

# Testing technology- industry concordances using linked micro-level data on patents and firms

Qifeng Weng, Zachary Brown  
North Carolina State University

# Better Technology-Industry Concordances

**Objective:** To know the likelihood that patented inventions with a given technology classification are associated with a given industry.

**Motivation:** To study industry-level economic dynamics of innovation, e.g. patenting and global value chains (Zolas and Lybbert, in prep.)

## Prevailing Concordances:

1. Algorithmic Links w/ Probabilities (ALP), Lybbert and Zolas (2014, Res. Policy)
2. DG concordance by Schmoch et al., 2003
3. Yale Technology Concordance (YTC) by Kortum and Putnam, 1997

## Problems with existing approaches:

- Labor intensive (combing through patent abstracts and applications)
- Not easy to automatically update
- Highly subjective

# Key Elements in Our Research

## We:

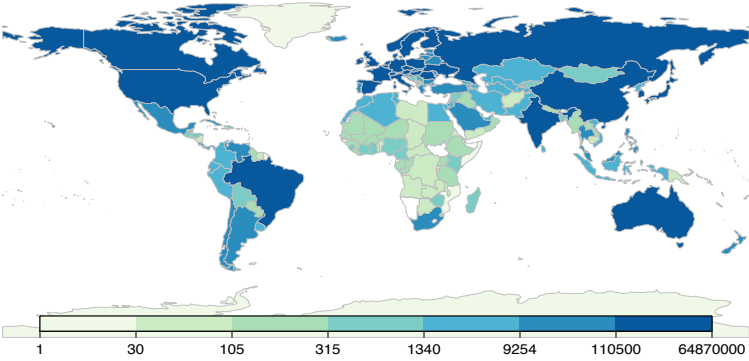
- Link patents (with technology classifiers) to industries by connecting patent applicants to the firm-level database (with associated industry classifiers).
  - Eliminates a significant amount of subjectivity.
- Propose a concordance using two large databases: PATSTAT and OECD-Orbis.

## Two possible approaches:

- Direct calculation: ratio between PATSTAT-Orbis matching patent family count with IPC4 “i” and linked to industry “j” and the raw PATSTAT patent family count with IPC4 “i”.
- Machine learning: Regress industry classifiers on large set of IPC4 covariates (there are 643 possible IPC4 covariates). Use machine learning (Bayesian model averaging) to identify a subset of these covariates most likely to be associated with a given industry.

# Only a fraction of PATSTAT patents can be matched to firms in Orbis: sample still representative?

Figure 3: PATSTAT Patent Count Global Distribution



*Raw*

Figure 1: Raw PATSTAT Patent IPC4 Distribution

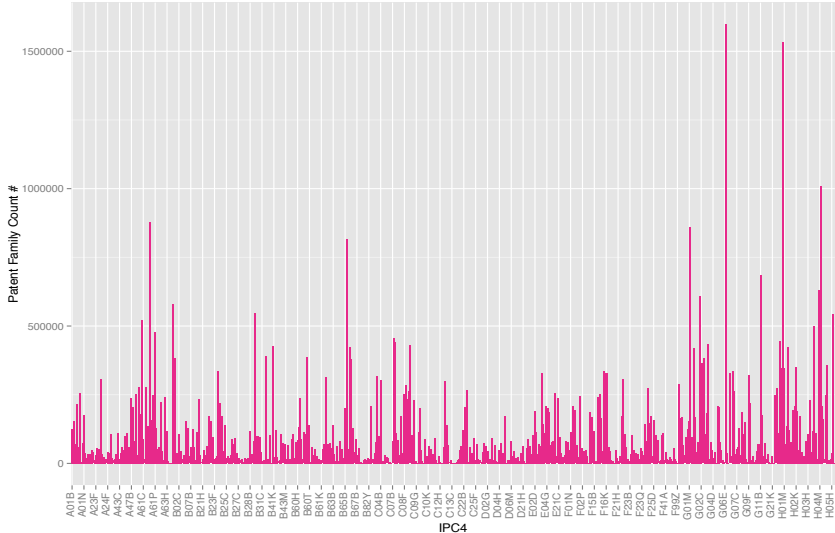
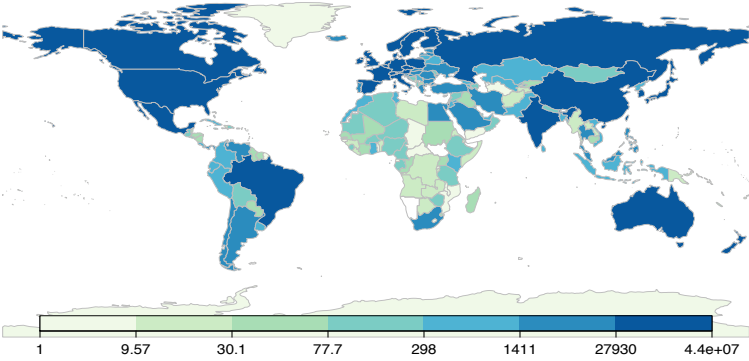
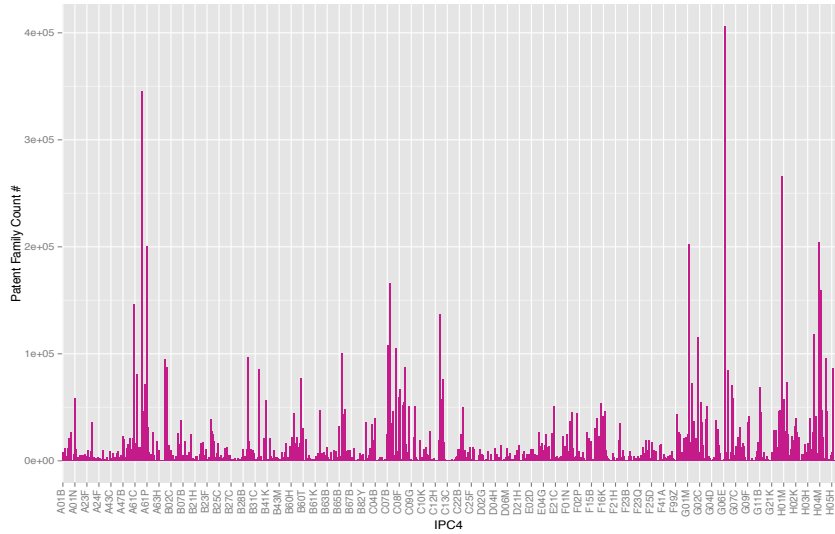


Figure 4: PATSTAT-Orbis Patent Count Global Distribution



*Orbis matched*

Figure 2: PATSTAT-Orbis Match IPC4 Distribution



# Results: ALP evaluation & new concordances

**ALP Concordance Significance Test**

3314	3315	3319	3320	3510	3520	3530	3600	3811	3812	3821	3822	3900	4100
3091	3092	3099	3100	3211	3212	3220	3230	3240	3250	3290	3311	3312	3313
2822	2823	2824	2825	2826	2829	2910	2920	2930	3011	3012	3020	3030	3040
2732	2740	2750	2790	2811	2812	2813	2814	2815	2816	2817	2818	2819	2821
2593	2599	2610	2620	2630	2640	2651	2652	2660	2670	2680	2710	2720	2731
2394	2395	2396	2399	2410	2420	2431	2432	2511	2512	2513	2520	2591	2592
2012	2013	2021	2022	2023	2029	2030	2100	2219	2220	2310	2391	2392	2393
1512	1520	1610	1621	1622	1623	1701	1702	1709	1811	1812	1820	1920	2011
1104	1200	1311	1312	1313	1391	1392	1393	1394	1399	1410	1420	1430	1511
1030	1040	1050	1061	1062	1071	1072	1073	1074	1075	1080	1101	1102	1103
710	721	729	810	891	892	893	899	910	990	1010	1020	1030	1030
163	164	170	210	220	240	311	312	321	322	510	520	610	620
127	128	129	130	141	142	143	144	145	146	149	150	161	162
111	112	113	114	115	119	121	122	123	124	125	126	126	126

Note: 1. Significance level is 0.05 2. Number is each cell indicates the ISIC Code

color Significant Non-Significant

**Correlation b/w ROC and OBC**

7219	7220	7490	8299	8541	8690	9499
6820	6910	7010	7022	7112	7210	7211
4773	4778	5829	6190	6201	6209	6420
4299	4321	4519	4614	4643	4690	4730
2910	2932	3250	3299	3319	4110	4120
2670	2751	2790	2824	2829	2892	2899
2229	2599	2611	2620	2630	2651	2660
1520	2014	2016	2041	2059	2110	2120

Correlation  
0.75  
0.50  
0.25  
0.00

Note: Number is each cell indicates the ISIC Code