Ongoing Research Projects on Patent Data

Research Division Innovation Economics – KOF Swiss Economic Institute, ETH Zurich

Florian Seliger

Summer School on Data & Algorithms for Science, Technology and Innovation Studies, September 2018

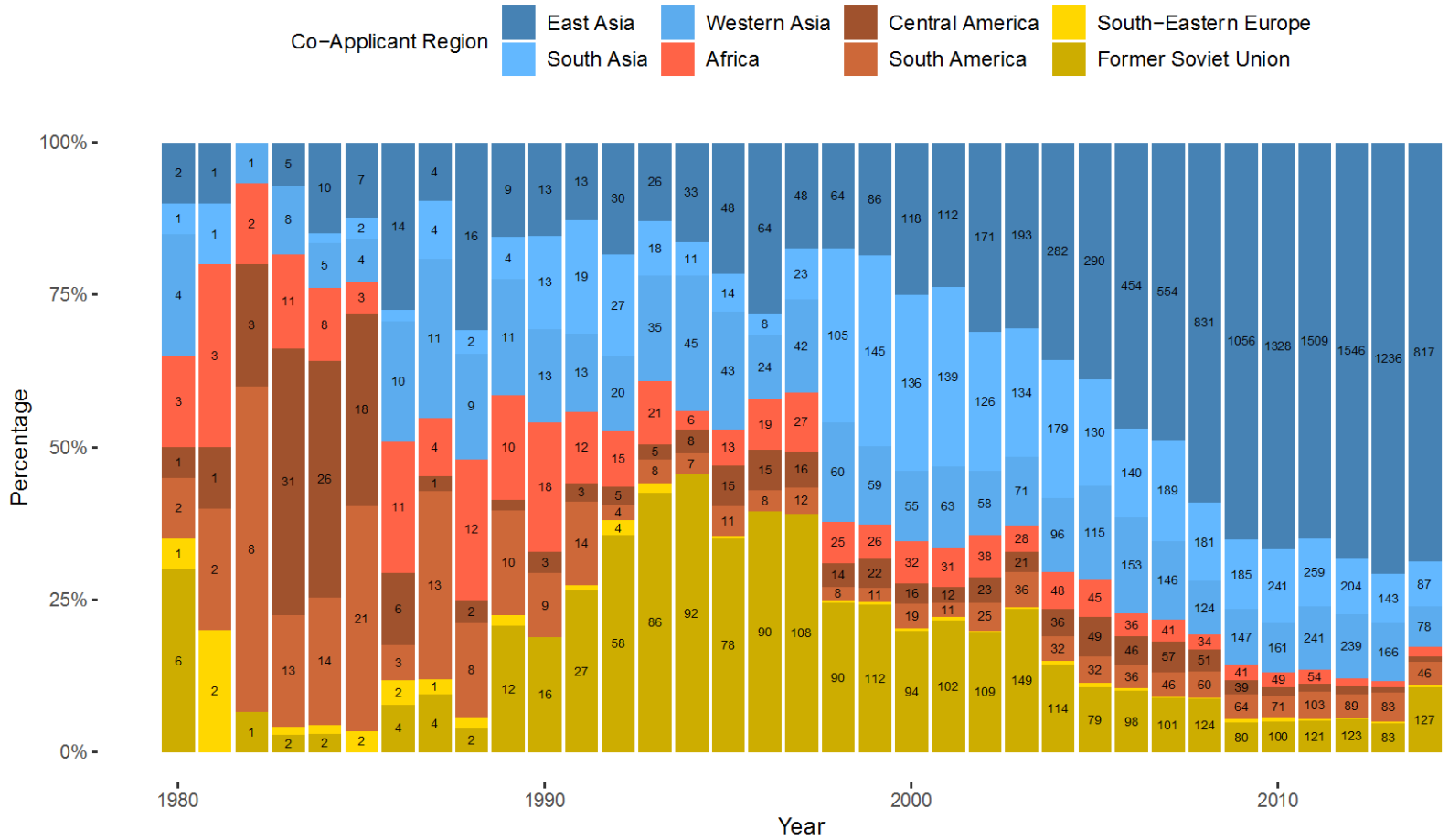
The Globalisation of R&D

Evidence on Technology Clusters

In collaboration with Gaétan de Rassenfosse and Jan Kozak (EPFL Lausanne),
funding from the Swiss National Science Foundation is acknowledged

Patents Co-Applied Between Core Europe and Developing Countries, by Region in (%)

Ciaramella, De Rassenfosse, Seliger 2018



Introduction

- Globalisation of R&D defining feature of innovation systems
- Cooperation and R&D offshoring across the globe in order to access pool of knowledge (talents, technologies) and to benefit from knowledge spillovers in technological clusters
- Although literature on “location choices” is broad, lack of evidence on **technological clusters**, their identification and evolution

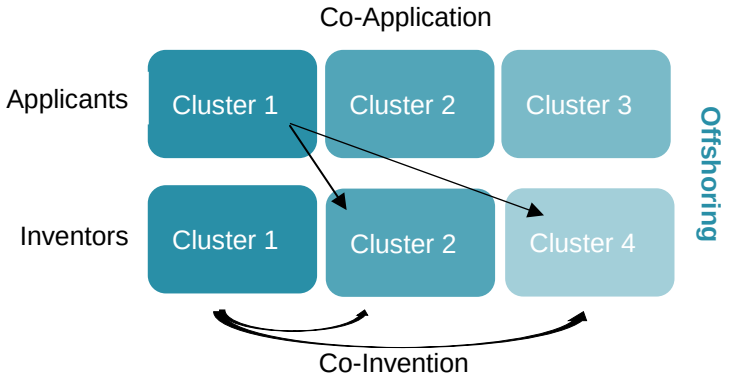
Introduction

Studies using patent data to trace internationalisation of R&D

- **Technology sourcing** – Patent applications with inventors abroad (R&D offshoring, market-seeking, technology-seeking) (Chung & Alcácer, 2002; Dunning & Lundan, 2008)
- **R&D collaboration** – Co-inventions and co-applications with partners abroad (Guellec & van Pottelsberghe de la Potterie 2001, Picci 2010, Picci & Savorelli 2018)
- **Knowledge flows** – Patent citations (Maruseth & Verspagen 2002, Peri 2005)
- **Knowledge transfer** – Patent transfers (De Marco et al. 2017, Bösenberg & Egger 2017), inventor movements (Oettl & Agrawal 2008, Singh & Agrawal 2011)

Introduction

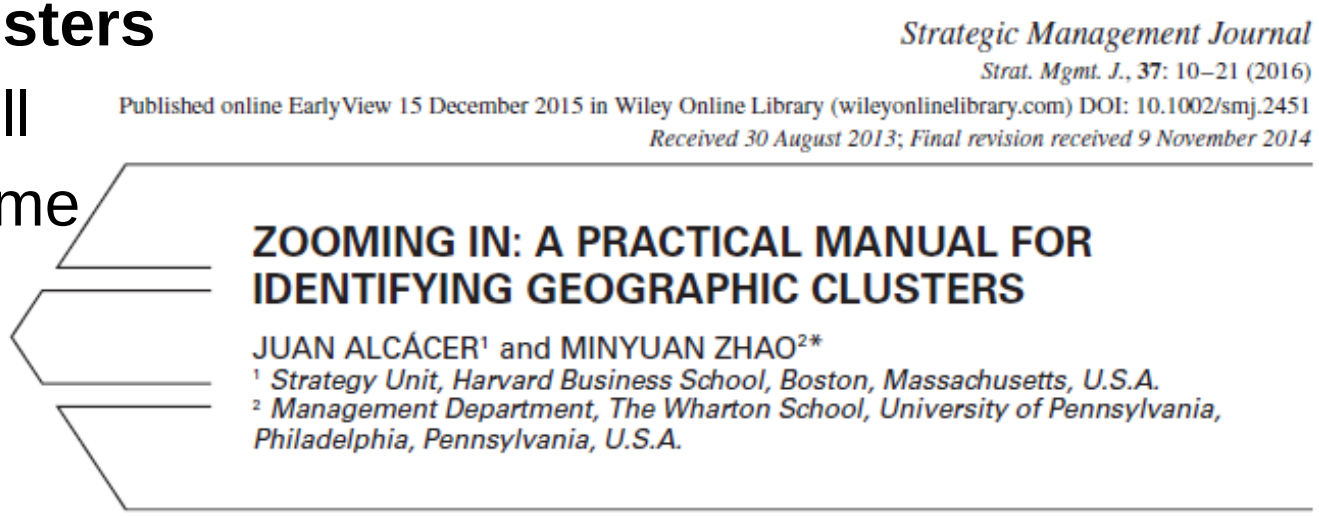
- Technology sourcing from one cluster to the other
- R&D collaboration between clusters
- Knowledge flows between clusters



Current state of research

Measurement of technology clusters

- Contributions in our field are still scarce and incomplete (data, time span, geographical and technological space)
- REGPAT etc. offer data by administrative regions,
  not very accurate, clusters defined by borders not technologies



Strategic Management Journal
Strat. Mgmt. J., **37**: 10–21 (2016)
Published online EarlyView 15 December 2015 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2451
Received 30 August 2013; Final revision received 9 November 2014

ZOOMING IN: A PRACTICAL MANUAL FOR IDENTIFYING GEOGRAPHIC CLUSTERS

JUAN ALCÁCER[1] and MINYUAN ZHAO[2]*
[1] *Strategy Unit, Harvard Business School, Boston, Massachusetts, U.S.A.*
[2] *Management Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.*

Research summary: *This paper advances strategic management research by taking a close look at the reasons, procedures, and results of cluster identification methods, focusing on a density-based algorithm that organically define clusters from actual locations of economic activities. Despite being a popular research topic and analytical tool, geographic clusters are often studied with little consideration given to the underlying economic activities, the unique cluster boundaries, or the appropriate benchmark of economic concentration. Our goal is to increase awareness of the complexities behind cluster identification, and to provide concrete insights and methodologies applicable to various empirical settings. The method we propose is especially useful when researchers work in global settings, where data available at different geographic units complicates comparisons across countries.*

Goals of the project

- Worldwide count of priority filings by inventor and applicant cluster and by industry/technology
- Dataset with geocoordinates and cluster assignment available for research community
- Observe development of technology clusters worldwide based on technology density
- Overcome limitations in other studies by including
  - More data: More countries and patent offices
  - A larger time span

Project stages

- Data collection
- Data preparation
- Geocoding of addresses
- Processing of geocoding results
- Imputation of missing information for priority filings
- Clustering and visualization

Data collection

- Intense and time-consuming
- PATSTAT/Regpat: incomplete with respect to coverage of inventor and applicant addresses

Data collection

- Patent applicant and inventor addresses (or at least city / post code) from

USPTO	Patstat	Patstat
EPO		
DPMA		Data contract
WIPO		Regpat
INPI		Personal contact
UK patent office		Contact by mail
JPO		Bulk download from homepage
KIPO		API
SIPO (applicant information)		Personal contact

-> extremely comprehensive address database (IP5 + further European patent offices)

Data preparation

- Data coverage and quality released to the public differ widely across and within patent offices
- Address data in patent documents many inconsistencies, spelling mistakes and coverage problems
  - Wrong country code: "ROMANLAAN 8 4527 AARDENBURG, BE”
  - Spelling mistake: “OLD MARK 13 3000 LOWEN, BE”
- Comprehensive cleaning procedures for all addresses (adjusted to country specificities, language etc.)
- Leads to about 7 mio. unique addresses

Data preparation

Rules for preprocessing

- Change everything to upper cases and delete all non-alpha numeric characters
- Remove all umlauts and accents etc.
- Delete trailing zeroes, repeated special characters or punctuation marks
- Delete trailing spaces or consecutive spaces within string or at the end
- Delete single characters
- Delete country names that appear in an address
- Make sure that abbreviations for counties/states/regions are used consistently or delete them

Geocoding of addresses

- Geocoding of all unique addresses with Google Maps API
  - Google Maps can geolocalize addresses all over the world in many languages
  - It can deal with typical problems such as spelling mistakes and ambiguities to a certain degree
  - Still wrongly assigned geocoordinates and about 11.7% null results

Geocoding of addresses

https://github.com/dportabella/BatchGeocodingInScalaUsingGoogleAPI

BatchGeocodingInScalaUsingGoogleAPI

This is a simple program for parsing a list of addresses using google maps api.

The main function for parsing a single address is implemented on src/main/scala/AddressParser.scala:

- It builds the proper URL google maps query with the requested address
- It uses Play Json to parse the json response
- It extracts the data we need and it builds an Address case class

BatchParsderCmd queries the addresses from a database, queries each address using google maps api, parses it and saves the result to the DB. It is implemented in Scala with akka streams and slick for optimal performance.

Geocoding of addresses

https://gist.github.com/shanealynn/033c8a3cacdba8ce03cbe116225ced31

Python script for batch geocoding of addresses using the Google Geocoding API.

This script allows for massive lists of addresses to be geocoded for free by pausing when the geocoder hits the free rate limit set by Google (2500 per day). If you have an API key for paid geocoding from Google, set it in the API key section.

Addresses for geocoding can be specified in a list of strings "addresses". In this script, addresses come from a csv file with a column "Address". Adjust the code to your own requirements as needed. After every 500 successul geocode operations, a temporary file with results is recorded in case of script failure / loss of connection later. Addresses and data are held in memory, so this script may need to be adjusted to process files line by line if you are processing millions of entries.

Shane Lynn

5th November 2016

Processing of Google results

- Manual inspection of thousands of addresses
- Most common problems in wrong or ambigous Google results
  - Google yields another country than in the queried address (1.4% of all results)
  - Significantly different postal code (1.1% of all results from addresses with postal codes)
  - Google yields multiple results (1.2% of all results)
  - Google yields too precise information (0.7% of the relevant results)
- If available extraction of postal codes with regular expressions from the queried addresses
- Matching of postal codes + different admin names combinations from geonames.org, use coordinates for critical cases and null results

Processing of Google results

Examples of critical results

Queried address "6350 070 CARAPICUIBA, BR"

Google result "R. Carapicuíba, 70 - Marilândia Jatobá (Barreiro), Belo Horizonte - MG, 30692-570, Brazil»

Queried address "SPOTORNO, IT»

Google result "via, Lungomare Guglielmo Marconi, 60, Spotorno SV, Italy»

Processing of Google results

Extraction of postal codes in PostgreSQL

update addresses_googleresponse_all set postal_code_ = ltrim(array_to_string(regexp_matches(address_, '\m[0-9]{3}\M\s+\m[0-9]{2}\M', 'g'), '')) where country_ in ('Slovakia','Sweden', 'Greece');

update addresses_googleresponse_all set postal_code_ = ltrim(array_to_string(regexp_matches(address_, '[0-9]{5}-[0-9]{3}', 'g'), '')) where country_ = 'Brazil';

Table 9-14. Regular Expression Quantifiers

Quantifier	Matches
*	a sequence of 0 or more matches of the atom
+	a sequence of 1 or more matches of the atom
?	a sequence of 0 or 1 matches of the atom
{m}	a sequence of exactly m matches of the atom
{m,}	a sequence of m or more matches of the atom
{m,n}	a sequence of m through n (inclusive) matches of the atom; m cannot exceed n
*?	non-greedy version of *
+?	non-greedy version of +
??	non-greedy version of ?
{m}?	non-greedy version of {m}
{m,}?	non-greedy version of {m,}
{m,n}?	non-greedy version of {m,n}

Processing of Google results

geonames.org

The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

New South Wales - postal codes

Australia | New South Wales | search

Either enter a postal code (eg. "9011", "AB1", "9980-999") or a city (eg. "London")

	Place	Code	Country	Admin1	Admin2	Admin3
1	Sydney -33.868/151.207	2000	Australia	New South Wales	SYDNEY STREETS	
2	Sydney -33.868/151.207	2001	Australia	New South Wales	SYDNEY BOXES	
3	The University Of Sydney -33.87/151.202	2006	Australia	New South Wales	SYDNEY STREETS	
4	Broadway -33.87/151.202	2007	Australia	New South Wales	SYDNEY STREETS	
5	Mosman -33.839/151.24	2088	Australia	New South Wales	CHATSWOOD	
6	Fairfield -33.867/150.95	1860	Australia	New South Wales	BANKSTOWN	
7	Liverpool -33.9/150.933	1871	Australia	New South Wales	CAMPBELLTOWN	
8	The Rocks -33.859/151.208	2000	Australia	New South Wales	SYDNEY STREETS	
9	Ultimo -33.877/151.197	2007	Australia	New South Wales	SYDNEY STREETS	

Research Policy 42 (2013) 720–737

Contents lists available at SciVerse ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

The worldwide count of priority patents: A new indicator of inventive activity

Gaétan de Rassenfosse[a,*], Hélène Dernis[b], Dominique Guellec[b], Lucio Picci[c,d], Bruno van Pottelsberghe de la Potterie[e]

[a] The University of Melbourne, Melbourne Institute of Applied Economic and Social Research, and Intellectual Property Research Institute of Australia, Level 7 Alan Gilbert Building, 161 Barry Street, Carlton, Victoria 3010, Australia
[b] Organisation for Economic Co-operation and Development, 2 rue A. Pascal, 75775 Paris Cedex 16, France
[c] Università di Bologna, Department of Economics, 45 Strada Maggiore, 40125 Bologna, Italy
[d] IPTS-JRC, Inca Garcilaso, 41092 Seville, Spain
[e] Université libre de Bruxelles, Solvay Brussels School of Economics and Management, tCite, ECARES and CEB, 50 Av. F. Roosevelt, 1050 Bruxelles, Belgium

ARTICLE INFO

Article history:
Received 28 June 2011
Received in revised form 22 October 2012
Accepted 22 November 2012
Available online 20 December 2012

JEL classification:
O30
O57

Keywords:
Patent count
Patent indicator
Patent statistics
Patstat
Priority count
Priority filing
Worldwide count

ABSTRACT

This paper describes a new patent-based indicator of inventive activity. The indicator is based on counting all the priority patent applications filed by a country's inventors, regardless of the patent office in which the application is filed, and can therefore be considered as a complete 'matrix' of all patent counts. The method has the advantage of covering more inventions than the selective Patent Cooperation Treaty (PCT) or triadic families counts, while at the same time limiting the home-country bias of single-country-based indicators (inventors from a particular country tend to file in their own country). The indicator is particularly useful to identify emerging technologies and to assess the innovation performance of developing economies.

Imputation of missing information for priority filings

- Priority filings are the first filings within a family of patents and thus closest to invention date
  - ➢ We want to measure inventiveness of clusters and need a timely indicator
- Many addresses missing for respective priority filing
  - Algorithm from de Rasenfosse, Dernis, Guellec, Picci, van Pottelsberghe de la Potterie (2013)
  - Exploit family linkages in order to recover missing information
  - Adapted to the imputation of coordinates

Browsing 6 sources of information

1 Priority document

2 Earliest direct equivalent

3 Earliest other second filing

4 Applicant information from priority document

5 Applicant information from earliest direct equivalent

6 Applicant information from earliest other second filing

Clustering and visualization

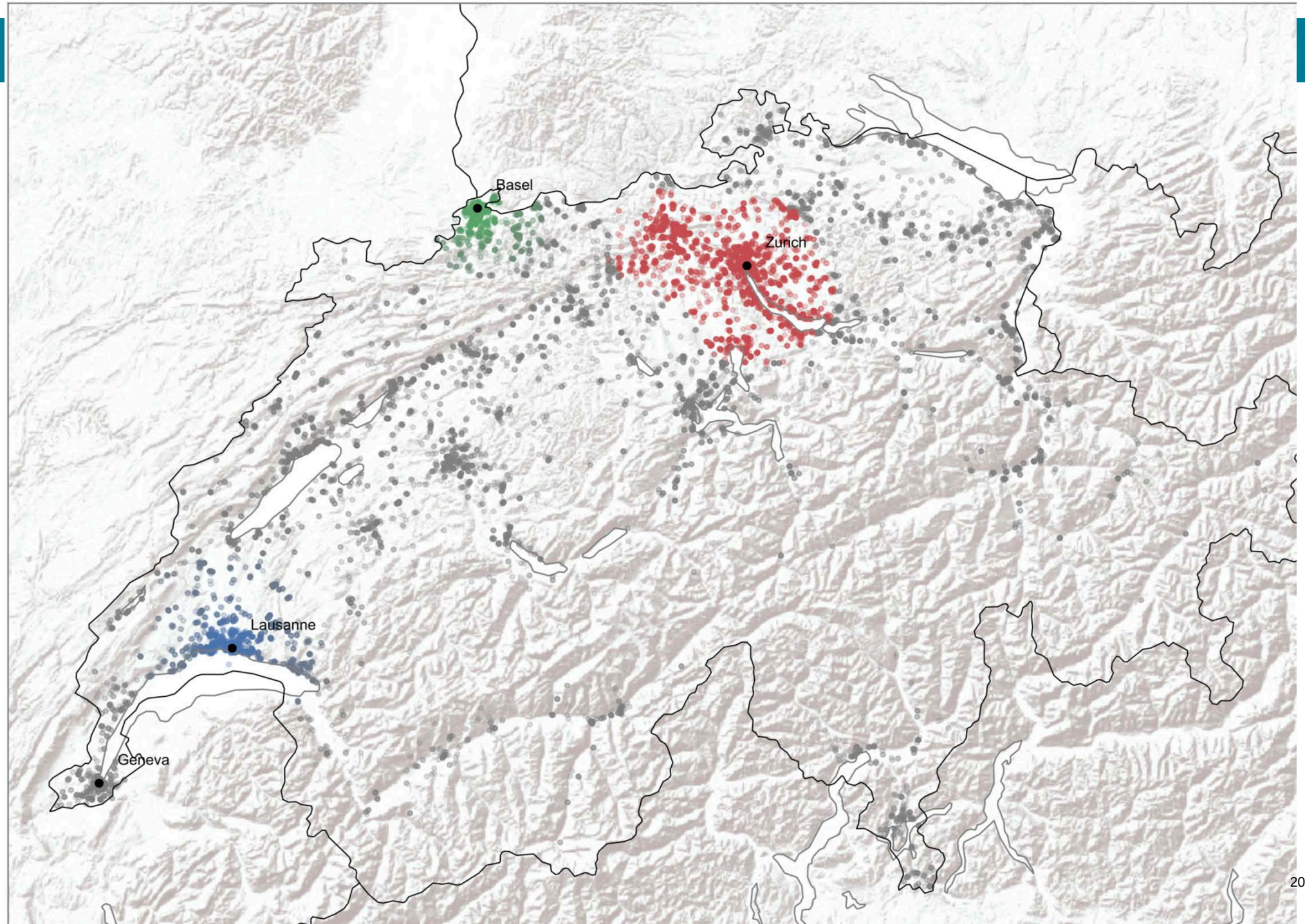
- Clustering algorithm (H)DBSCAN (Ester et al. 1996, Campello, Moulavi, Zimek, 2013) with haversine distances
  - Does not require number of clusters as parameter
  - Clusters of arbitrary shape
  - Two parameters: radius and the minimum of data points in neighborhood to define a cluster
  - Identify core points and expand the cluster by adding all directly-reachable points and find all density-reachable points
  - HDBSCAN searches over all radiuses to find clusters that persist for many values of them
- By world region, time window
- Eventually by country, year, technology
- Results depend on ex-ante assumptions and choice of parameters

Clustering and visualization

Switzerland

Based on patent families assigned to inventor addresses, 2010

(USPTO and EPO, minimum cluster size = 1’000 data points)

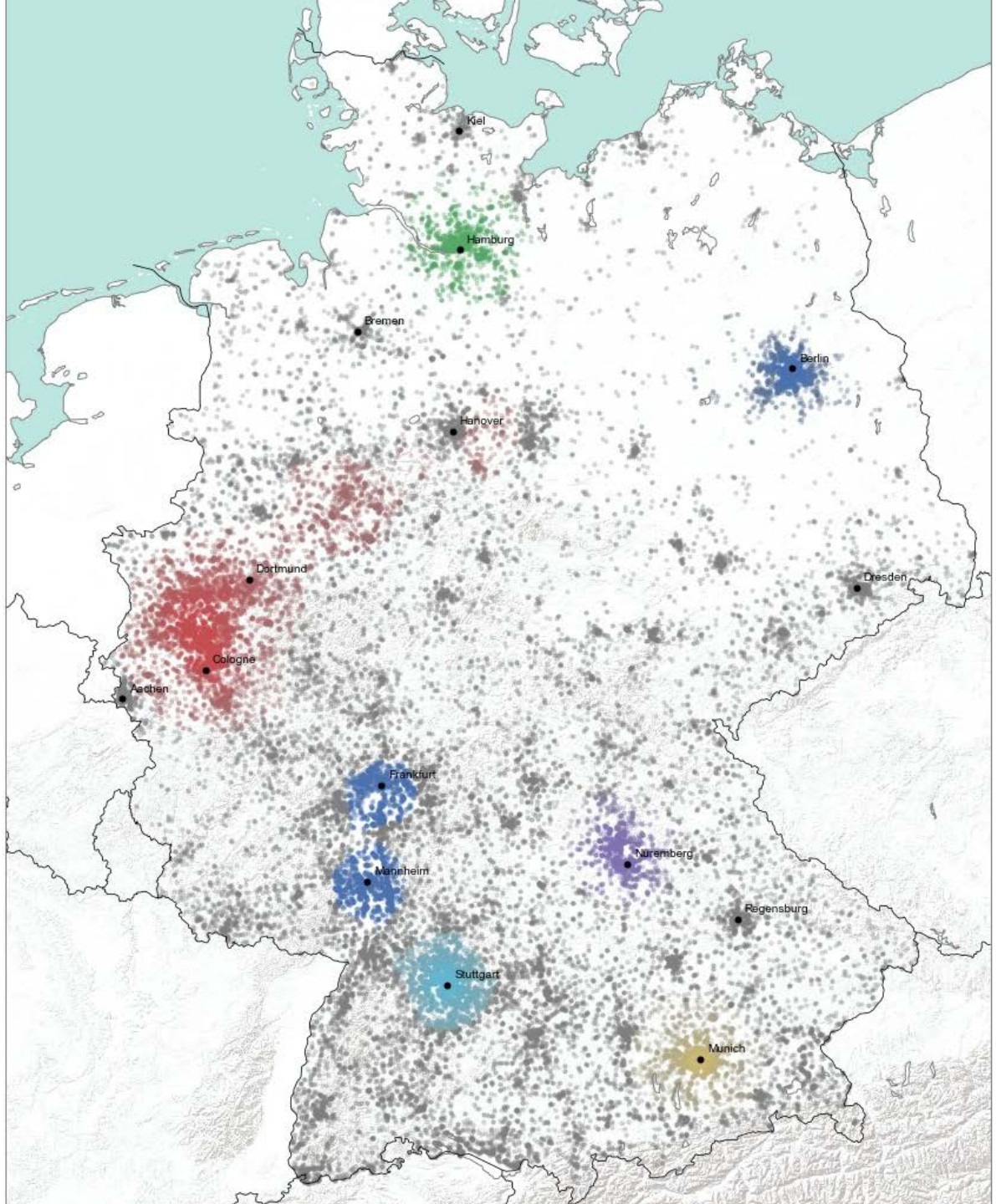

Basel
Zurich
Lausanne
Geneva

Clustering and visualization

Germany

Based on inventor addresses, 2010

(USPTO and EPO)


Kiel
Hamburg
Bremen
Berlin
Hanover
Dortmund
Dresden
Cologne
Aachen
Frankfurt
Mannheim
Nuremberg
Regensburg
Stuttgart
Munich

Research output – in preparation

Data

- Application identifier from PATSTAT, corresponding coordinates, cluster identifiers

appln_id	latitude	longitude	cluster
1	47.5146	7.6039	1
2	47.2729	8.7205	2
3	47.3983	8.4488	3
4	47.5848	7.6499	1
5	46.4761	6.4302	0

- The data will be made publicly available

Data visualization

- Maps by world regions showing evolution over time
- Published on a project website

Product and Process Inventions in Patent Data

Knowledge Spillovers from Product and Process Inventions

Sebastian Heinrich, Florian Seliger, Martin Wörter (KOF Swiss Economic Institute, ETH Zurich)

In collaboration with

Vanessa Behrens & Christian Rammer (ZEW Mannheim)

Ulrich Schmoch (Fraunhofer ISI, Karlsruhe)

Stefan Feuerriegel (ETH Zurich)

Funding from the European Patent Office Academic Research Program is acknowledged

Introduction

Process inventions – efficiency-enhancing activities that are aimed at

- lowering the cost of producing a good or service
- improving product quality
- in the language of patent examiners, **process claims** include
  - «all kind of activities in which the **use** of some material product for effecting the process is implied» (EPO),
  - «they define **steps, acts, or methods** (…) and include a **new use** of a known process, machine, manufacture, or compositions of matter» (USPTO)
- More difficult to appropriate, slower knowledge leakage to competitors (Levin, Klevorick, Nelson, Gilbert, Griliches, 1987)

Introduction

Lack of data

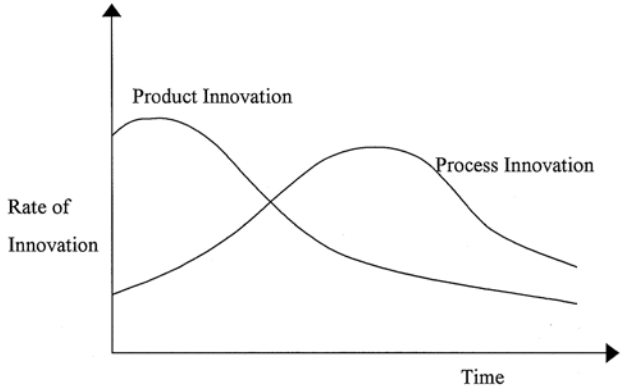
- Apart from CIS, lack of empirical data / insufficient data with respect to process inventions
- Few attempts to distinguish product and process patents (Scherer 1982, Cohen & Klepper 1982, Bena & Simintzi 2017)

Lack of studies

- Most of the scientific debate is theoretical (e.g. Rosenkranz 2003, Boone 2000, Bonanno & Harworth 1998)

Goals of the project

- Categorization of patents (product vs. process patents)
- Trace technological life-cycles



From Utterback and Abernathy (1975)

- Bridge the gap between process innovations in economics and processes in patents (patent attorney/examination view)

Project stages

- Identification of product and process patents
  - Keyword search in titles, abstracts, and claims
  - Manual classification and text mining
- Detailed descriptive analysis by country and technology fields
- Analysis of lifecycles
- Econometric analysis

Keyword search in title, abstract, and claims

- PATSTAT: Abstracts and titles, at least one entry per family
- Full-text Claims:
  - EPO backfile containing EP-A and EP-B documents (1978 – 2016)
  - USPTO Claims Research Dataset (1976 – 2014): US patents granted between 1976 and 2014 and US patent applications pub- lished between 2001 and 2014
- Data available in XML and CSV format
- Parsed and imported to a PostgreSQL database using a Python program
- Keyword search implemented in PostgreSQL.

Keyword search in title, abstract, and claims

- A set of keywords was derived through manual search process in EPO patents
- Identification of product patents incomplete as in many cases the labeling of the specific product is used instead of more abstract terms
- Identification of process patents quite complete

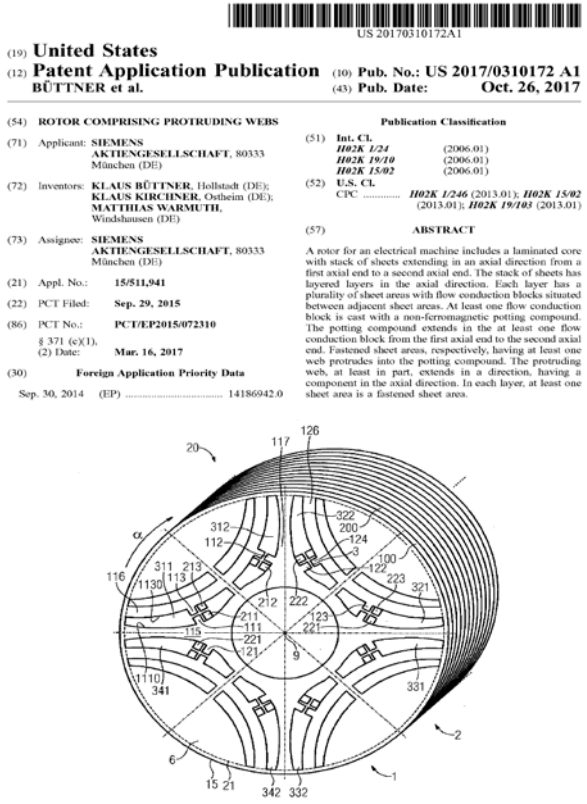
Product keywords	**Process keywords**
device	method
machine	process
material	procedure
tool	use of
apparatus	application of
vehicle	
compound	
composition	
substance	
article	

Keyword search in title, abstract, and claims

- Each text field was checked fo the occurrence of one of the process keywords
  - EPO patents: English, French, German
  - **Abstracts and Titles**: Classified as Process or Product
  - **Claims**: Share of Process Claims per Patent (Number of Claims classified as Process divided by Total Number of Claims per Patent)
    - Unambigious classification of individual claims is possible
    - Not the case for abstracts and titles: in many abstracts and titles both product and process descriptions can be found
- Main advantage of the **Claims Process Share**: Shows to which degree a patent is a product and process at the same time
- Be aware that often information is only published once and that missing data must be looked up in equivalents or other family members

Keyword search in title, abstract, and claims

Example US20173101 72A1 Siemens (published Oct. 26, 2017)



US 20170310172A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2017/0310172 A1**

BÜTTNER et al. (43) **Pub. Date: Oct. 26, 2017**

(54) **ROTOR COMPRISING PROTRUDING WEBS**

(71) Applicant: **SIEMENS AKTIENGESELLSCHAFT**, 80333 München (DE)

(72) Inventors: **KLAUS BÜTTNER**, Hollstadt (DE); **KLAUS KIRCHNER**, Ostheim (DE); **MATTHIAS WARMUTH**, Windshausen (DE)

(73) Assignee: **SIEMENS AKTIENGESELLSCHAFT**, 80333 München (DE)

(21) Appl. No.: **15/511,941**

(22) PCT Filed: **Sep. 29, 2015**

(86) PCT No.: **PCT/EP2015/072310**

§ 371 (c)(1),
(2) Date: **Mar. 16, 2017**

(30) **Foreign Application Priority Data**

Sep. 30, 2014 (EP) 14186942.0

Publication Classification

(51) **Int. Cl.**
H02K 1/24 (2006.01)
H02K 19/10 (2006.01)
H02K 15/02 (2006.01)

(52) **U.S. Cl.**
CPC ***H02K 1/246*** (2013.01); ***H02K 15/02*** (2013.01); ***H02K 19/103*** (2013.01)

(57) **ABSTRACT**

A rotor for an electrical machine includes a laminated core with stack of sheets extending in an axial direction from a first axial end to a second axial end. The stack of sheets has layered layers in the axial direction. Each layer has a plurality of sheet areas with flow conduction blocks situated between adjacent sheet areas. At least one flow conduction block is cast with a non-ferromagnetic potting compound. The potting compound extends in the at least one flow conduction block from the first axial end to the second axial end. Fastened sheet areas, respectively, having at least one web protrudes into the potting compound. The protruding web, at least in part, extends in a direction, having a component in the axial direction. In each layer, at least one sheet area is a fastened sheet area.

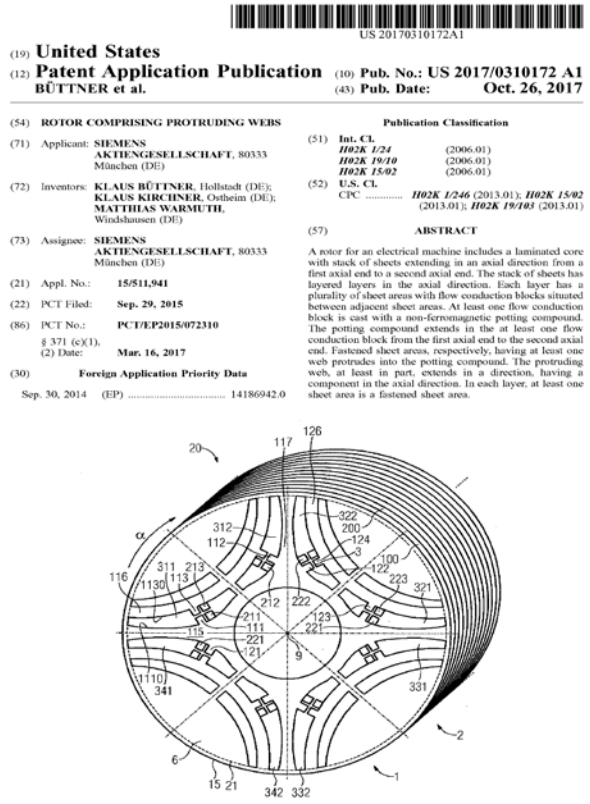
Title: Rotor comprising protruding webs **-> PRODUCT**

Abstract **-> PRODUCT**

A rotor for an electrical machine includes a laminated core with stack of sheets extending in an axial direction from a first axial end to a second axial end. The stack of sheets has layered layers in the axial direction. Each layer has a plurality of sheet areas with flow conduction blocks situated between adjacent sheet areas. At least one flow conduction block is cast with a non-ferromagnetic potting compound. The potting compound extends in the at least one flow conduction block from the first axial end to the second axial end. Fastened sheet areas, respectively, having at least one web protrudes into the potting compound. The protruding web, at least in part, extends in a direction, having a component in the axial direction. In each layer, at least one sheet area is a fastened sheet area.

Keyword search in title, abstract, and claims

Example US20173101 72A1 Siemens (published Oct. 26, 2017)



US 20170310172A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2017/0310172 A1**

BÜTTNER et al. (43) **Pub. Date: Oct. 26, 2017**

(54) **ROTOR COMPRISING PROTRUDING WEBS**

(71) Applicant: **SIEMENS AKTIENGESELLSCHAFT**, 80333 München (DE)

(72) Inventors: **KLAUS BÜTTNER**, Hollstadt (DE); **KLAUS KIRCHNER**, Ostheim (DE); **MATTHIAS WARMUTH**, Windshausen (DE)

(73) Assignee: **SIEMENS AKTIENGESELLSCHAFT**, 80333 München (DE)

(21) Appl. No.: **15/511,941**

(22) PCT Filed: **Sep. 29, 2015**

(86) PCT No.: **PCT/EP2015/072310**

§ 371 (c)(1), (2) Date: **Mar. 16, 2017**

(30) **Foreign Application Priority Data**

Sep. 30, 2014 (EP) 14186942.0

Publication Classification

(51) **Int. Cl.**
H02K 1/24 (2006.01)
H02K 19/10 (2006.01)
H02K 15/02 (2006.01)

(52) **U.S. Cl.**
CPC ***H02K 1/246*** (2013.01); ***H02K 15/02*** (2013.01); ***H02K 19/103*** (2013.01)

(57) **ABSTRACT**

A rotor for an electrical machine includes a laminated core with stack of sheets extending in an axial direction from a first axial end to a second axial end. The stack of sheets has layered layers in the axial direction. Each layer has a plurality of sheet areas with flow conduction blocks situated between adjacent sheet areas. At least one flow conduction block is cast with a non-ferromagnetic potting compound. The potting compound extends in the at least one flow conduction block from the first axial end to the second axial end. Fastened sheet areas, respectively, having at least one web protrudes into the potting compound. The protruding web, at least in part, extends in a direction, having a component in the axial direction. In each layer, at least one sheet area is a fastened sheet area.

Claims

1.-10. A rotor for an electrical machine, comprising a laminated core extending in an axial direction from a first axial end to a second axial end (…) **-> PRODUCT**

11. An electrical machine comprising a rotor mounted for rotation about a rotation axis extending in an axial direction (…) **-> PRODUCT**

12. A motor vehicle, comprising a rotor including a laminated core extending in an axial direction from a first axial end to a second axial end (…) **-> PRODUCT**

13. A laminated core for a rotor, comprising layers stacked in an axial direction to define a first axial end and a second axial end (…)

14. A method for manufacturing a rotor, comprising stacking layers in an axial direction to establish a lamination core, with each layer having a plurality of lamination regions (…) **-> PROCESS**

15. A method of using laminations for a rotor, comprising:
stacking layers having a plurality of fastened lamination regions in an axial direction of a laminated core extending in the axial direction from a first axial end to a second axial end (…) **-> PROCESS**

-> the process share is 2/15 = 13 %

Manual classification of claims and abstracts and text-mining

Random sample of about 1‘100 patents

- USPTO and EPO, only granted patents
- Abstracts and individual claims were classified manually by student helpers
- One third of the patents were classified twice (interrater reliability)
- Detailed guidelines on what defines a product and process patent according to
  - our list of keywords
  - the EPO and USPTO examination guidelines

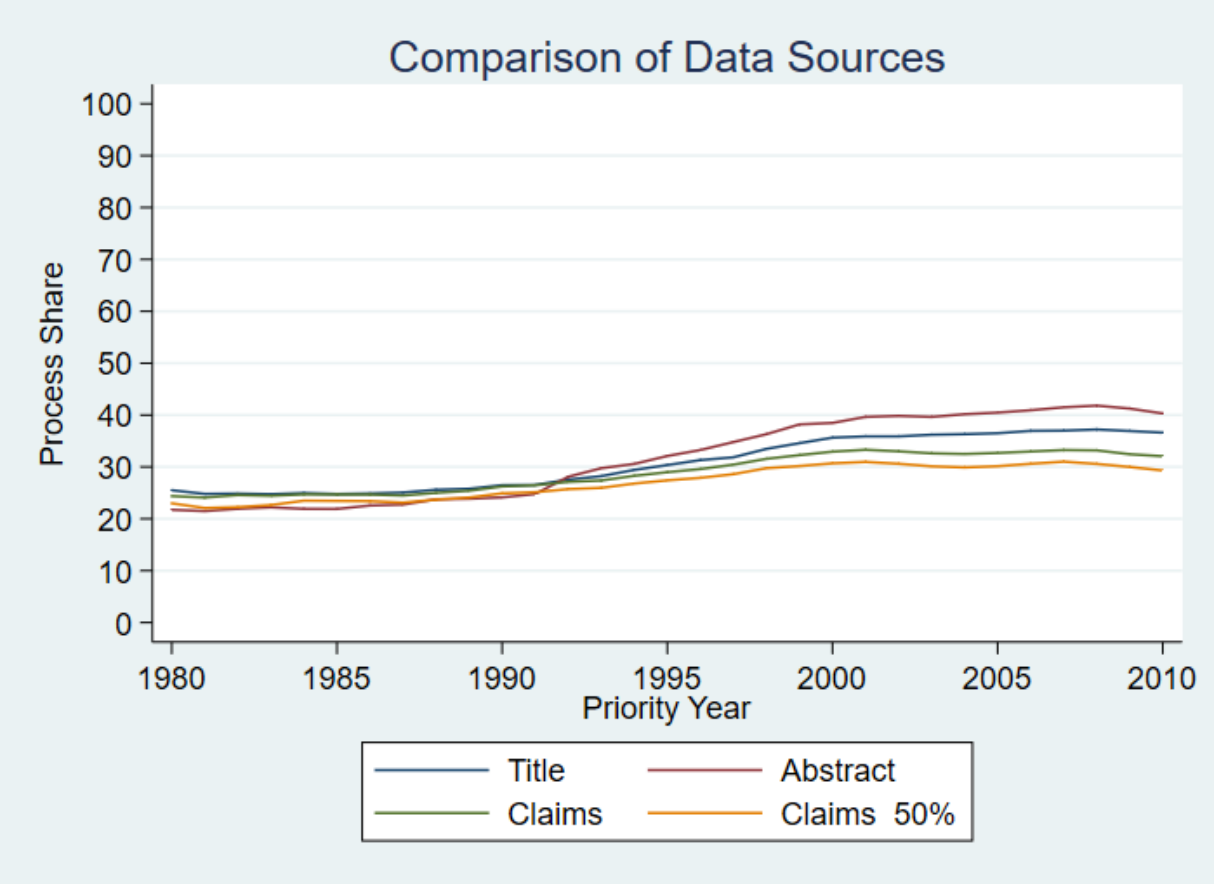
Manual classification of claims and abstracts, and text-mining

Preliminary results for pre-sample of 100 patents

- Classification of patents with about 80% balanced accuracy for the product category using a Random Forest approach
- Regression task in order to predict the ratio of product to process claims

Descriptive analysis

Share of process claims / patents in % for different text fields over time



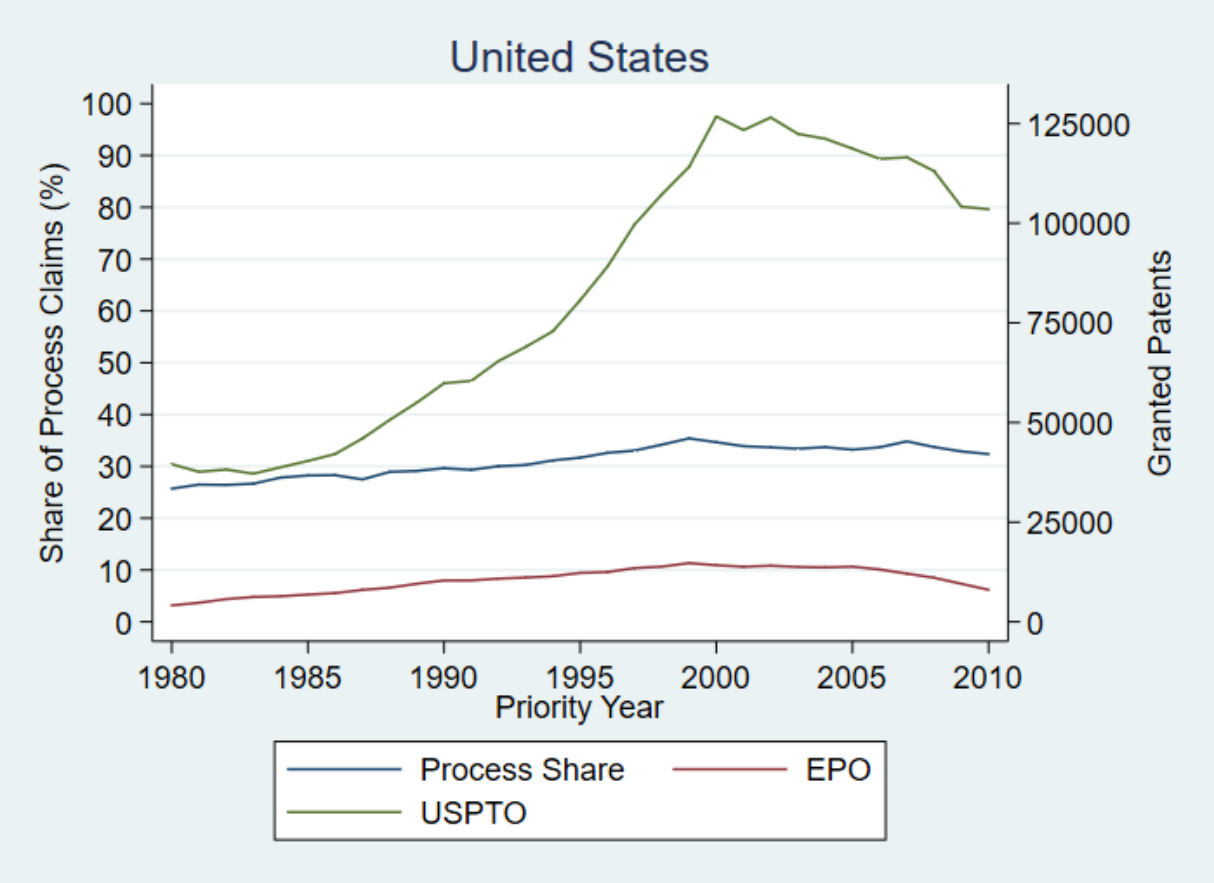
Descriptive analysis

Comparison by country of inventor

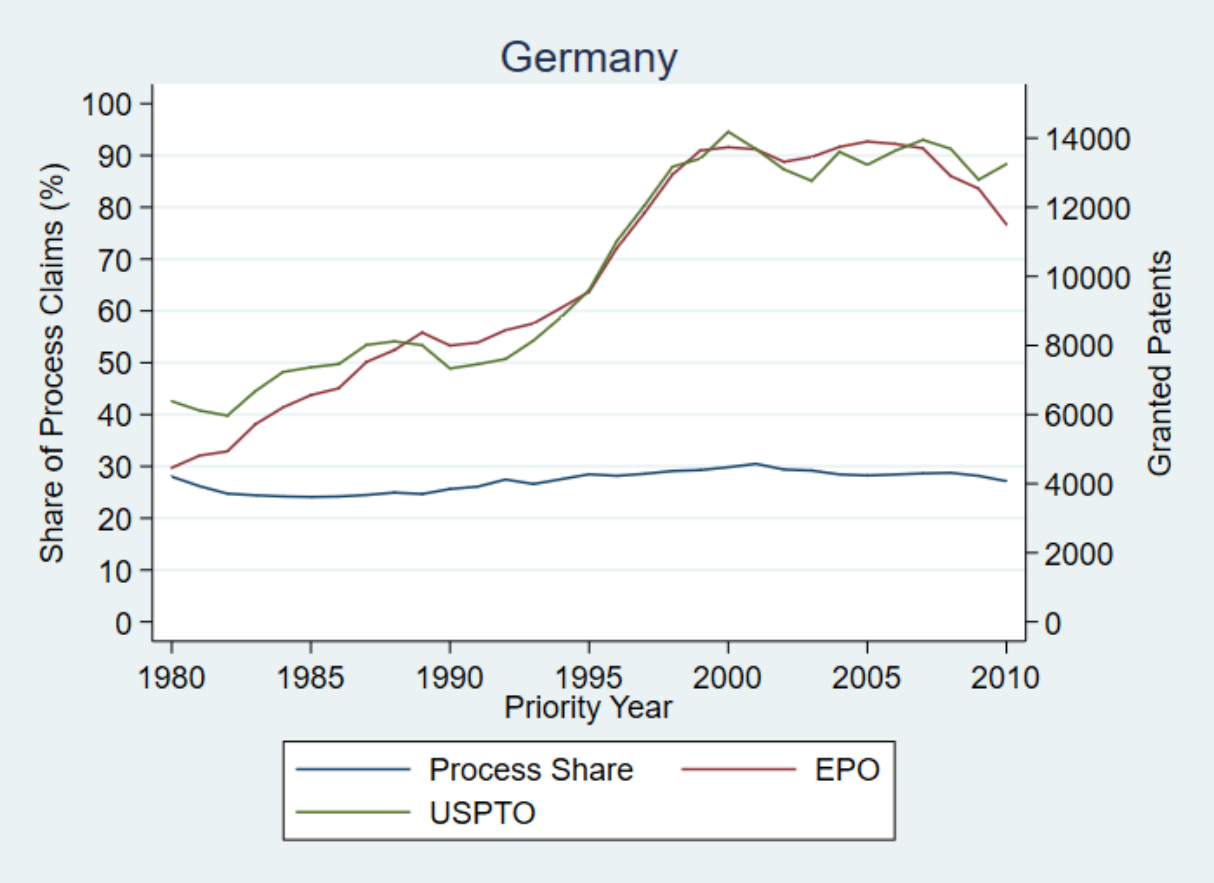
- EPO and USPTO
- Granted Patents
- Priority Year 1980 – 2010
- Patstat-Version: Autumn 2017

- For the number of patents per country a fractional count of inventors is used.

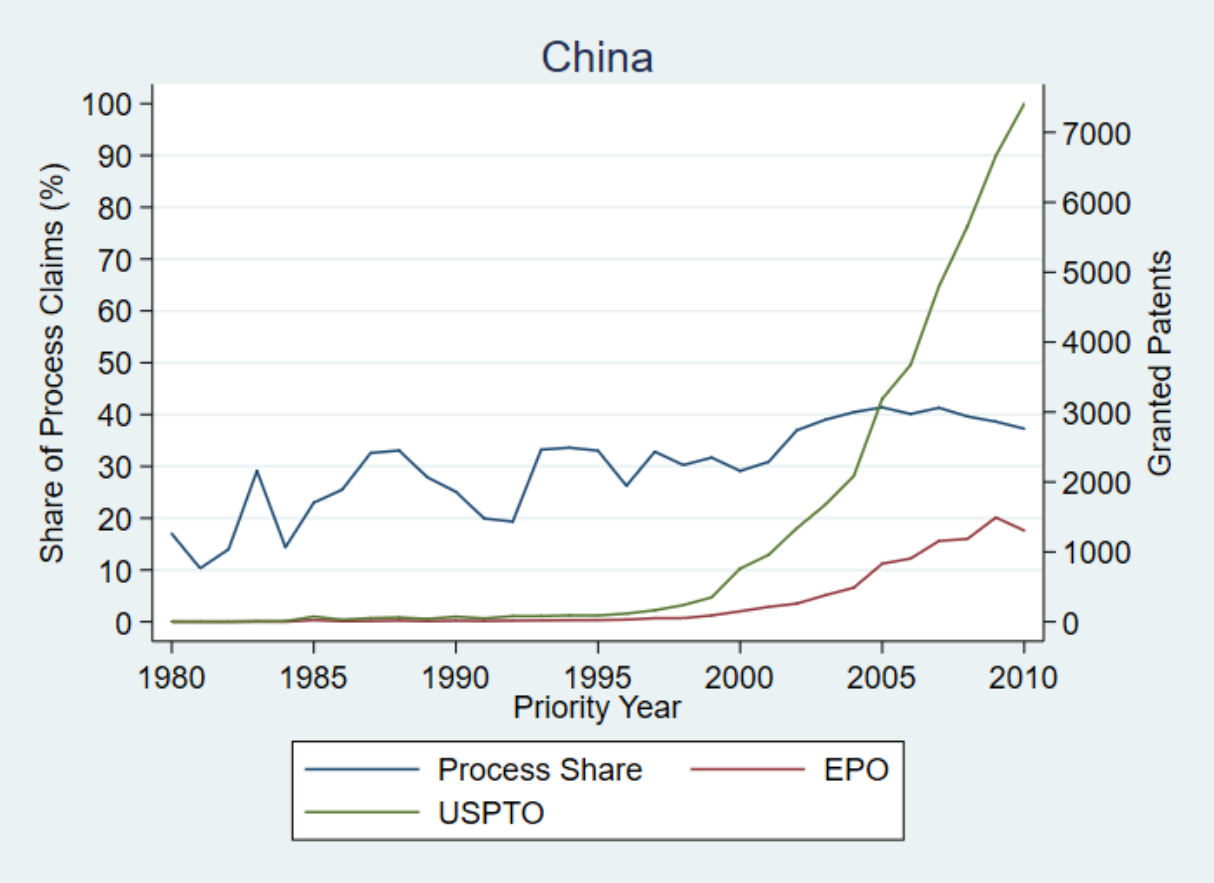
Descriptive analysis



Descriptive analysis



Descriptive analysis



Descriptive analysis

Preliminary summary of country analysis

- Potential relationship between economic / technological maturity and the process share
  - Laggards might need to focus on processes during catch-up
  - Big mature economies also seem to have increasing process shares (at low level)

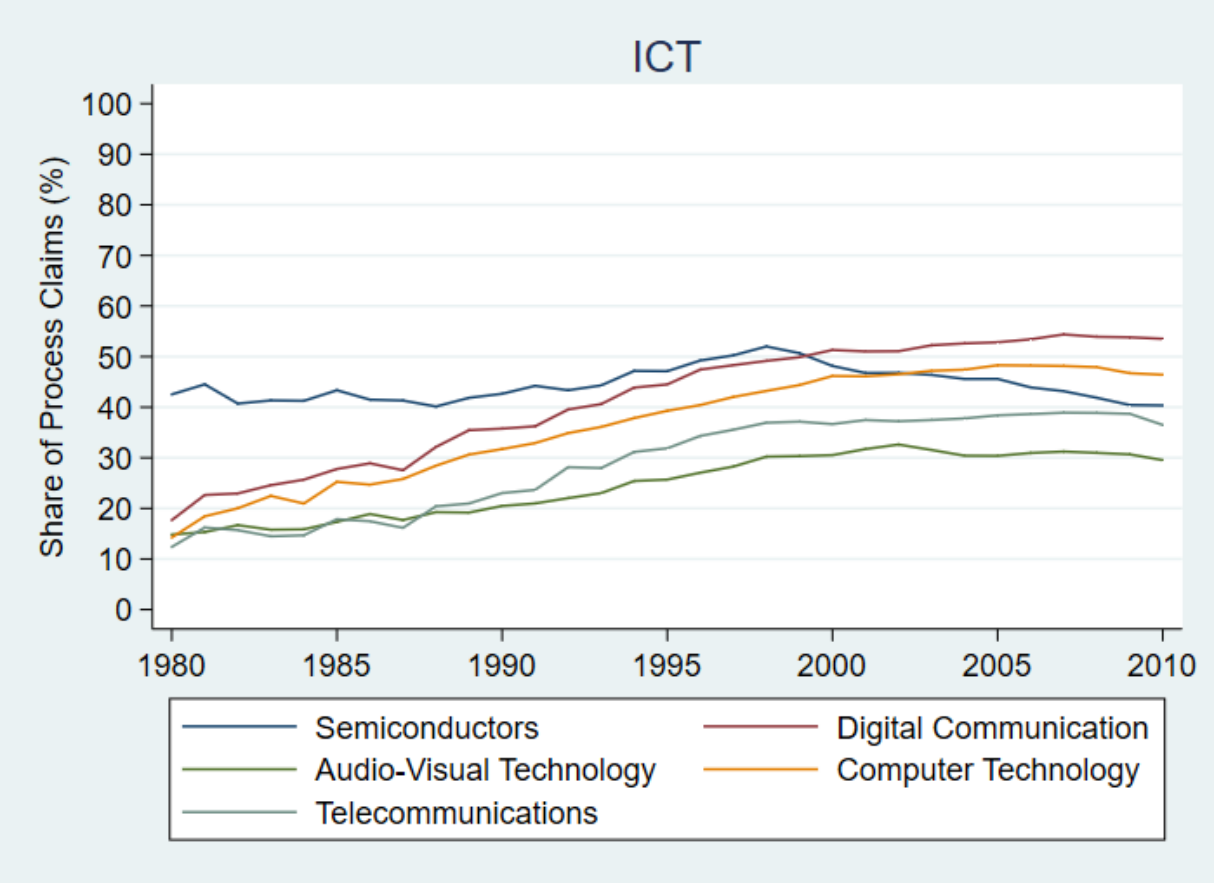
Descriptive analysis

Comparison by technological fields

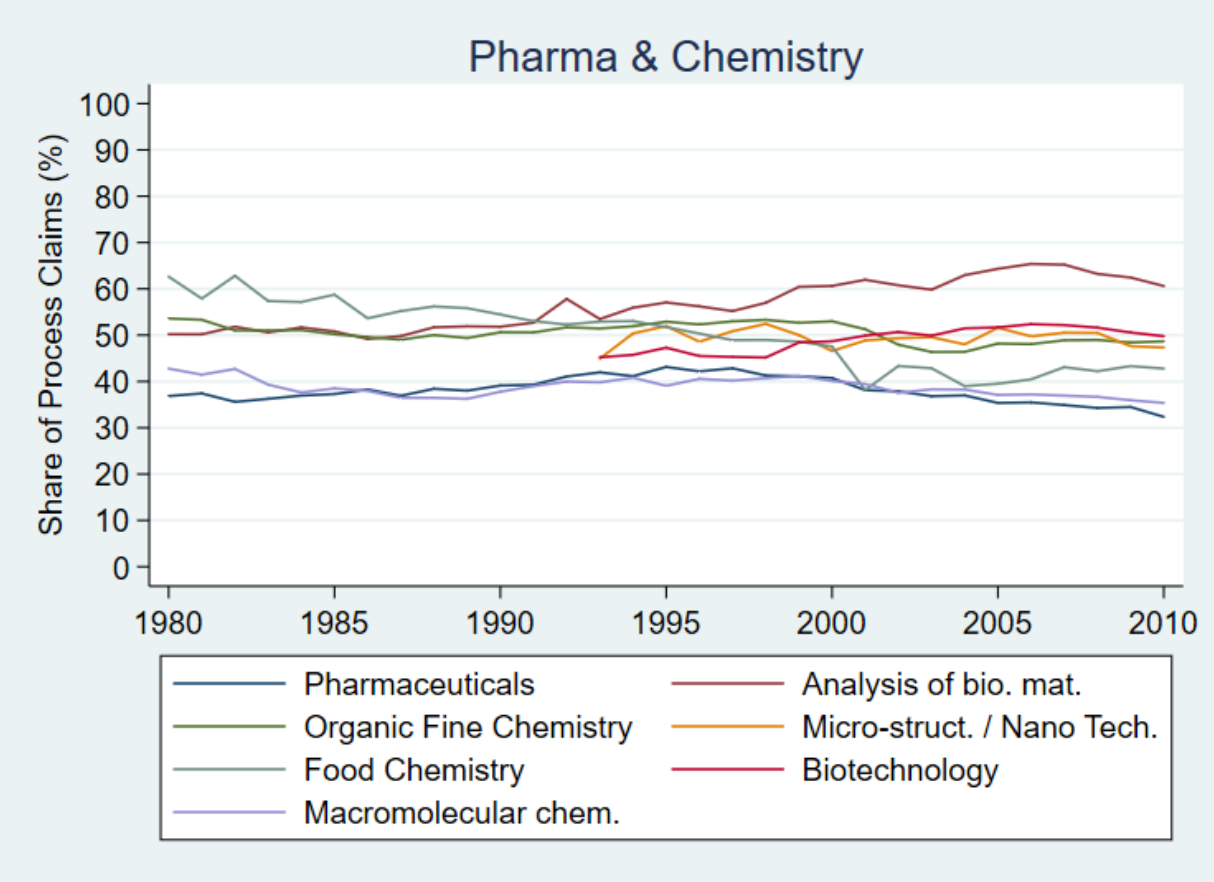
- EPO and USPTO
- Granted Patents
- Priority Year 1980 – 2010
- Patstat-Version: Autumn 2017

- Technological fields come from Patstat table tls901_techn_field containing a mapping between 35 technical fields and the much more detailed IPC classification (Schmoch 2008)
- For technological fields, weights at application level provided by Patstat are used (tls230_appln_techn_field.weight)

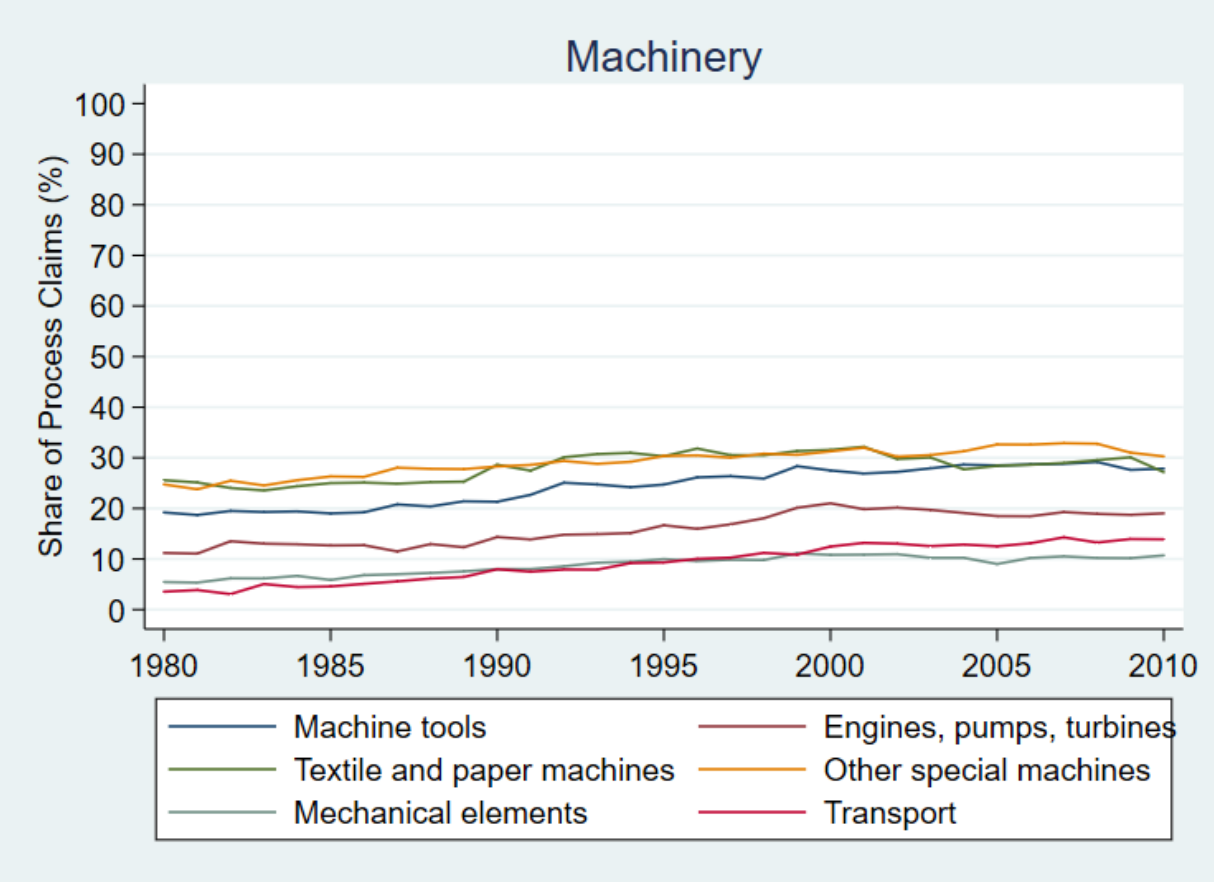
Descriptive analysis



Descriptive analysis



Descriptive analysis



Descriptive analysis

Preliminary summary of comparison by technological fields

- Most of the ICT fields show strongly growing process shares
  - Growth rate strongest in 80s and 90s
  - Greater complexity of computer programs or increasing concentration?
- Some technologies from the Chemistry & Pharmaceuticals field seem to be rather process-driven, Machinery not (but increasing shares)

Relationship of process patenting and process innovations

Marginal effects of Probit regressions – dependent variable: firm is a process innovator

Claims Process Share	0.063***
	(0.022)
Process Title	0.009
	(0.017)
Process Abstract	-0.005
	(0.015)
Firm Size	0.073***
	(0.002)
Log(Applications)	0.021***
	(0.002)
Observations	38,594
Sector FE	YES

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Data

- Mannheim Innovation Panel matched with EP patents and process indicators
- 7,329 firms that filed at least one EP patent between 1992 and 2016
- 235,178 EP patents out of a total of around 2,614,000 EP applications are covered

Relationship of process patenting and process innovations

Refinement of patent-based measures

- Goal: Focusing on economically relevant product and process claims
  - Exclusion of use claims
  - Only inclusion of independent claims

Planned econometric analyses

- Matched firm-level data from ZEW and KOF
- We will estimate something like

$$PERF_{it} = \beta_0 + \beta_1 SPILL\_PROD_{it-1} + \beta_2 SPILL\_PROC_{it-1} + \beta_3 PROD_{it-1} + \beta_4 PROC_{it-1} + X\beta + t_t + u_i + \varepsilon_{it}$$

where

$PERF_{it}$: Some kind of innovative performance (sales with innovations, productivity)

$SPILL\_PROD_{it-1}$: Spillovers related to rivals' product patents,

$SPILL\_PROC_{it-1}$: Spillovers related to rivals' process patents,

$PROD_{it-1}$: Firm *i*'s 'own' product patents,

$PROC_{it-1}$: Firm *i*'s 'own' process patents

Planned econometric analyses

Guiding research questions

- Is knowledge about processes more difficult to appropriate by firms? (-> higher spillover effect)
- Is knowledge leakage to competitors related to process inventions (s)lower compared to product inventions? (-> lower spillover effect)
- Are there complementarities between product and process inventions?

Research output – in preparation

Data

- Publication identifier from PATSTAT, several claims process shares

pat_publn_id	Process_share_ keyword_search	Process_share_ text_mining	Process_share_ only_ independent_ claims
1	0.8	0.7	0.7
2	0.2	0.2	0.2
3	0	0	0
4	0.3	0.4	0.3
5	0.1	0.2	0.2

- The data will be made publicly available

Finally…

The data from the projects will be made publicly available.

- Data on Geolocalization around end of 2018/beginning of 2019
- Data on process and product patents end of 2019
- You will be able to find the data and supplementary material via Google search
- If you have recommendations, research ideas, comments, questions or specific data needs, don‘t hesitate to contact us

seliger@kof.ethz.ch

Thank you for your attention!

seliger@kof.ethz.ch