

European Patent Citations - How to count and how to interpret them?

Dietmar Harhoff

Ludwig-Maximilian University Munich (INNO-tec),
ZEW and CEPR

Karin Hoisl

Ludwig-Maximilian University Munich (INNO-tec)

2

Overview

- Motivation
- Historical Aspects
- EPO Search Process
- Empirical Literature
- Data and Descriptive Statistics
- Multivariate Analysis
- Summary of Results
- Caveats and More Plans

Motivation

- Patent citations are widely used by economists (and others) in empirical analysis.
- EP patent citations are treated analogously to US citations. Justified?
- Most studies disregard problems arising from the presence of *equivalent* documents. How to correct for them?
- All citations are equal. Really true?
 - experiment with classification of references
 - experiment with second-order citations
- How to interpret citation variables in multivariate settings?

Historical Aspects

- **1950s** Eugene Garfield proposes the Science Citation Index and compiles the first Patent Citation Index on chemical patents.
- **1961** Sher/Garfield analyze citation counts of Nobel Prize winners.
- **1972** National Science Foundation publishes its first Science Indicators Report.
- **1976** Francis Narin publishes a report describing the state of the art of citation analysis.
- **1981** Carpenter/Narin/Wolf conduct the first large scale study on patent citations.
- **2002** NBER Patent Citation Data File is made available to interested researchers - containing data on approximately 3 million US patents and more than 16 million citations).

EPO Search Process Institutional Framework

- The Search Process follows the *Guidelines for Examination in the European Patent Office*.
- Overall responsibility lies with the Directorate General for Searching in The Hague.
- European Searches are carried out by the Search Divisions of the EPO.
- Main objective: discovering prior art relevant for determining whether the invention meets the novelty and inventive step requirements.
- The search is conducted in the basis of the claims.

EPO Search Process Termination and Report

- **The Search will be terminated if**
 - the probability of discovering further relevant documents is very low compared to the effort needed, or
 - documents are discovered which doubtlessly demonstrate a violation of patentability requirements.
- **The Search Report should only include**
 - the most important documents,
 - one of several documents of equal importance,
 - the earlier document of two equally important documents.



According to EPO philosophy a good search report contains all relevant information within a minimum number of citations (Michel/Bettels, 2001)

EPO Search Process

Classification of Patent References

X	particularly relevant documents when taken alone (a claimed invention cannot be considered novel or cannot be considered to involve an inventive step)
Y	particularly relevant if combined with another document of the same category
A	documents defining the general state of the art
O	documents referring to non-written disclosure
P	intermediate documents (documents published between the date of filing and the priority date)
T	documents relating to theory or principle underlying the invention (documents which were published after the filing date and are not in conflict with the application, but were cited for a better understanding of the invention)
E	potentially conflicting patent documents, published on or after the filing date of the underlying invention
D	document already cited in the application
L	document cited for other reasons (e.g., a document which may throw doubt on a priority claim)

EPO Search Process

Consequences for Citation Analysis (1/2)

- Difference in interpretation between examiner and applicant citations
 - ? the examiner has to ensure the novelty and inventive step of the invention, the applicant cites work that is related but significantly different from his invention
- The date of filing of the EP application is used as a reference date for the search
 - ? references published between the priority date and the filing date may lead to negative citation lags

EPO Search Process Consequences for Citation Analysis (2/2)

- Citing no more than what is absolutely necessary and obligation to favor early documents over later ones
 - ? other documents important for the economic analysis may not appear in the list of references
- Examiners should preferably cite documents in the language of the application
 - ? overestimation of the influence of the applicant's home country and distortion in citation counts
- The Search Division only cites documents to which it has access
 - ? underestimation of documents not accessible in databases

Empirical Literature Selected Articles

Knowledge Spillovers

Applicability of patent citations to measure knowledge spillovers

- Jaffe et al. (2000) – placebo patents

Geographic localization of knowledge spillovers

- Jaffe et al. (1993) – ITM-method

Determinants of knowledge spillovers

- Almeida/Kogut (1999) – labor mobility

Patent Value

- Trajtenberg (1990) – social value (US citation data)
- Harhoff et al. (1999) – private value (German and US citation data)
- Jaffe/Hall/Trajtenberg (2005) – market value (US citation data)

Duration of Examination/Processes

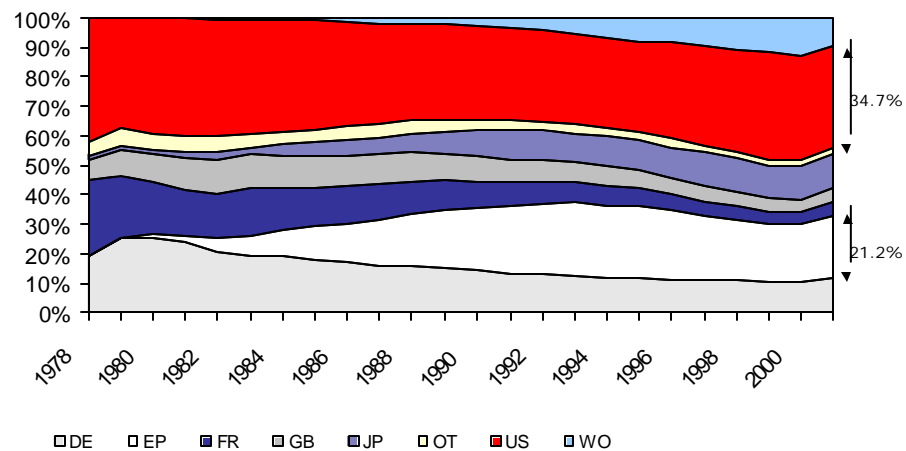
- Regibeau/Rockett (2003) – Harhoff/Wagner (2003) – Johnson/Popp (2003)

Data and Descriptive Statistics

- Data provided by EPO/OECD team
- 5,026,957 references (4,816,212 references without XP entries) in 1,079,365 EPO patent documents
- last entry as of Nov. 28th, 2001
- extensive consolidation work
 - EP/ISA search reports
 - XP references
 - duplicates at the source document level
- Preliminary results available - work continuing on
 - NPL references
 - complete identification of self-references and self-citations
 - complete computation of citations lags

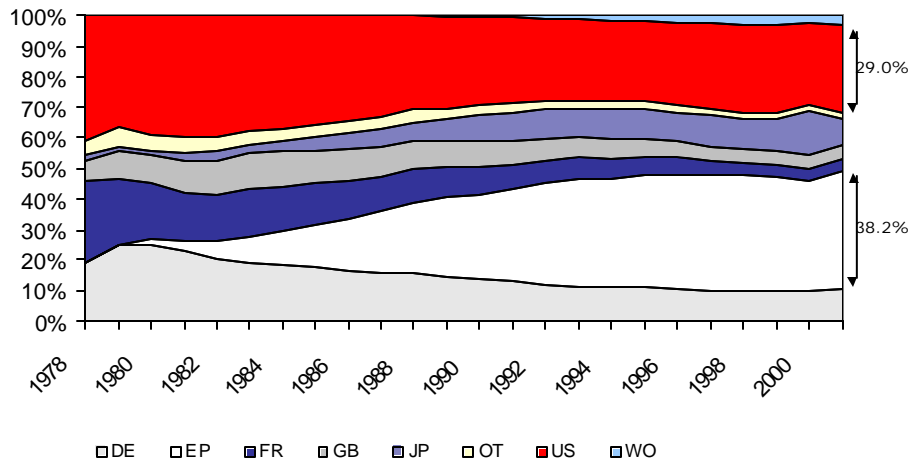
Data and Descriptive Statistics

Source of EP Patent References – Raw Data



Data and Descriptive Statistics

Source of EP Patent References - EP Equivalents Imposed



Data and Descriptive Statistics

Source of EP Patent References – Overall Comparison

Source of Reference

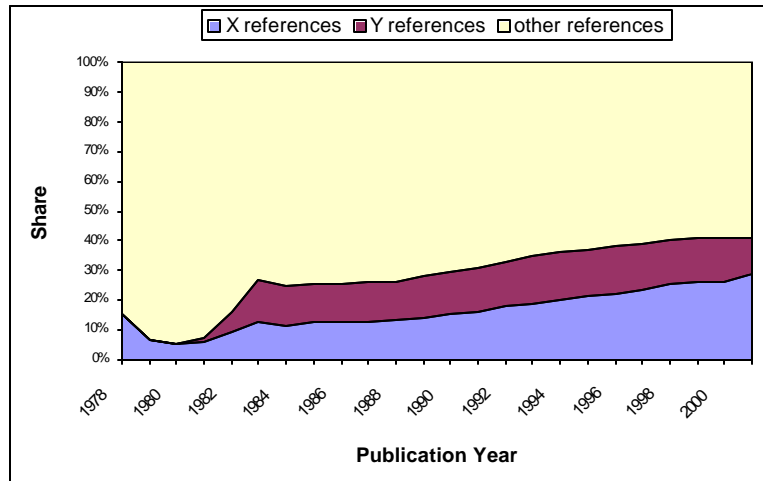
	DE	EP	FR	GB	JP	OT	US	WO
without	13.7	19.0	8.2	7.3	8.8	3.1	33.9	6.0
with	13.2	28.8	8.2	7.3	8.2	3.1	29.9	1.3

correction for equivalents

After correcting for equivalents, the share of within-EP referencing increases from 19.0% to 28.8% of all patent references

Data and Descriptive Statistics

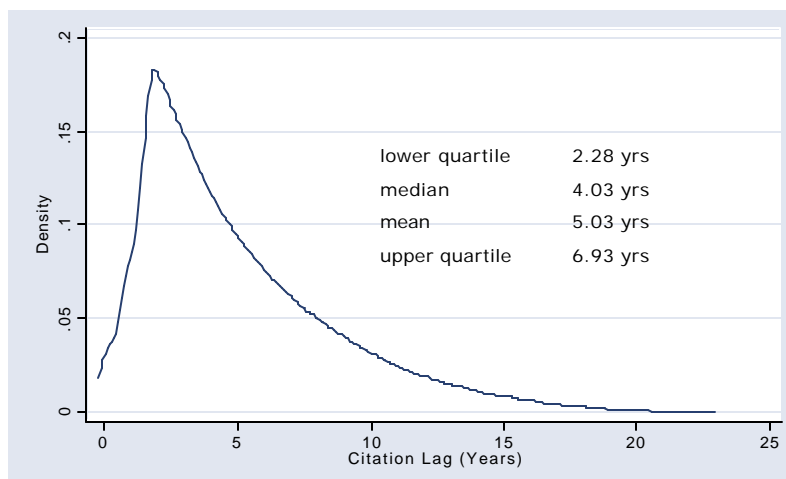
Classification of EP Patent References, 1978-2001



Source: Harhoff/Hoisl (2004) – authors' computations based on EPO/OECD citation database.

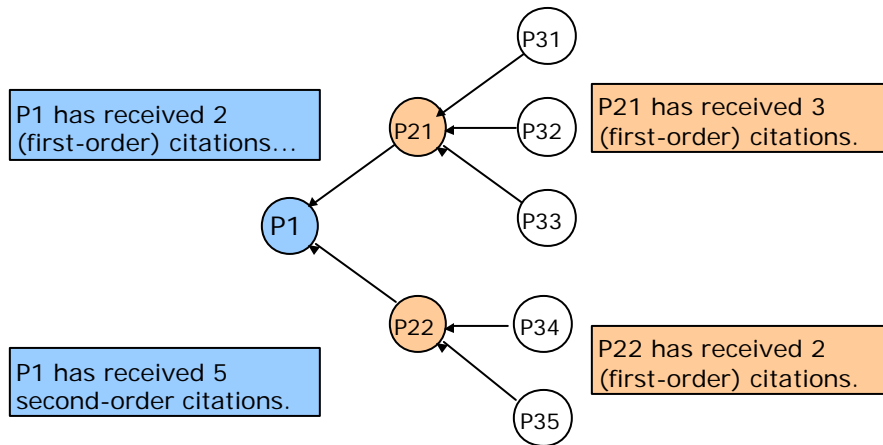
Data and Descriptive Statistics

Citation Lag Distribution (Within-EP Citations Only)



Source: Harhoff/Hoisl (2004) – authors' computations based on EPO/OECD citation database.

2nd Order Citations



Data and Descriptive Statistics

Citation Statistics by Type of Patent Holder

Variable	Independent Inventor (N = 39,671)	University (N = 5,434)	Corporate (N = 550,144)
references	4.88	3.58	4.15
X-references	0.68	0.73	0.56
self-references	0.04	0.05	0.14
NPL references	0.36	2.37	0.73
citations	1.30	2.73	1.81
X-citations	0.10	0.39	0.17
self-citations	0.03	0.05	0.13
2nd order citations	3.83	11.50	6.85
claims	9.3	10.9	7.3
grant lag	4.2 yrs	5.5 yrs	4.3 yrs

Source: Harhoff/Hoisl (2004) – authors' computations based on EPO/OECD citation database.

Multivariate Analysis

Results from three types of multivariate analysis

- Value equation (German PATVAL Data, European PATVAL results to be developed in joint work with an international team)
- Opposition incidence (Harhoff 2004)
- Revocation incidence (Harhoff 2004)

Relevant variables

- number of citations (first and second-order)
- number of references to the patent literature
- number of references to the non-patent literature
- composition of citations (A, X, Y, self)
- composition of references (A, X, Y, self)
- *lots of control variables (incl. patent holder type, not considered here)*

Multivariate Analysis

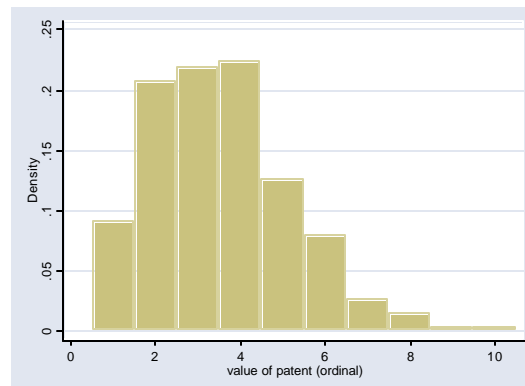
Value equation

Variable	patent value - German PATVAL study (ordered probit, N=3,350)
references	n.s.
X share	n.s.
Y share	n.s.
share self references	n.s.
citations	0.0482 (3.74)
X share	n.s.
Y share	n.s.
share self citations	n.s.
2nd order citations	0.0044 (2.52)
NPL references	n.s.
control variables	technical field, year dummies, add. variables
pseudo R squared	0.0062

Source: Harhoff/Hoisl (2004) – marginal effects (t-stats) authors' computations based on EPO/OECD citation database.

Multivariate Analysis

Value equation



Source: Harhoff/Hoisl (2004) – based on German PATVAL data.

Multivariate Analysis

Opposition and Revocation Incidence

Source: Harhoff/Hoisl (2004) – marginal effects (t-stats) authors' computations based on EPO/OECD citation database.

Variable	opposition incidence (binary probit) N=594,647	revocation incidence (binary probit) N=27,530
references	0.0004 (3.5)	n.s.
X share	0.0270 (22.9)	0.0411 (3.5)
Y share	0.0088 (6.8)	n.s.
share self references	0.0394 (14.4)	n.s.
citations	0.0070 (60.9)	0.0066 (7.3)
X share	0.0137 (8.7)	n.s.
Y share	0.0125 (6.5)	n.s.
share self citations	0.0128 (6.6)	0.0463 (2.4)
2nd order citations	0.0003 (16.0)	n.s.
NPL references	-0.0009 (3.3)	n.s.
control variables	technical field, year dummies, add. variables	dto. plus additional variables
pseudo-R-squared	0.0693	0.0269

Summary of Results

- value equation
 - only citation counts matter, composition not relevant
 - 2nd order citations appear to be relevant
- opposition incidence equation
 - not number, but composition of references matters considerably
 - number of citations proxies for value
 - X-citations indicate anticipated interaction between opponents
 - 2nd-order citations have some predictive power
 - self-citations and references have a strong negative effect on opposition incidence – appear to indicate idiosyncratic research paths
- revocation incidence equation
 - weak predictive power (as should be!)
 - residual information still in X-references and 1st-order citations

Caveats and More Plans

- Caveat: these are preliminary results.
- Results for 2nd-order citations warrant attention (and more experimentation).
- Results for self-referencing surprisingly strong – should be extended using a complete coding of patent holders (for non-EP references).
- Still far cry from a structural model of value, legal robustness and other latent variables – but some progress.
- Inclusion of more refined NPL indicators to commence shortly.