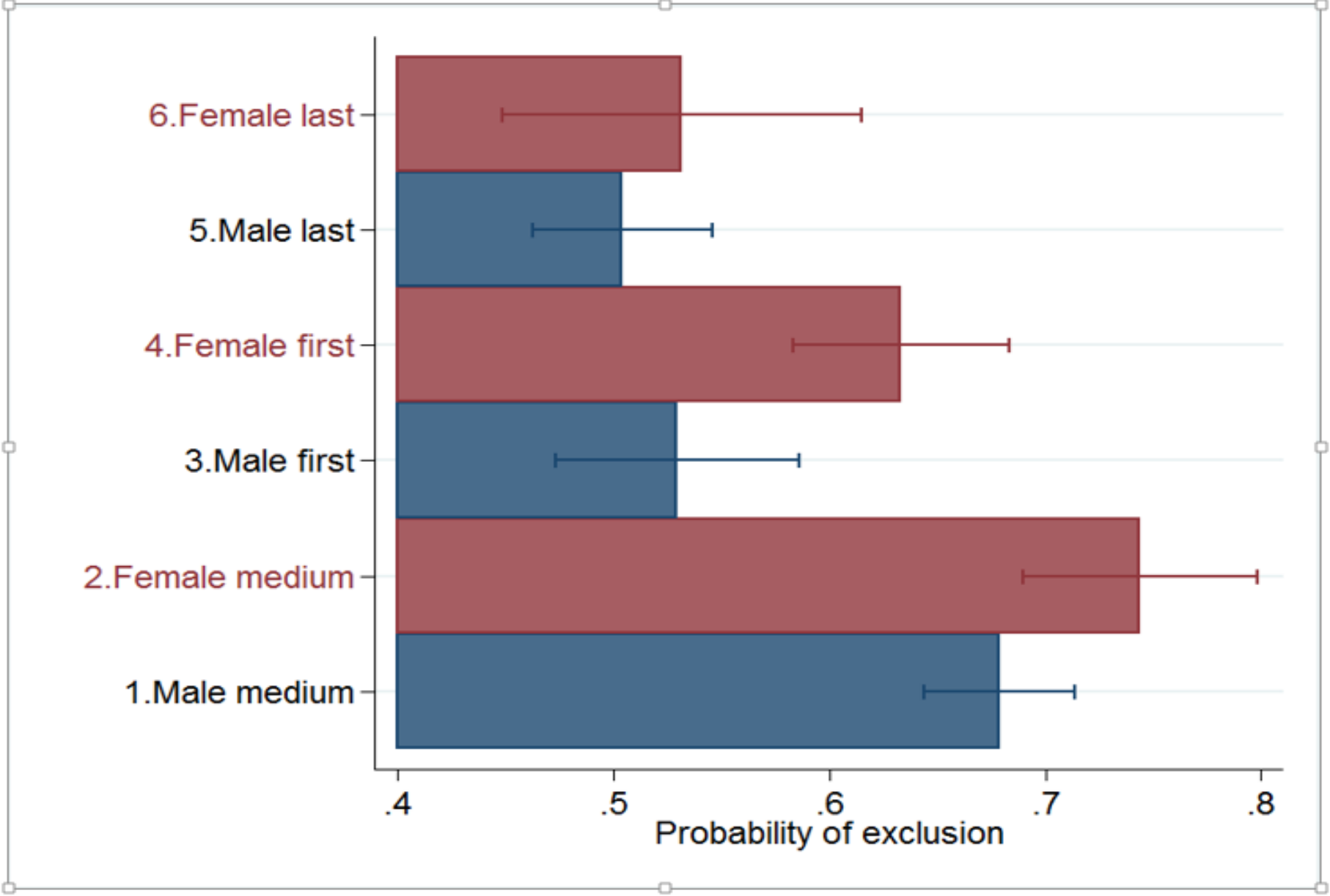
    where $first\ year_{ij}$ = year of first publication by author *i*

Figure B.2 – Predicted probability of exclusion from patents, by author's position in the by-line and relative seniority (scholarship)



As from regression (1), table B.1 (95% confidence intervals)

Figure B.4 – Predicted probability of exclusion from patents, by author's position in the by-line and gender



As from regression (3), table B.1 (95% confidence intervals)

“Renewed co-authorship” regression

OBSERVATIONS: author-PPS dyads

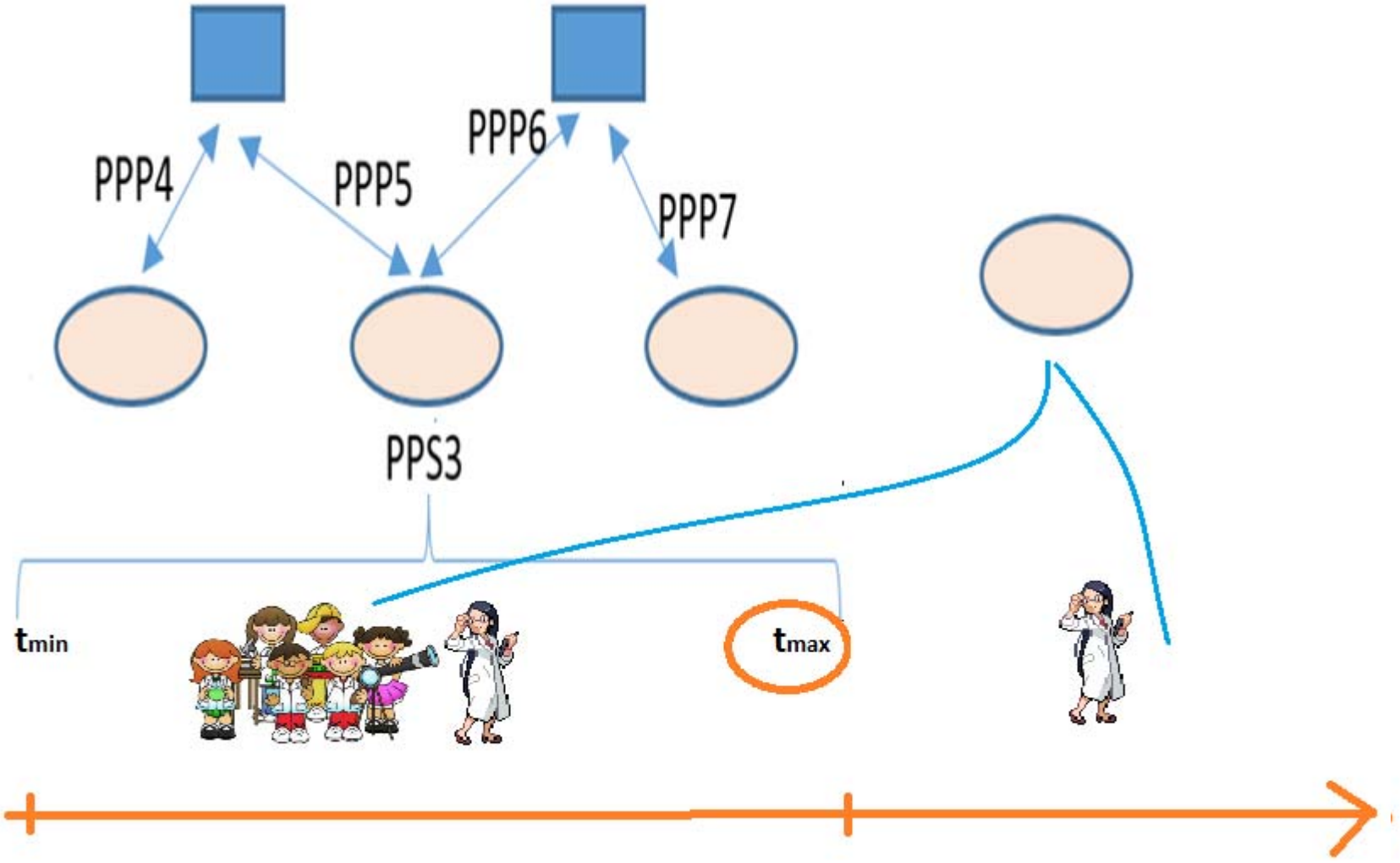
SAMPLE:

- PPSs from the 3p90 class, except:
  - PPPs with number of authors < total number of inventors
  - Publications with alphabetical by-lines
  - Authors with extremely common surnames

  → ~10149 observations

- DEPENDENT VARIABLE *renewed co-authorship* : =1 if the author co-author
  - at least 2 papers
  - with at least 1 PPS-author (whether strictly a co-author or not)
  - after the most recent paper in the PPS

“Renewed co-authorship” regression



“Renewed co-authorship” regression

EXPLANATORY VARIABLES OF INTEREST:

- ✓ Author’s position in the byline:
  - ➢ *First* (at least one paper in PPS as first author
  - ➢ *Last* (at least one paper as last author; none as first)
  - ➢ *Middle* (never get first or last author → ref case)

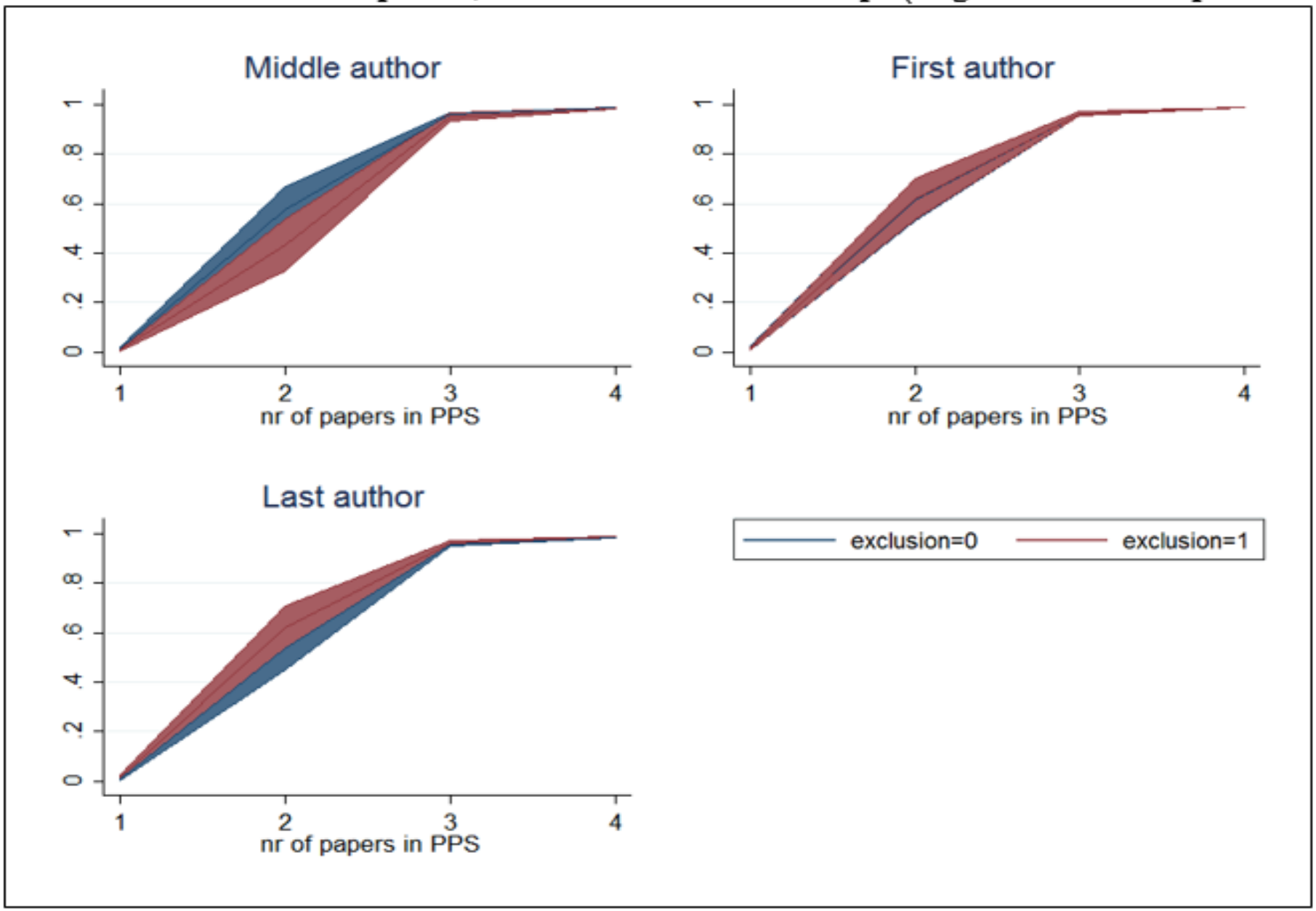
- ✓ *Exclusion :* =1 if author was never inventor on patents in the PPS

- ✓ *Nr of authorships in PPS :* nr of papers each author has contributed to the PPS → seniority/stability of position with the PPS team

CONTROLS:

- ✓ *Last year in PPS*: dummies to control for right truncation
- ✓ Author’s seniority, gender etc as in previous regressions

Figure 5.2.1 – Predicted probability of renewed co-authorship, by author's position, inclusion in patent, and ex-ante co-authorships (Logit estimates - 3p90 PPS)



[1] As from regression (3), table 5.2.1 (95% confidence intervals) / Other variables at means

Policy implications (further research)

1. Authorship & inventorship used by evaluation exercises ARE NOT "objective" measures of contributions to team research

2. Welfare implications:
   - ✓ Negotiation outcome can be optimal for individuals AND necessary for team management, but send biased signals to third parties
   - ✓ Negotiation counters Matthews effect in science, but it boasts it in technology transfer/innovation.

BACK-UP SLIDES

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

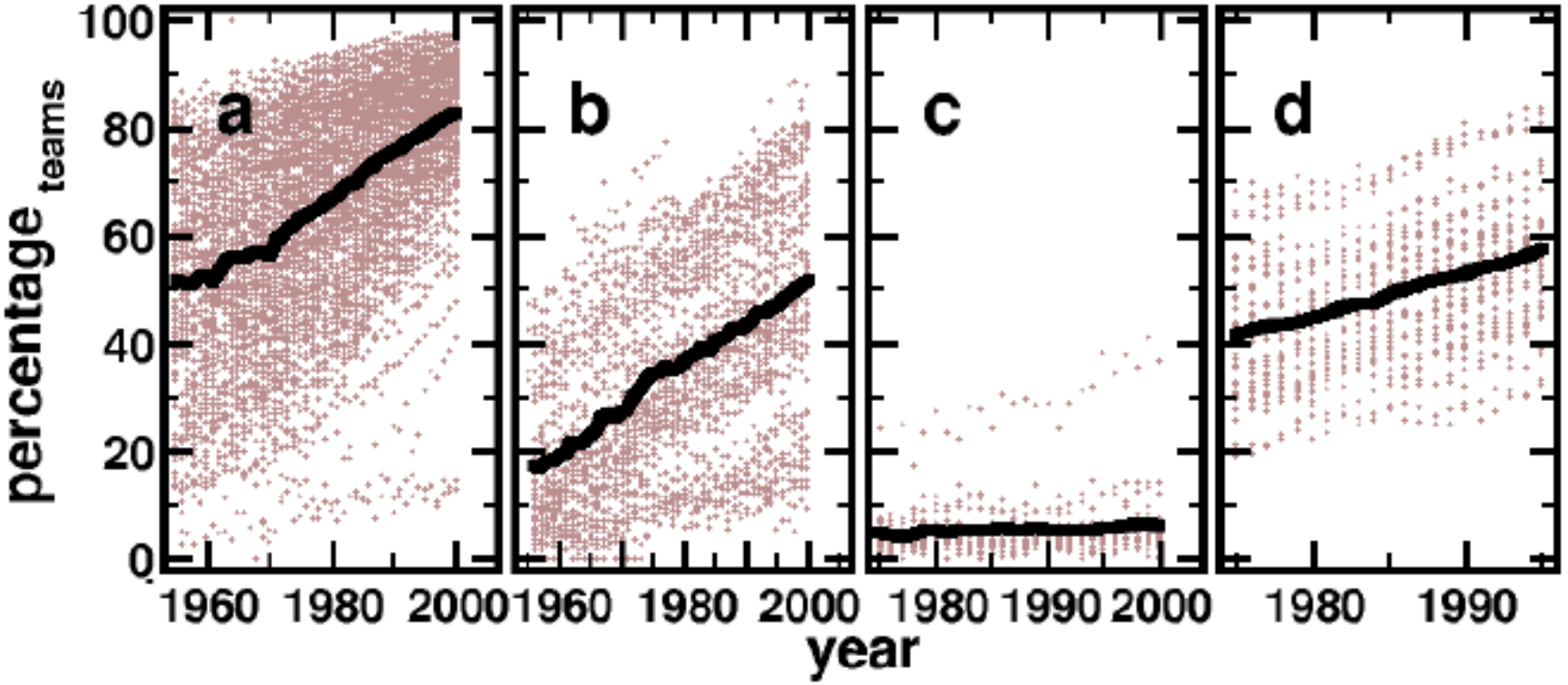
université de BORDEAUX

Motivation: Team science and scientific credit

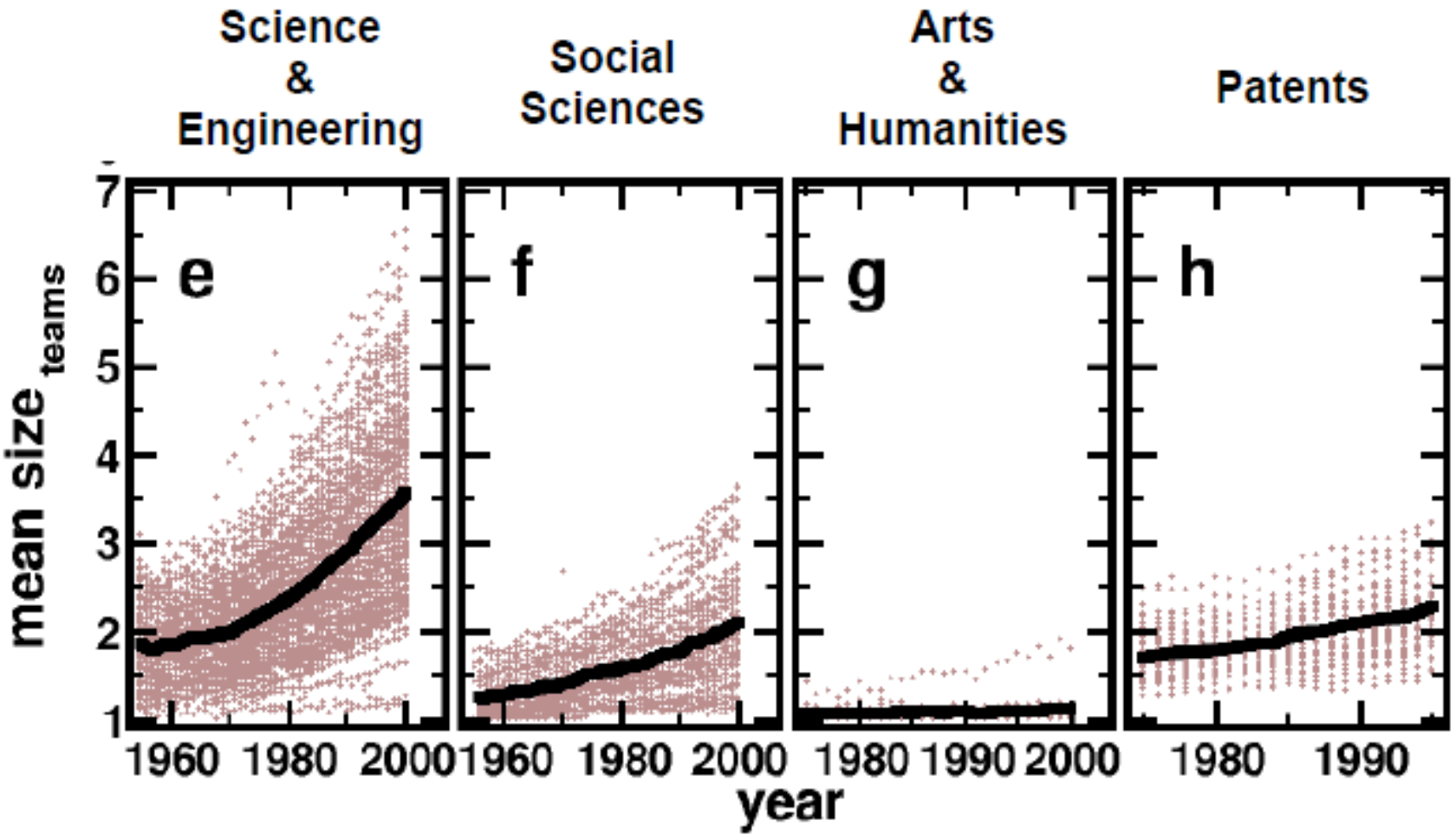
- "Teams in science" as a research topic *(Wuchty et al., 2007; Jones et al., 2008)*
  - Increasing % of multi-authored papers & patents
  - Increasing avg size of co-authorship
  - Increasing geographical dispersion of teams
- Candidate economic determinants:
  - Technology of scientific production (increasing fixed costs)
  - Lower communication costs (*Agrawal et al., 2013)*
  - "Burden of knowledge": longer studies vs. specialization & team work *(Jones, 2009)*

% of team-authored publications and patents, by field (various years) - Source: Wutchy et al. (2007)

Science & Engineering	Social Sciences	Arts & Humanities	Patents
Subfields: 171	54	27	36



Mean size of teams – Source; Wutchy et al., 2007



An example of prequel study *(Gans and Murray, 2013 → adaptation of literature on the economics of patents)*

- 2 researchers: Pioneer (Senior), Follower (Junior)
- 2-step (cumulative) research → objective social value of each stage: x & y
- F has comparative advantage in step-2, but cannot perform step-1
- Individual credit = fractional attribution of social value of research
- P faces chooses from 3 strategies

Available strategies	Costs
Integration (P →1 paper on both steps)	Knowledge acquisition
Collaboration (P recruits F for joint paper on both steps)	Coordination
Publication (P → paper step-1 , F → paper step-2 & cites)	--

- Credit distribution

	P	F
Integration	x+y	0
Collaboration	$\alpha_p(x+y)$	$\alpha_f(x+y)$
Publication	$x+\beta_p y$	$\beta_f y$

Potential inefficiencies for:
$\alpha_p+\alpha_f>1$ or $\beta_p+\beta_f>1$

Patent-publication pairs / example

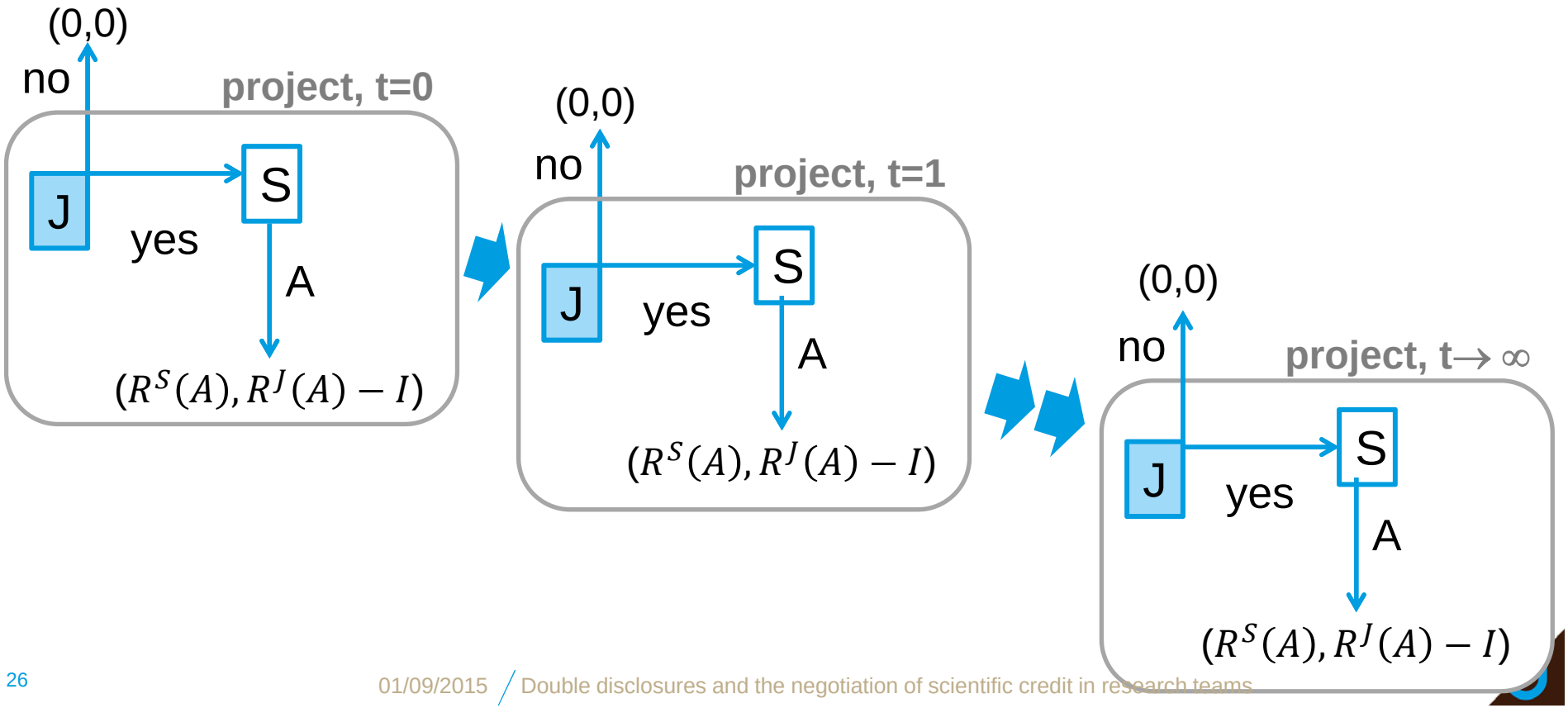
PATENT	PUBLICATION
EP1012301	ISI:000074208600018
Title	**Title**
Total synthesis and functional overexpression of a $i(Candida rugosa) lip1 gene coding for a major industrial lipase	Design. total synthesis and functional overexpression of the Candida rugosa lip1 gene coding for a major industrial lipase
Abstract	**Abstract**
The dimorphic yeast Candida rugosa has an unusual codon usage which hampers the functional expression of genes derived from this yeast in a conventional heterologous host. Lipases produced by this yeast are extensively used in industrial bioconversions, but commercial lipase samples contain several different isoforms encoded by the lip gene family. In a first laborious attempt the lip1 gene. encoding the major isoform of the C. rugosa lipases (crls) was systematically modified [...]	*The dimorphic yeast Candida rugosa has an unusual codon usage that hampers the functional expression of genes derived from this yeast in a conventional heterologous host. Commercial samples of C. rugosa lipase (crl) are widely used in industry. but contain several different isoforms encoded by the lip gene family, among which the isoform encoded by the gene lip1 is the most prominent. In a first laborious attempt. the lip1 gene was systematically modified. [...]*
Inventors	**Authors**
Brocca S., Schmidt-Dannert C., Lotti M., Alberghina L., Schmid R.	Brocca S., Schmidt-Dannert C., Lotti M., Alberghina L., Schmid R.

Theoretical model

- Team composed by a senior scientist (S) and junior one (J)
- Sequence of 1-period projects, each of which:
  - ✓ requires the participation of both S and J
  - ✓ produces one paper and one patent
- S has full control over the allocation of attribution rights
  - ✓ Take-it-or-leave-it offer *A(first authorship, inventorship)*
    e.g. $A \equiv (J, S)$ *: J is first author but is excluded from the patent.*
  - ✓ S cannot commit to any A when the team is formed → final decision may not respect A
- J's participation rules:
  - ✓ per-period cost of participation =*I*
  - ✓ S's final decision cannot be litigated (irrelevance of legal norms) → J's only bargaining power: end-of-period exit !

- Economic value of attribution rights → $R^S(A)$ ($R^J(A)$) is the generic payoff :
  - R_1^S (R_1^J) = value of first authorship for *S (J)*
  - $R_{N1}^S(R_{N1}^J)$ = value of non-first authorship for *S (J)*
  - Value of inventorship: v/n, where $n = 1{,}2$ is the number of inventors listed in the patent

- GAME STRUCTURE (STRATEGIES) & PAYOFFS



«Exclusion» equilibrium

› $A = (J, S)$ is a "sustainable" equilibrium if

$$\delta \geq max \left\{ \frac{I - \mu R_1^J}{I - \mu I}, \frac{R_1^S - R_{N1}^S}{R_1^S + v} \right\}$$

Condition for J Condition for S

➢ *The higher the value of inventorship* → the more S is willing to trade first authorship for it

➢ The higher the value of first authorship for J

➢ The low

➢ The high

➢ The low

NB1 - When $A = (J, S)$ is an equilibrium, also $A = (J, JS)$ is, but S would prefer the former!

NB2 - Other equilibria may exist for small regions of the parameters (e.g. $A = (S, J)$ for small *v.)*

Table A.2 – Patent-publication sets (PPSs) : % distribution of observations, by sample (from less to more restricted) and values of similarity index*

	Similarity index (value intervals)							
	0.2-0.3	*0.3-'0.4*	*0.4-0.5*	*0.5-0.6*	*0.6-0.7*	*0.7-0.8*	*>0.8*	*TOT*
4q3	58.1	24.4	11.0	4.6	1.2	0.5	0.2	100
3p90	19.0	47.6	21.3	8.6	2.3	0.8	0.3	100
2p95	--	42.1	37.3	14.8	3.9	1.4	0.5	100

* similarity index = inverse of cosine distance between title/abstract of each patent and each publication in the PPS

Data & methodology

PPP identification: 2 strategies

- Expert opinion:
  - selected journals → articles → patent DB search → expert validation
- Text analysis:
  - “academic inventors” (APE-INV) → paper DB search → text analysis (contents similarities)
  - transitivity issues: from patent-publication pairs (PPPs) to patent-publication sets (PPSs)

Data & methodology

PATENT DATA: APE-INV project ("Academic patenting in Europe")

- name-based matching of inventors listed on EPO patents and academics listed in various sources (ministerial records, university records, Web of Science records)

Table A.1 Patent sample and methodological reference, by country

	patents		years		reference paper
	nr	%	from	to	
Austria	572	6.4	1997	2007	Stummer et al. (2013)
Belgium	457	5.1	1997	2007	Mejer (2013)
Spain	1820	20.4	1999	2007	Maraut and Martinez (2014)
Italy	3217	36.0	1997	2007	Pezzoni et al. (2014)
Sweden	579	6.5	1997	2007	Ljungberg et al. (2014)
UK	917	10.3	1997	2006	Banal-Estanol et al. (2010)
France	1371	15.3	1997	2007	Cassi (xxx)
Total	8933	100			

Figure A.3 Selected PPPs: % distribution by technological field

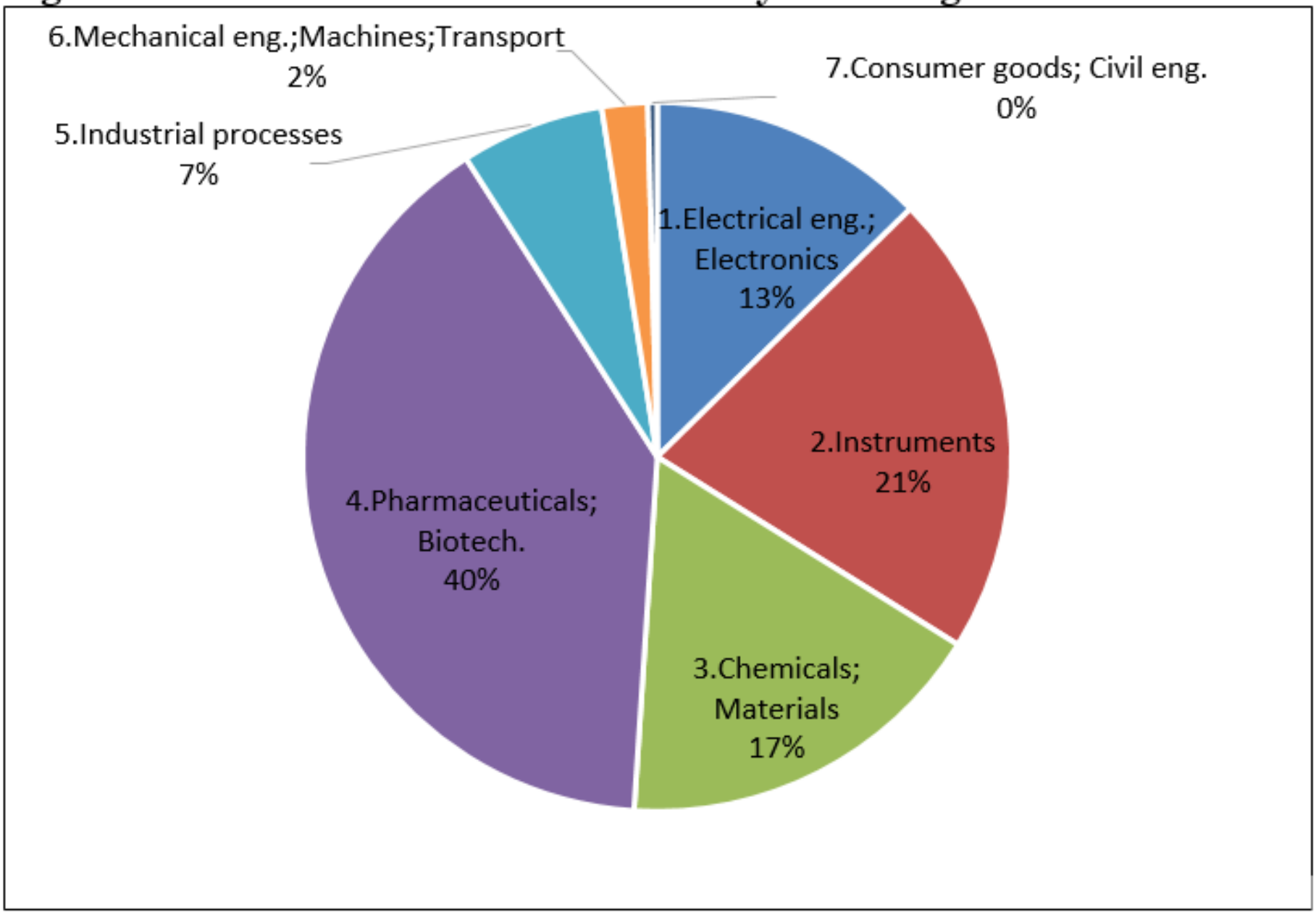
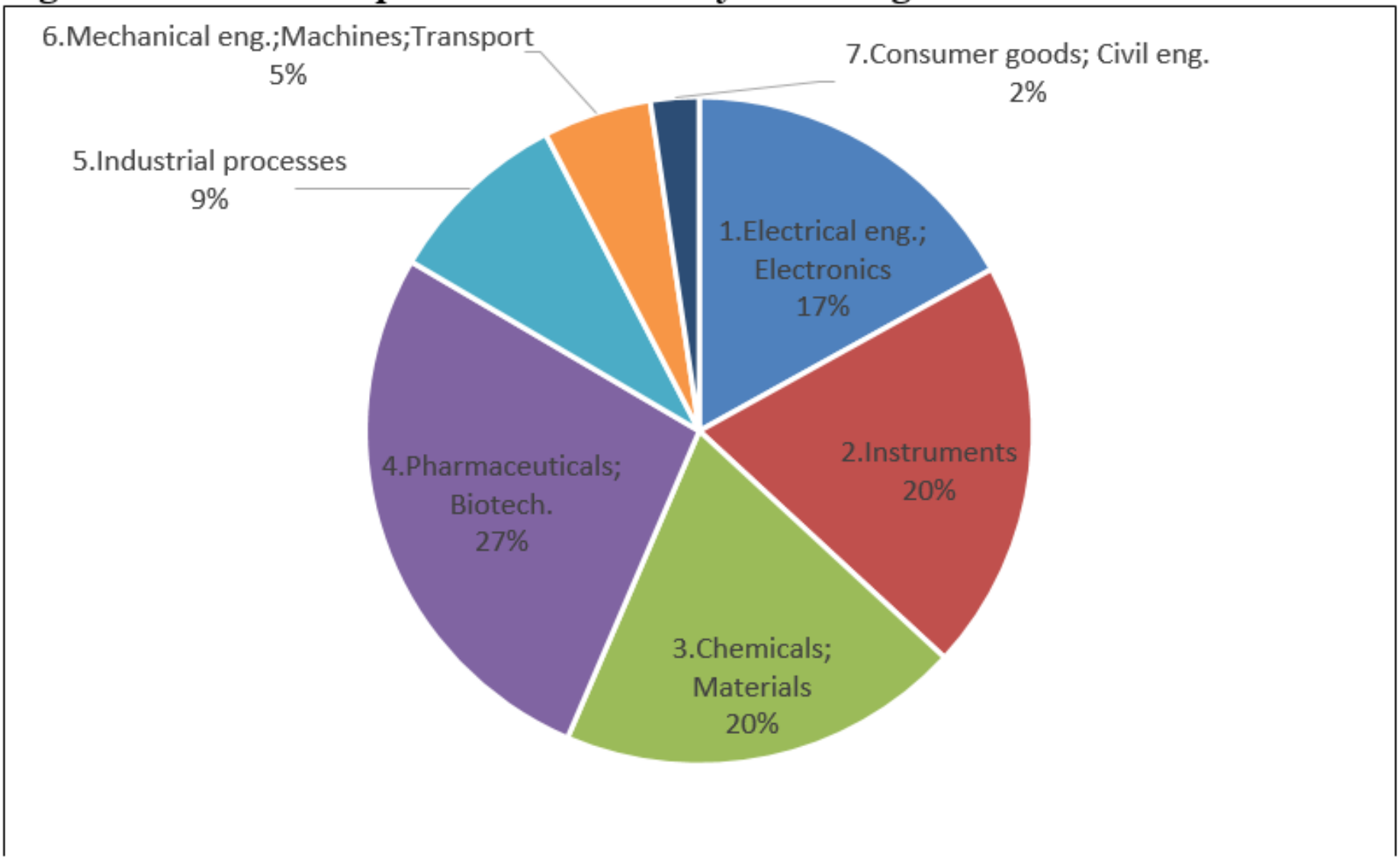


Figure A.2 Patent sample: % distribution by technological field



Note: Observations are patent-in-technological field, with several patents being classified in more than one field (on average, 1.33 fields per patent). Total nr of observations = 12187

Table A.3 Number of patents in the PPS - Frequency distribution

nr of patents in the PPS	*Nr of observations (in PPS class)*		
	pps_2p95	*pps_3p90*	*pps_4q3*
1	483	813	1471
2	62	109	198
3	13	23	56
4		2	15
5	1	2	8
>5	2	3	13
Total PPSs	561	952	1761
Total nr of patents	668	1154	2260

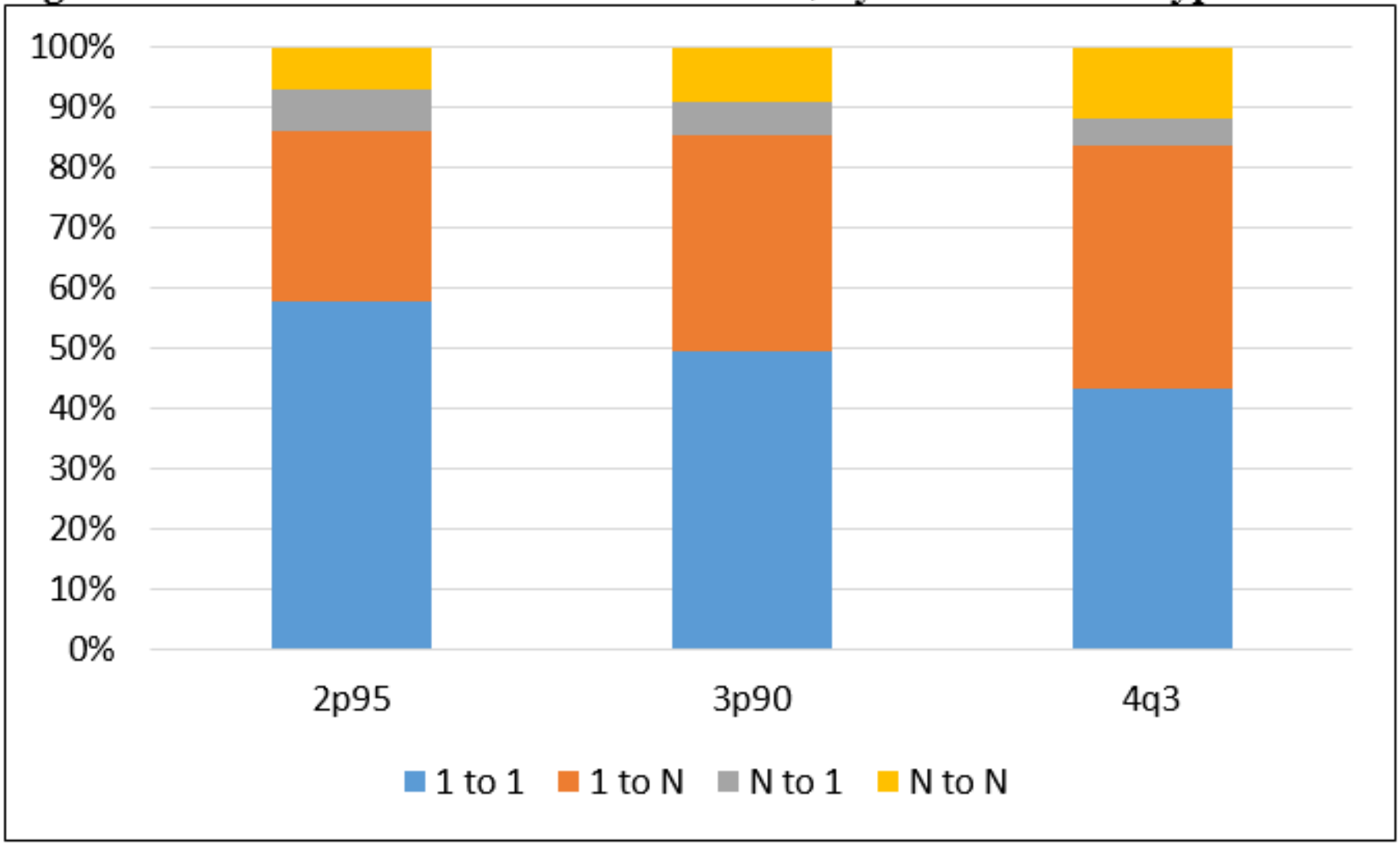
Table A.4 Number of publications in the PPS - Frequency distribution

nr of publications in the PPS	*Nr of observations (in PPS class)*		
	pps_2p95	*pps_3p90*	*pps_4q3*
1	363	524	843
2	96	187	332
3	44	96	205
4	24	48	118
5	10	30	53
6-20	21	62	190
21-100	3	5	19
101-200			1
Total PPS	561	952	1761
Total nr of publications	1142	2299	5570

Table A.6 - % distribution of authors/publication couples, by technology of the patent in the PPP and author's gender (missing observation excluded)

	Electrical eng. & Electronics	Instruments	Chemicals & Materials	Pharma & Biotech.	Other techn.	Total
Male	1,441	3,080	1,271	6,833	385	13,010
%	*85.06*	*74.22*	*68.89*	*62.11*	*84.06*	*67.94*
Female	253	1,070	574	4,169	73	6,139
%	*14.94*	*25.78*	*31.11*	*37.89*	*15.94*	*32.06*
Total	1,694	4,150	1,845	11,002	458	19,149
%	*100*	*100*	*100*	*100*	*100*	*100*

Figure A.4 – % distribution of observations, by PPS class and type



PPS class: 4q3, 3p90, 2p95 (from less to more restrictive values of the similarity indx)
PPS type: 1-to-1, 1-to-N, N-to-1, N-to-N (nr patents –to- nr of publications in the PPS)

Figure A.5 – % distribution of observations, by value of the difference between number of inventors and authors in PPPs and three classes of PPSs

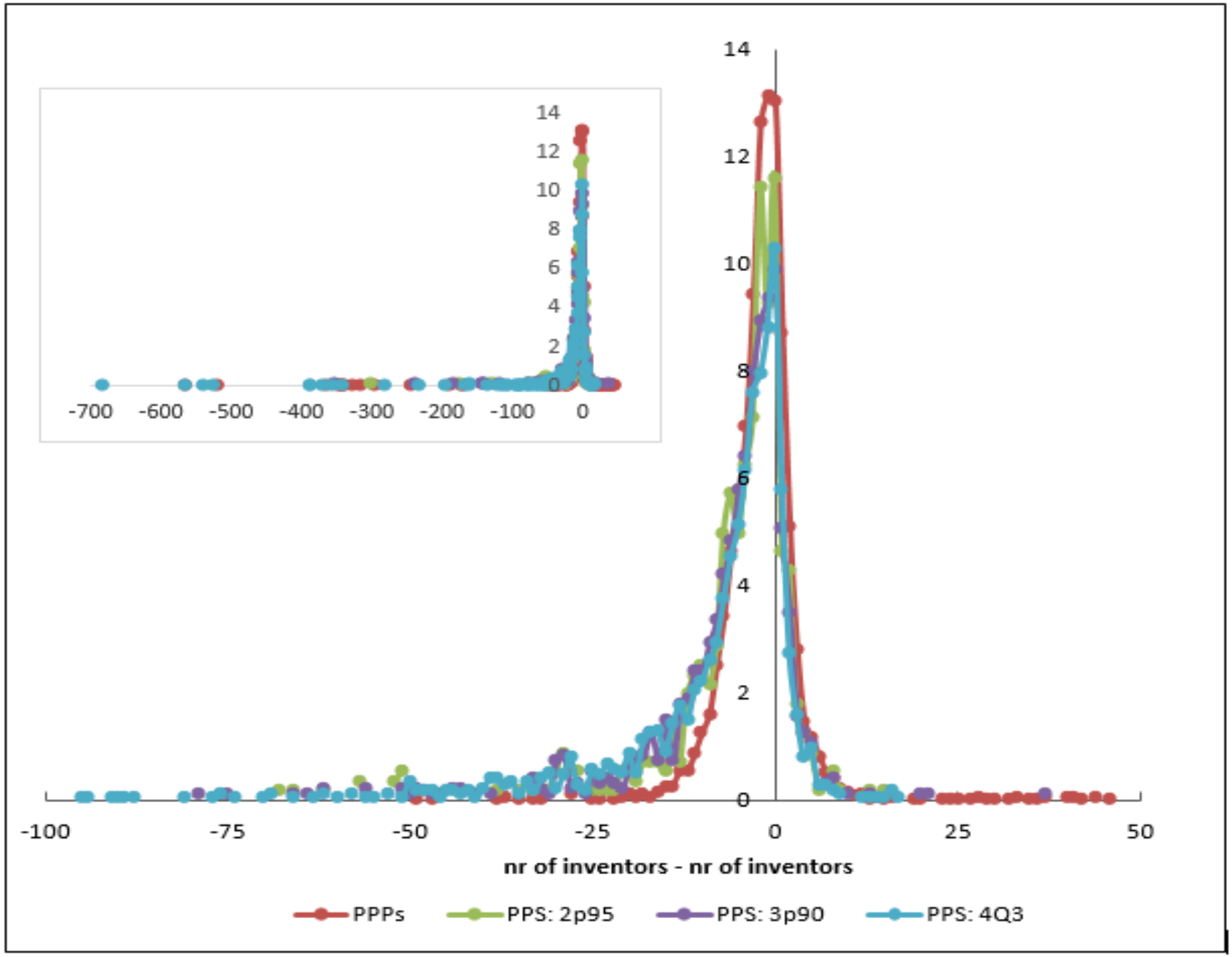
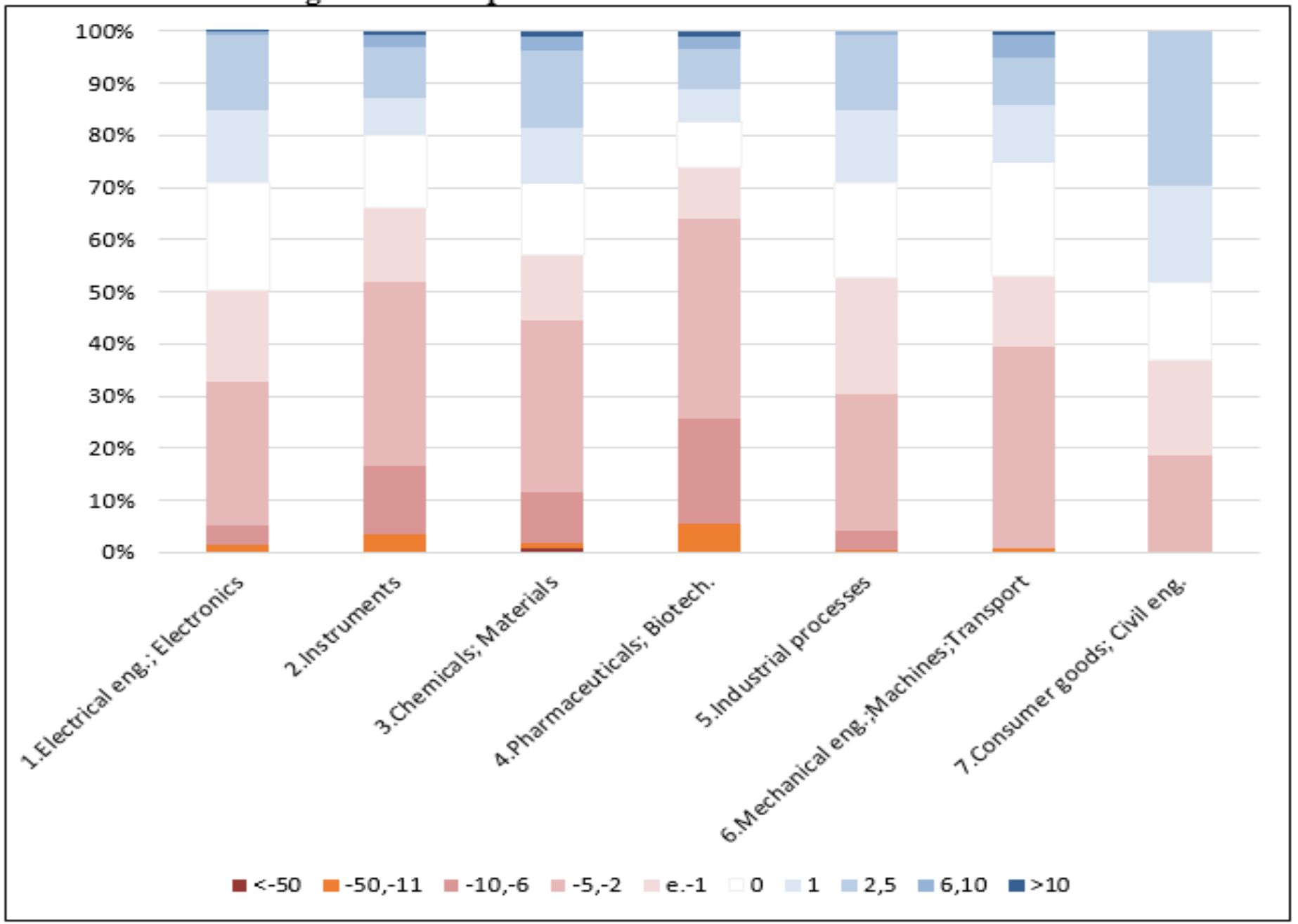


Figure A.6 % frequency distribution of PPPs by difference between number of inventors and authors, and technological fields of patents



EMPIRICAL TEST: 2 REGRESSIONS /1

1. Exclusion regression:

$$Prob(exclusion)$$

$$\boldsymbol{Prob(no\ inventor|author)_{ij}} = \\ = \boldsymbol{f(seniority; gender; byline\ position; controls)}$$

i=individual ; j=patent-related publication

→ Hypotheses:

$$\frac{\partial Prob(exclusion)}{\partial seniority} < 0 \; ; \; \frac{Prob(exclusion|female)}{Prob(exclusion|male)} > 1 \; ;$$

$$\frac{Prob(exclusion|first)}{Prob(exclusion|last)} > 1 \; ;$$

$$\frac{Prob(exclusion|first)}{Prob(exclusion|medium)} \lessgtr ? \; 1 \text{ (from previous research: <1)}$$

“Exclusion” regression

OBSERVATIONS: author-publication dyads

SAMPLE:

- PPSs from the 3p90 class, except:
  - PPPs with number of authors < total number of inventors
  - Publications with alphabetical by-lines
  - Authors with extremely common surnames

  → ~14000 observations (~1000 patents and 1820 publications)

  → Sampling issue 1: >60% observations from Pharma&Biotech

  → Sampling issue 2: >50% observations from Italy

- DEPENDENT VARIABLE *exclusion* : =1 if the author is not an inventor in the PPS

“Exclusion” regression

OBSERVATIONS: author-publication dyads

DEPENDENT VARIABLE *exclusion* : =1 if the author is not an inventor in the PPS

EXPLANATORY VARIABLES OF INTEREST:

- ✓ Author’s position in the byline: *First*, *Last*; and *Middle* (reference case)
- ✓ Author’s gender: *Female* (unavailable for ~30% observations)
- ✓ Author’s seniority:
  - ➢ $relative\ scholarship_{ij} = \frac{stock_{ij} - \min(stock_j)}{\max(stock_j) - \min(stock_j)} = [0,1]$

    where $stock_{ij}$ = nr publications by author *i* at time of publication *j*

    where $first\ year_{ij}$ = year of first publication by author *i*
  - ➢ $relative\ seniority_{ij} = 1 - \frac{first\ year_{ij} - \min(first\ year_j)}{\max(first\ year_j) - \min(first\ year_j)} = [0,1]$

    where *first year*$_{ij}$ = year of first publication by author *i*

“Exclusion” regression

CONTROLS:

- ✓ Text distance between patent and publication in the PPP
  → *Proximity* = 1/*Cos(pat,pub)*, with:

$$Cos(pat, pub) = \sum_i x_{pat,i} x_{pub,i} \Big/ \sqrt{\sum_i x^2{}_{pat,i}} \sqrt{\sum_i x^2{}_{pub,i}}$$

  x_{ij} =1 if word *i* is in document *j=(pat,pub)* and =0 otherwise (max distance if >1 patents or publications in the PPS)

- ✓ Time distance between between patent and publication: $year_{pub}$ – $year_{pat}$ → five dummies from: -2 (ref case) to +2
- ✓ Number of authors in the publication
- ✓ Total number of inventors in the PPP
- ✓ Technology and country dummies (Electronics and Austria as references)

Table 4.2.1.1 – Exclusion regression: summary statistics

	Obs	Mean	Std. Dev.	Min	Max
exclusion	14261	0.68	0.47	0	1
Middle	14261	0.74	0.44	0	1
First	14261	0.13	0.34	0	1
Last	14261	0.13	0.34	0	1
Female	9148	0.35	0.48	0	1
relative_scholarship	14244	0.26	0.35	0	1
relative_seniority	14247	0.38	0.36	0	1
top_scholar	14261	0.13	0.34	0	1
bottom_scholar	14261	0.20	0.40	0	1
most_senior	14261	0.14	0.34	0	1
most_junior	14261	0.22	0.42	0	1
proximity	14261	0.38	0.09	0.28	0.80
-2 years	14261	0.11	0.32	0	1
-1 year	14261	0.13	0.34	0	1
0 years	14261	0.19	0.39	0	1
+1 year	14261	0.28	0.45	0	1
+2 years	14261	0.28	0.45	0	1
n_authors	14261	9.60	5.44	2	55
n_inventors	14261	4.01	2.82	1	40

“Exclusion” regression

Table 4.2.1.1 – Exclusion regression: summary statistics (cont.)

	Obs	Mean	Std. Dev.	Min	Max
Electrical eng; Electronics	14261	0.08	0.27	0	1
Instruments	14261	0.19	0.39	0	1
Chemicals; Materials	14261	0.07	0.26	0	1
Pharmaceuticals; Biotech	14261	0.63	0.48	0	1
Other technologies	14261	0.03	0.16	0	1
Austria	14261	0.08	0.28	0	1
Belgium	14261	0.05	0.23	0	1
Spain	14261	0.14	0.34	0	1
France	14261	0.12	0.32	0	1
Italy	14261	0.55	0.50	0	1
Sweden	14261	0.02	0.15	0	1
UK	14261	0.04	0.18	0	1

Table 5.1.1 – Exclusion regression: OLS estimates (3p90 PPS class)

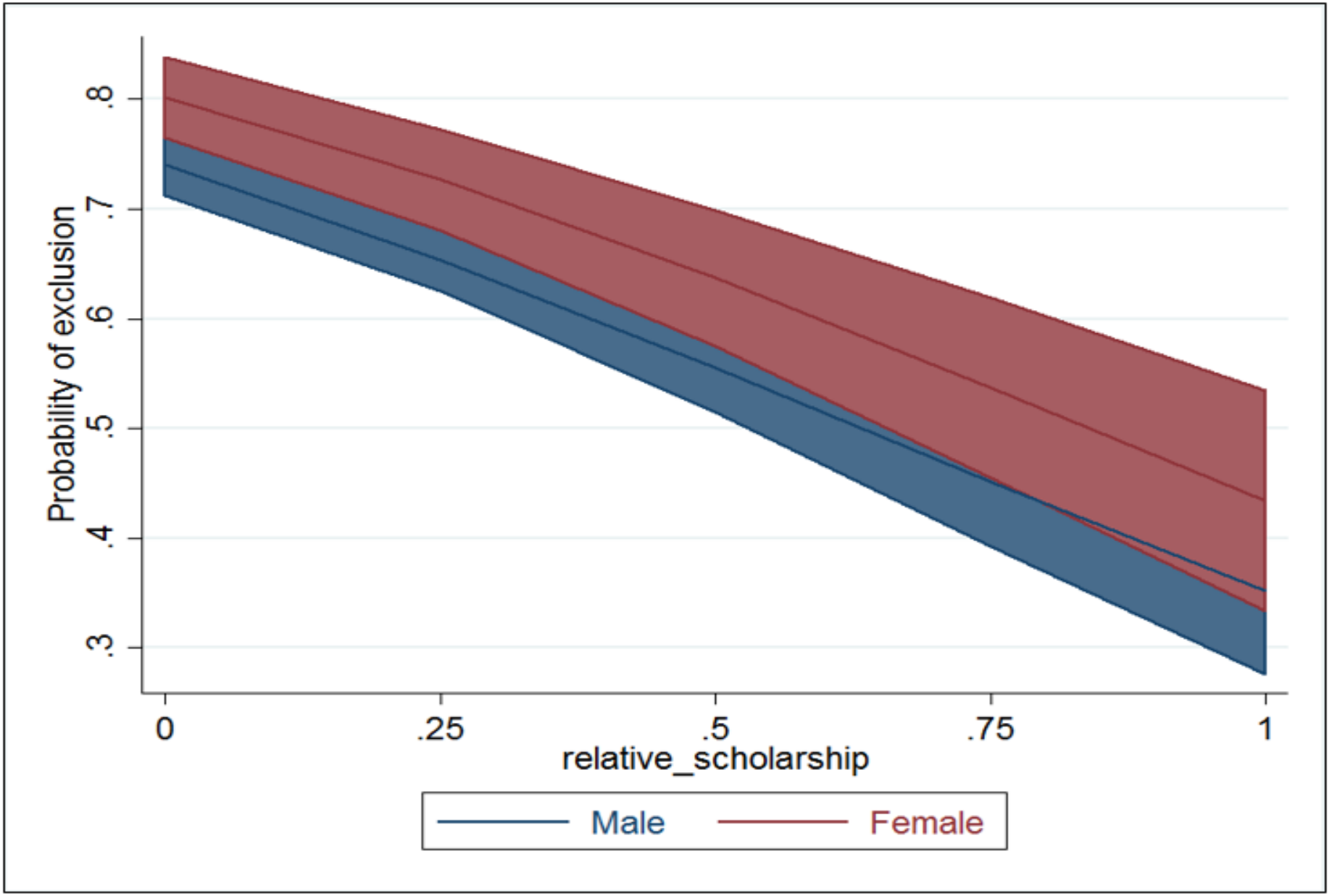
	(1)	(2)	(3)	(4)	(5)
First	-0.14***	-0.13***			
	(0.028)	(0.031)			
Last	-0.21***	-0.19***			
	(0.025)	(0.026)			
Female		0.06***			
		(0.016)			
Female Middle			0.06***	0.04	0.07***
			(0.014)	(0.027)	(0.025)
Male first			-0.16***	-0.17***	-0.21***
			(0.038)	(0.047)	(0.038)
Female first			-0.04	0.01	-0.05
			(0.031)	(0.053)	(0.044)
Male last			-0.19***	-0.14***	-0.23***
			(0.029)	(0.036)	(0.031)
Female last			-0.16***	-0.17***	-0.16**
			(0.041)	(0.059)	(0.061)
relative_scholarship	-0.41***	-0.40***	-0.40***	-0.36***	-0.32***
	(0.036)	(0.043)	(0.043)	(0.053)	(0.039)
most_junior	0.05***	0.05***	0.05***	0.03	0.06***
	(0.014)	(0.017)	(0.017)	(0.026)	(0.022)
proximity	-0.13***	-0.17**	-0.17**	-0.17	-0.11
	(0.038)	(0.078)	(0.078)	(0.138)	(0.119)
-1 year	0.01	0.02	0.02	0.02	-0.04*
	(0.019)	(0.020)	(0.020)	(0.028)	(0.024)
0 years	-0.04**	-0.02	-0.02	0.00	-0.11***
	(0.021)	(0.030)	(0.030)	(0.037)	(0.028)
+1 year	-0.05*	-0.03	-0.03	-0.03	-0.10***
	(0.028)	(0.031)	(0.031)	(0.037)	(0.023)
+2 years	-0.06*	-0.06**	-0.06**	-0.08**	-0.14***
	(0.030)	(0.031)	(0.031)	(0.039)	(0.027)

Table 5.1.1 – Exclusion regression: OLS estimates (3p90 PPS class) / cont.

n_authors	0.01***	0.01***	0.01***	0.02***	0.02***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
n_inventors	-0.03***	-0.03***	-0.03***	-0.04***	-0.03***
	(0.004)	(0.005)	(0.005)	(0.010)	(0.004)
Instruments	0.09**	0.15***	0.15***	0.13***	0.17**
	(0.041)	(0.055)	(0.054)	(0.049)	(0.071)
Chemicals; Materials	0.12***	0.20***	0.20***	0.21***	0.22***
	(0.044)	(0.056)	(0.056)	(0.052)	(0.072)
Pharmaceuticals; Biotech	0.13***	0.20***	0.20***		0.17**
	(0.040)	(0.054)	(0.054)		(0.071)
Other technologies	0.07	0.13*	0.13*	0.13**	0.05
	(0.054)	(0.070)	(0.070)	(0.067)	(0.086)
Belgium	-0.04	-0.00	-0.00	-0.03	0.02
	(0.023)	(0.034)	(0.034)	(0.066)	(0.034)
Spain	-0.07***	-0.01	-0.01	-0.09	-0.00
	(0.025)	(0.033)	(0.033)	(0.057)	(0.031)
France	-0.04*	-0.02	-0.02	-0.03	-0.02
	(0.020)	(0.027)	(0.027)	(0.045)	(0.026)
Italy	-0.03***	-0.05*	-0.05	-0.12***	
	(0.025)	(0.030)	(0.030)	(0.036)	
Sweden	0.07*	0.10*	0.10*	0.06	0.12***
	(0.042)	(0.051)	(0.051)	(0.075)	(0.043)
UK	-0.06	-0.12**	-0.12**	-0.18***	-0.11*
	(0.044)	(0.056)	(0.056)	(0.057)	(0.057)
Constant	0.89***	0.73***	0.73***	0.77***	0.73***
	(0.052)	(0.071)	(0.071)	(0.094)	(0.094)
Observations	14,244	9,141	9,141	2,943	3,872
R-squared	0.216	0.233	0.239	0.223	0.246
F-test	106.4	70.74	67.35	30.99	41.39

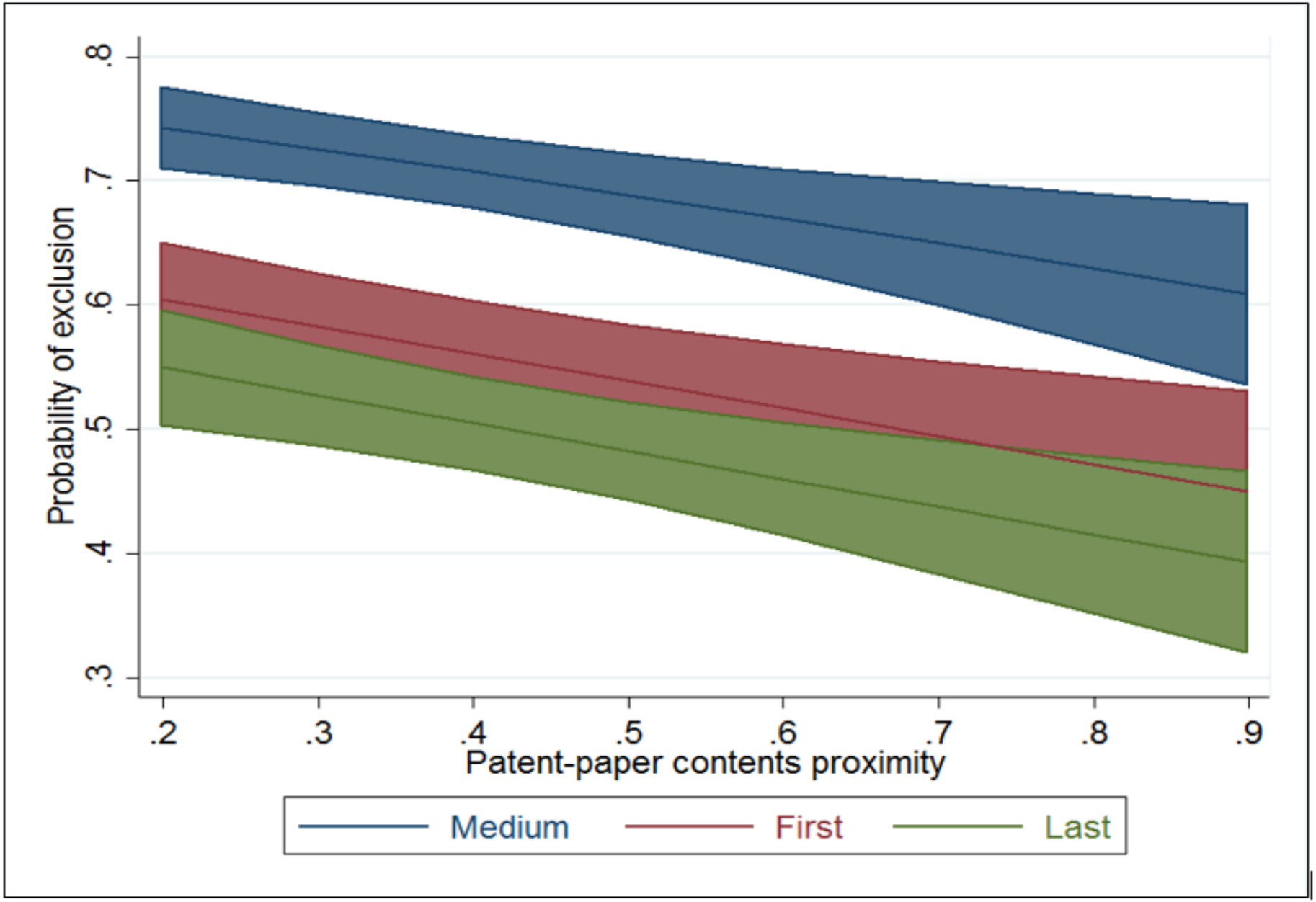
Robust standard errors in parentheses ; *** p<0.01, ** p<0.05, * p<0.1

Figure B.3 – Predicted probability of exclusion from patents, by author's gender and relative seniority (scholarship)



As from regression (2), table B.1 (95% confidence intervals)

Figure B.1 – Predicted probability of exclusion from patents, by author's position in the by-line and contents proximity of paper and patent



As from regression (1), table B.1 (95% confidence intervals)

Country and technology effects

1. Pure country effects:

	(1)	(2)	(3)	(4)
(Austria)	…	…	…	…
Belgium	-0.09**	-0.08	-0.08	-0.06
Spain	-0.10***	-0.05	-0.05	-0.03
France	-0.08***	-0.06	-0.06	0.03
Italy	-0.13***	-0.11**	-0.11**	-0.11**
Sweden	0.03	0.05	0.05	0.16***
UK	-0.1	-0.16**	-0.16**	-0.12*

2. Gender-country interaction: NO significant results (BUT country coefficients affected by gender controls)

3. Gender-technology interaction: higher probability of exclusion in Pharma&Biotech (also Pharma&Biotech+Instruments) BUT low presence of women in other fields (→ limit the significance of estimates)

4. Seniority interactions (with countries or technologies): NO significant results

Table 5.1.1 – Exclusion regression: OLS estimates (3p90 PPS class)

	(1)	(2)	(3)	(4)	(5)
First	-0.14***	-0.13***			
Last	-0.21***	-0.19***			
Female		0.06***			
Female Middle			*0.06****	*0.04*	*0.07****
Male first			*-0.16****	*-0.17****	*-0.21****
Female first			*-0.04*	*0.01*	*-0.05*
Male last			*-0.19****	*-0.14****	*-0.23****
Female last			*-0.16****	*-0.17****	*-0.16***
relative_scholarship	-0.41***	-0.40***	-0.40***	-0.36***	-0.32***
most_junior	0.05***	0.05***	0.05***	0.03	0.06***
Technology f.e.	Y	Y	Y	Y (4)	Y
Country f.e.	Y	Y	Y	Y	Y (5)
Other controls	Y	Y	Y	Y	Y
Constant	0.89***	0.73***	0.73***	0.77***	0.73***
Observations	14,244	9,141	9,141	2,943	3,872
R-squared	0.216	0.238	0.239	0.228	0.246
F-test	106.4	70.74	67.35	30.99	41.39

(4) no Pharma & Biotech ; (5) no Italy

Robust standard errors in parentheses ; *** p<0.01, ** p<0.05, * p<0.1

EMPIRICAL TEST: 2 REGRESSIONS /2

2. Renewed co-authorship regression :

$$\boldsymbol{Prob(new\ co-authorship)_{ij} =}$$
$$\boldsymbol{= f(exclusion; byline\ position; ; controls)}$$

i=*individual* ; j=*patent-related publication*

→ Hypotheses:

$$\frac{Prob(new\ co-authorship|first)}{Prob(new\ co-authorship|middle)} > 1\ ;$$

$$\frac{Prob(new\ co-authorship|first, exclusion)}{Prob(new\ co-authorship|first, non-exclusion)} = 1\ ;$$

$$\frac{Prob(new\ co-authorship|middle, exclusion)}{Prob(new\ co-authorship|middle, non-exclusion)} < 1\ ;$$

“Renewed co-authorship” regression

Table 4.3.1– Nr of renewed co-authorships per author, and % of first/last authorship – frequency distribution

Nr of co-authorships	Nr of authors	% of authors
none	9	0.09
1	8,722	85.94
>1	1,418	13.97
Total	*10,149*	*100*

“Renewed co-authorship” regression

Table 4.3.3 – Renewed co-authorship regression: summary statistics

	Obs	Mean	Std. Dev.	Min	Max
Position:					
Medium	10149	0.67	0.471	0	1
First	10149	0.17	0.374	0	1
Last	10149	0.16	0.370	0	1
Exclusion	10149	0.69	0.463	0	1
Nr authorships in PPS	10149	1.43	1.107	1	25
Female	5886	0.32	0.467	0	1
First publication year	10021	1992.81	11.986	1950	2014
Publication stock	10021	26.59	57.457	0	1318
Proximity	10149	0.42	0.102	0.28	0.93
N_authors in PPS	10149	20.50	17.455	2	80
N_inventors in PPS	10133	4.68	3.352	1	17
Last year in PPS	10149	2004.87	2.454	1999	2009

"Renewed co-authorship" regression

Table 5.2.1 – Renewed co-authorship regressions: OLS (3p90 PPS)

	(1)	(2)	(3)
First	0.07***	0.02	0.01
Last	0.11***	-0.01	-0.01
Exclusion			
First*Exclusion			
Last*Exclusion			
Female			
Nr authorships in PPS			
Nr authorships in PPS (sq)			
First publication year			
Publication stock			
Proximity			
N_authors in PPS			
N_inventors			
Last year in			
Constant			
Observations			
R2/Pseudo-R2	0.077	0.499	0.515
F-test	21.67	63.46	53.71

Columns (1) to (3): OLS ; Columns (4) to (6) Logit (estimated coefficients reported)

Robust standard errors in parentheses ; *** p<0.01, ** p<0.05, * p<0.1

Regression (2):

$$H0: \beta_{Exclusion} + \beta_{First*Exclusion} = 0$$

→ cannot reject: First is indifferent to exclusion

Regression (2):

$$H0: \beta_{Exclusion} + \beta_{Last*Exclusion} = 0$$

lusion!!!

Also problematic: wrong sign!

BUT it disappears with robustness checks (2p95 sample)

Discussion

1. Credit attribution as outcome of complex negotiations

   → *social norms that matter are not so much abstract ones (fractional attribution etc) but those that affect the "returns on credit"*

   → *"returns" on credit depend on the individual characteristics: seniority and gender*

✓ <u>Empirical results compatible with theory</u>

2. "Exclusion" regression:

   ➔ *Large marginal effect of seniority : +40% probability*

   ➔ *Smaller (but significant) gender effect: +6%*

   ➔ *Previous results, obtained for Italy, only hold for other countries*

3. "Renewed co-authorship" regression: results are compatible with theory

   ➔ *Exclusion does not affect team fidelity of first authors*

   ➔ *BUT it affects middle authors (those who claimed first authorship or inventorship and got none)*