Research using the *USPTO's Trademark Case File Dataset*

Stuart Graham

Georgia Tech & USPTO

6th Conference on Innovation and Entrepreneurship
Searle Center, Northwestern University
6 June 2013

The USPTO Trademark Case Files Dataset:

Descriptions, Lessons, and Insights

Stuart Graham, Galen Hancock, Alan Marco, Amanda Fila Myers

United States Patent and Trademark Office
An agency of the Department of Commerce
Alexandria, VA, USA

Available at: http://ssrn.com/abstract=2188621

Motivation

- Data as precondition / facilitator of research
  - E.g., NBER Patent Citation Database, HJT, (2001)
- Support for embryonic field of TM study
  - Pillar of the USPTO economics research agenda
- Responsibility of government
  - Transparency
  - Data as a public good = underprovided
- Obama Administration information initiatives
  - E.g., data.gov umbrella

Brief Backgrounder

- Trademark
  - "a word, phrase, symbol, design, color, smell, sound, or combination thereof that identifies and distinguishes the goods and services of one party from those of others." 15 U.S.C. § 1127.
- Bases
  - State common law
  - US federal register
    - Since 1881, modern Lanham Act dates from 1946
    - Benefit of prima facie evidence of ownership and exclusive right to use the mark on goods specified
- Represents the goodwill an enterprise and its goods and services maintain with the public.
  - Consumers rely on trademarks to reduce search costs, to distinguish among competing producers, and as guarantees of product quality.

Prior, ongoing research using TM data:
Where's the action

- Economics / Management
  - Graevenitz / LMU
  - Complements
- NGOs
  - OECD, Millot
- Empirical legal studies
  - Beebe - NYU
- Marketing
  - Tucker – MIT / Krasnikov – GWU
- IP Offices
  - WIPO, Fink / OHIM, Wajsman / UK, Clayton / USPTO

Google

USPTO Bulk Downloads: Trademarks

The following USPTO trademark products are available for free download.

Trademark Registration Images (1870 – present)

Trademark Application Text (1884 – present)

Trademark Assignment Text (1980 – present)

Trademark Trial and Appeal Board Text (1955 – present)

Trademark Application Images (2010 – present)

---

www.uspto.gov/ip/officechiefecon/tm_casefiles.jsp

uspto.gov

search for patents | search for trademarks

Search our site →

The United States Patent and Trademark Office
an agency of the Department of Commerce

PATENTS | TRADEMARKS | IP LAW & POLICY | PRODUCTS & SERVICES | INVENTORS | NEWS & NOTICES | FAQs | ABOUT US

Home Page » IP LAW & POLICY » Office of Chief Economist

Rulemaking

IP Policy and Enforcement

Protecting IP Overseas

Training and Education

Training Programs & Conferences

Office of Governmental Affairs

Office of Chief Economist

Boards and Counsel

Trademark Case Files

The Trademark Case Files Dataset contains detailed information on 6.7 million trademark applications filed with or registrations issued by the USPTO between January 1870 and January 2012. It is derived from the USPTO main database for administering trademarks and includes data on mark characteristics, prosecution events, ownership, classification, third-party oppositions, and renewal history.

A document describing these data is available here: "The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights."

This document is also available as: Graham, S., Hancock, G., Marco, A. and Myers, A. (2013). "The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights." SSRN working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2188621.

The OCE developed these data files for public use, and encourage users to identify fixes and improvements.

Please provide all feedback to: EconomicsData@uspto.gov

Documentation

Trademark Case Files Dataset Documentation

- Appendix I: Variable Tables
- Appendix II: Event description [.dta format] [.csv format]
- Appendix III: Madrid Event description [.dta format] [.csv format]

● Appendix II: Event description [.dta format] [.csv format]
● Appendix III: Madrid Event description [.dta format] [.csv format]

Data Files

A map of all data files can be found here (same as Appendix I).

Download full set of data files [.dta format] [.csv format]

Download individual data files:

● case_file [.dta format] [.csv format]
● correspondent_domrep_attorney [.dta format] [.csv format]
  ● event [.dta format] [.csv format]
  ● owner [.dta format] [.csv format]
    ● owner_name_change [.dta format] [.csv format]
  ● classification [.dta format] [.csv format]
    ● intl_class [.dta format] [.csv format]
    ● us_class [.dta format] [.csv format]
  ● statement [.dta format] [.csv format]
  ● design_search [.dta format] [.csv format]
  ● foreign_app [.dta format] [.csv format]
  ● prior_mark [.dta format] [.csv format]
  ● madrid_intl_file [.dta format] [.csv format]
    ● madrid_event [.dta format] [.csv format]

Additional Resources

● Trademark Applications Documentation (TAD)
● Trademark Manuals, Guides, Official Gazette

» Accessibility
» Privacy Policy
» Terms of Use
» Security
» Emergencies/Security Alerts
» Information Quality Guidelines

» Federal Activities Inventory Reform (FAIR) Act
» Notification and Federal Employee Antidiscrimination and Retaliation (NoFEAR) Act
» Budget & Performance
» Freedom of Information Act (FOIA)

» Department of Commerce NoFEAR Act Report
» Regulations.gov
» STOP!Fakes.gov
» Department of Commerce
» USA.gov

» Strategy Targeting Organized Piracy (STOP!)
» Careers
» Site Index
» Contact Us
» USPTO Webmaster

This page is owned by Office of the Chief Economist.

Last Modified: 2/4/2013 8:55:40 AM

Figure 5: Trademark Case Files Dataset organizational structure

event	
obs: 79,226,385	
var: 5	

	owner_name_change
owner	obs: 507,246
unique: own_id	var: 3
obs: 13,813,383	
var: 25	

	intl_class
classification	obs: 7,318,747
unique: class_id	var: 3
obs: 7,238,351	
var: 11	

	us_class
case_file	obs: 23,421,483
unique: serial_no	var: 3
obs: 6,707,708	
var: 80	

statement
obs: 11,779,403
var: 4

correspondent_domrep_attorney
unique: serial_no
obs: 6,707,708
var: 11

design_search
obs: 3,721,378
var: 2

foreign_app
obs: 316,702
var: 13

prior_mark
obs: 1,627,902
var: 4

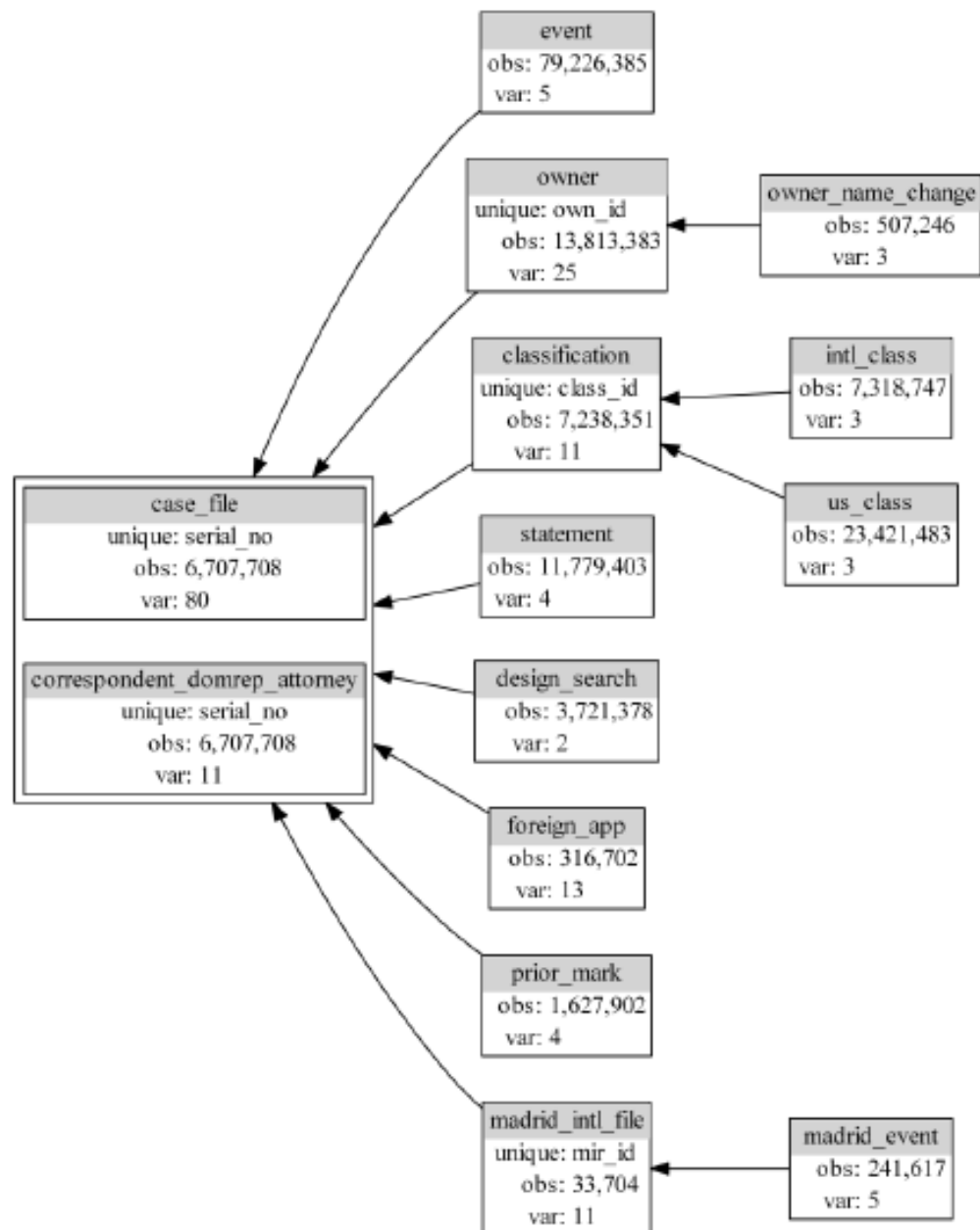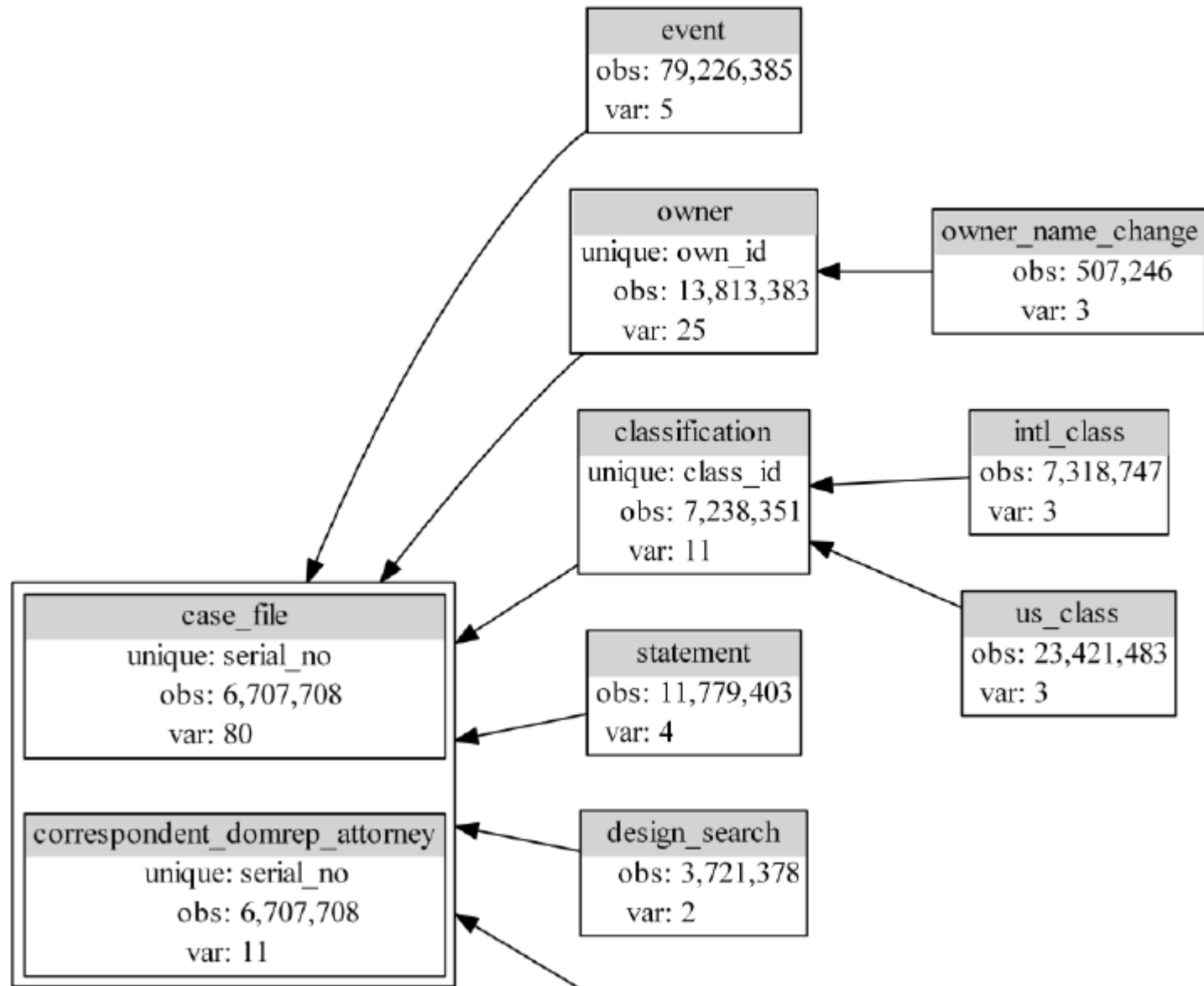|	**madrid_event**
madrid_intl_file	obs: 241,617
unique: mir_id	var: 5
obs: 33,704	
var: 11	

Figure 5: Trademark Case Files Dataset organizational structure

```
┌─────────────────────────────────┐
│          case_file              │
│       unique: serial_no         │                    ┌──────────────────┐
│       obs: 6,707,708            │        statement   │    us_class      │
│          var: 80                │     obs: 11,779,403│  obs: 23,421,483 │
├─────────────────────────────────┤        var: 4     │     var: 3       │
│ correspondent_domrep_attorney   │                    └──────────────────┘
│       unique: serial_no         │
│       obs: 6,707,708            │     design_search
│          var: 11                │     obs: 3,721,378
└─────────────────────────────────┘        var: 2

                                          foreign_app
                                          obs: 316,702
                                            var: 13

                                           prior_mark
                                          obs: 1,627,902
                                             var: 4

                                          madrid_intl_file        madrid_event
                                          unique: mir_id          obs: 241,617
                                           obs: 33,704              var: 5
                                            var: 11
```

case_file
unique: serial_no
obs: 6,707,708
var: 80

correspondent_domrep_attorney
unique: serial_no
obs: 6,707,708
var: 11

statement
obs: 11,779,403
var: 4

us_class
obs: 23,421,483
var: 3

design_search
obs: 3,721,378
var: 2

foreign_app
obs: 316,702
var: 13

prior_mark
obs: 1,627,902
var: 4

madrid_intl_file
unique: mir_id
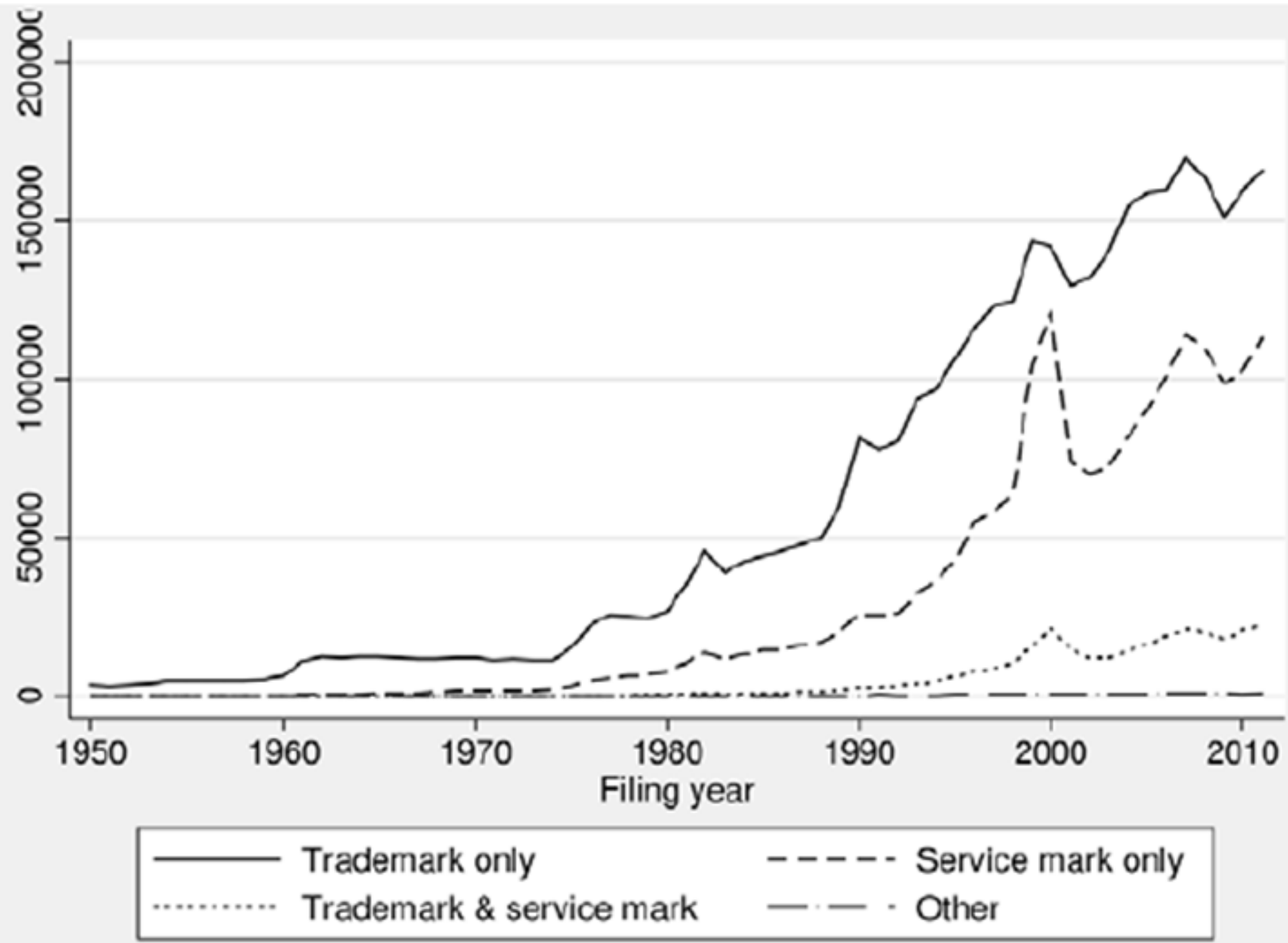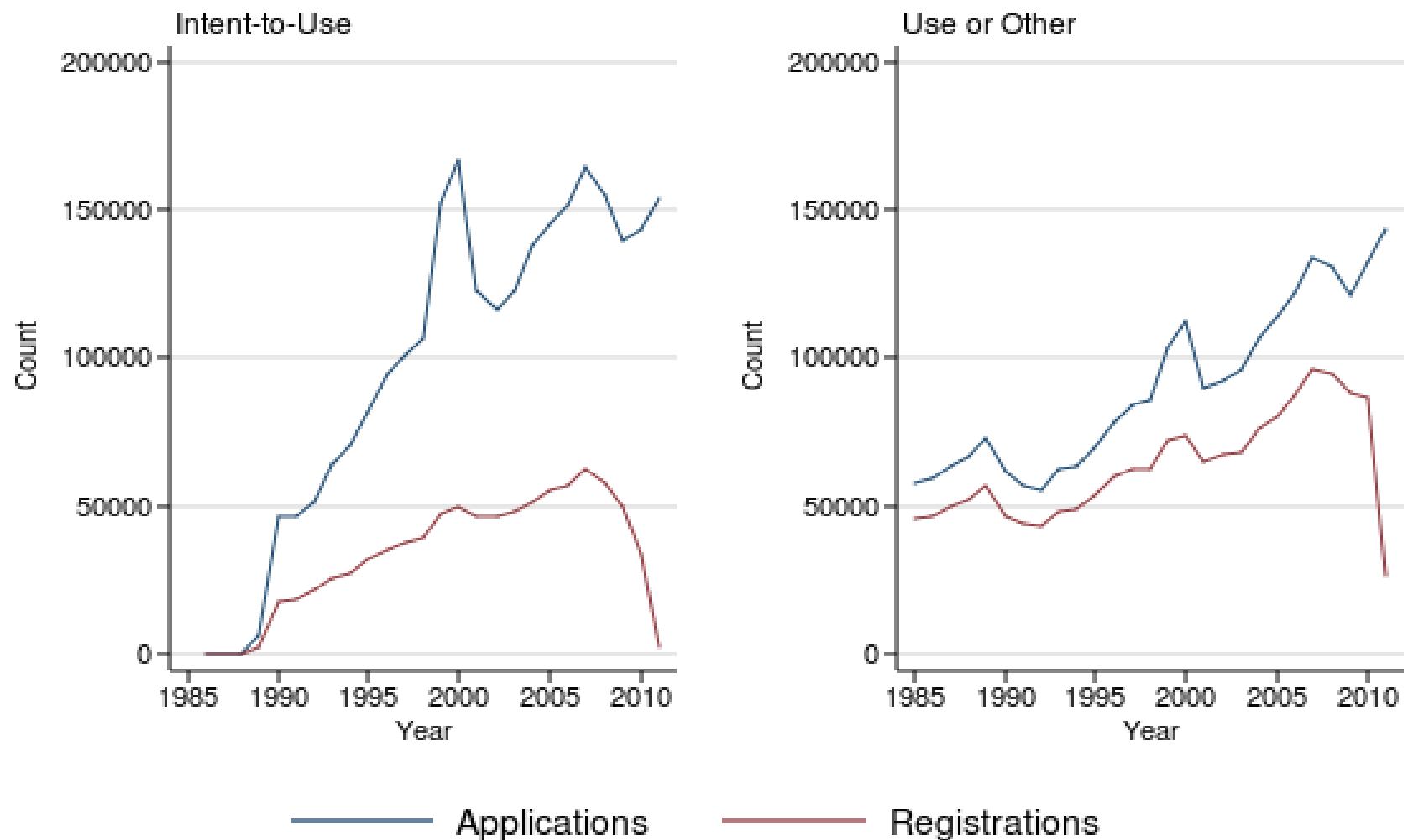obs: 33,704
var: 11

madrid_event
obs: 241,617
var: 5

Applications and Registrations by Year of Filing at USPTO

Mark type by filing year

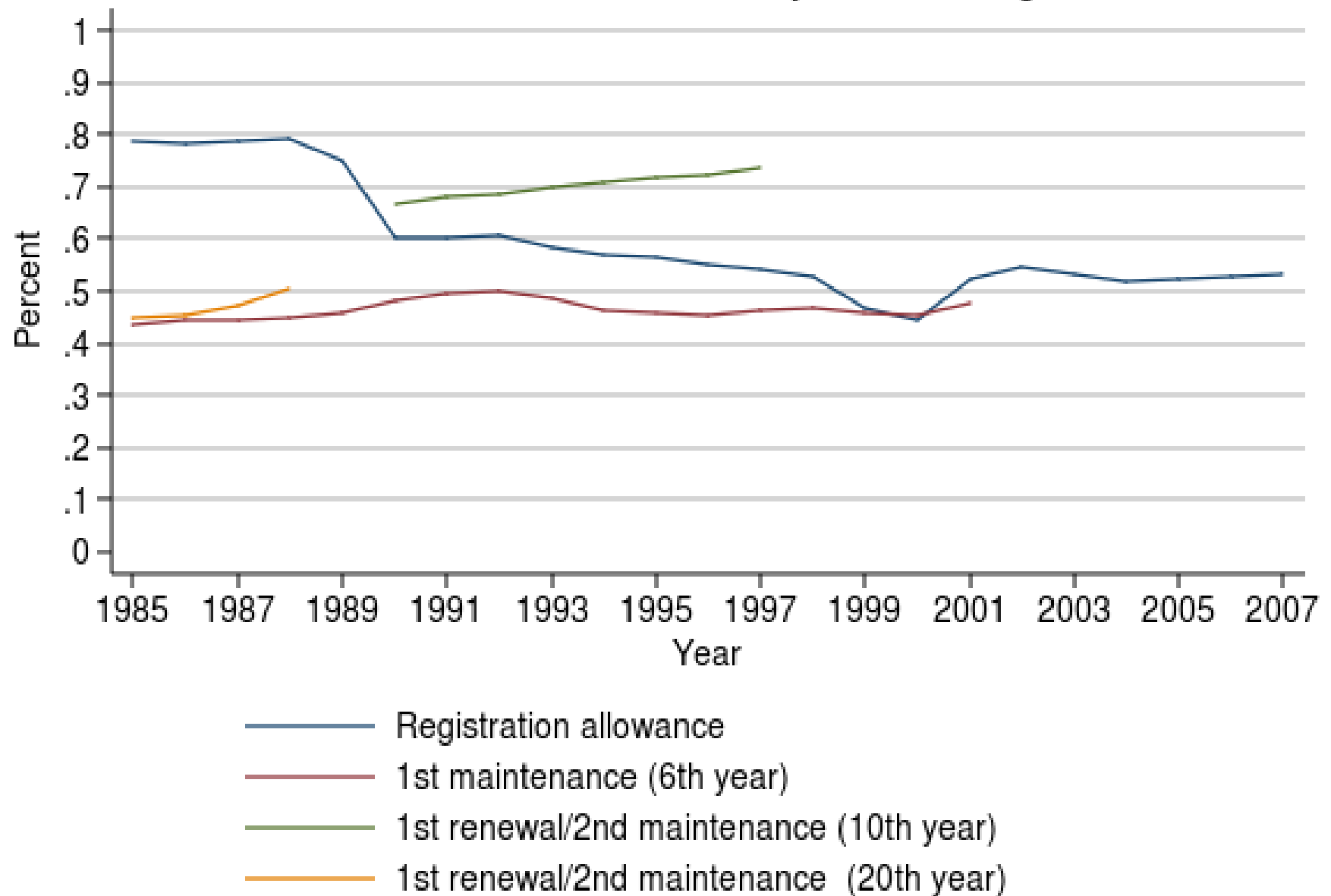# Applications and Registrations by Year of Filing at USPTO

Intent-to-Use | Use or Other

Applications ————— Registrations

Use or other includes all applications not filed solely on intended use legal basis

Allowance, Maintenance, and Renewal by Year of Filing at USPTO

Legend:
- Registration allowance
- 1st maintenance (6th year)
- 1st renewal/2nd maintenance (10th year)
- 1st renewal/2nd maintenance (20th year)

Observations per unique *serial number* in data file

Data file	(1) count	(2) mean	(3) sd	(4) median
		per *serial number*[1]		
event	79,226,385	13.475	9.075	12
owner	13,813,383	2.363	1.109	3
classification	7,238,351	1.237	0.833	1
statement	11,779,403	2.024	1.272	2
design_search	3,721,378	2.949	2.016	2
prior_mark	1,627,902	2.101	2.260	2

1) Calculated only for unique *serial numbers* observable in each data file

Registrations by owner legal type and registration year
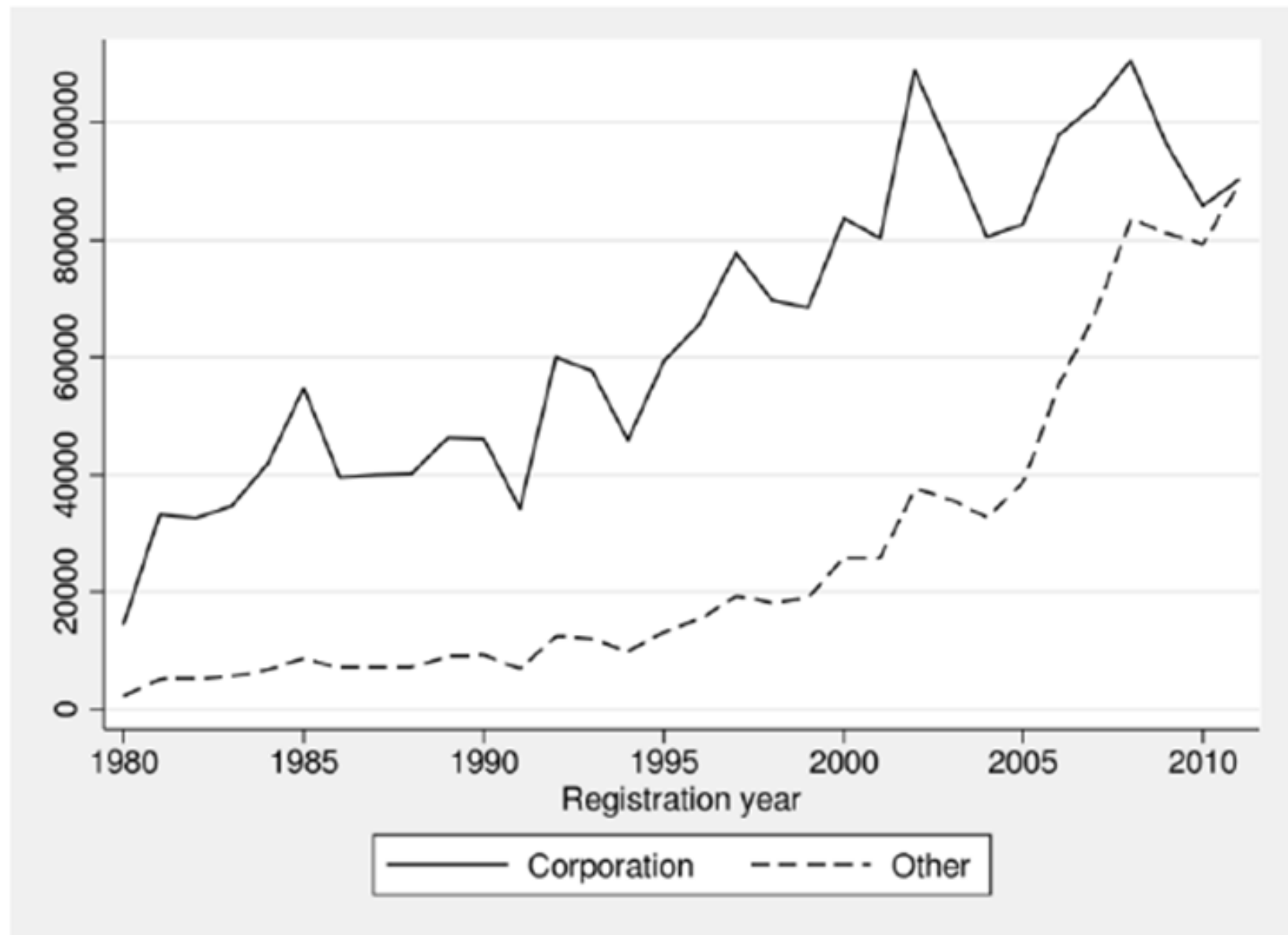
**Registrations by (non-corporation) owner legal type and registration year**
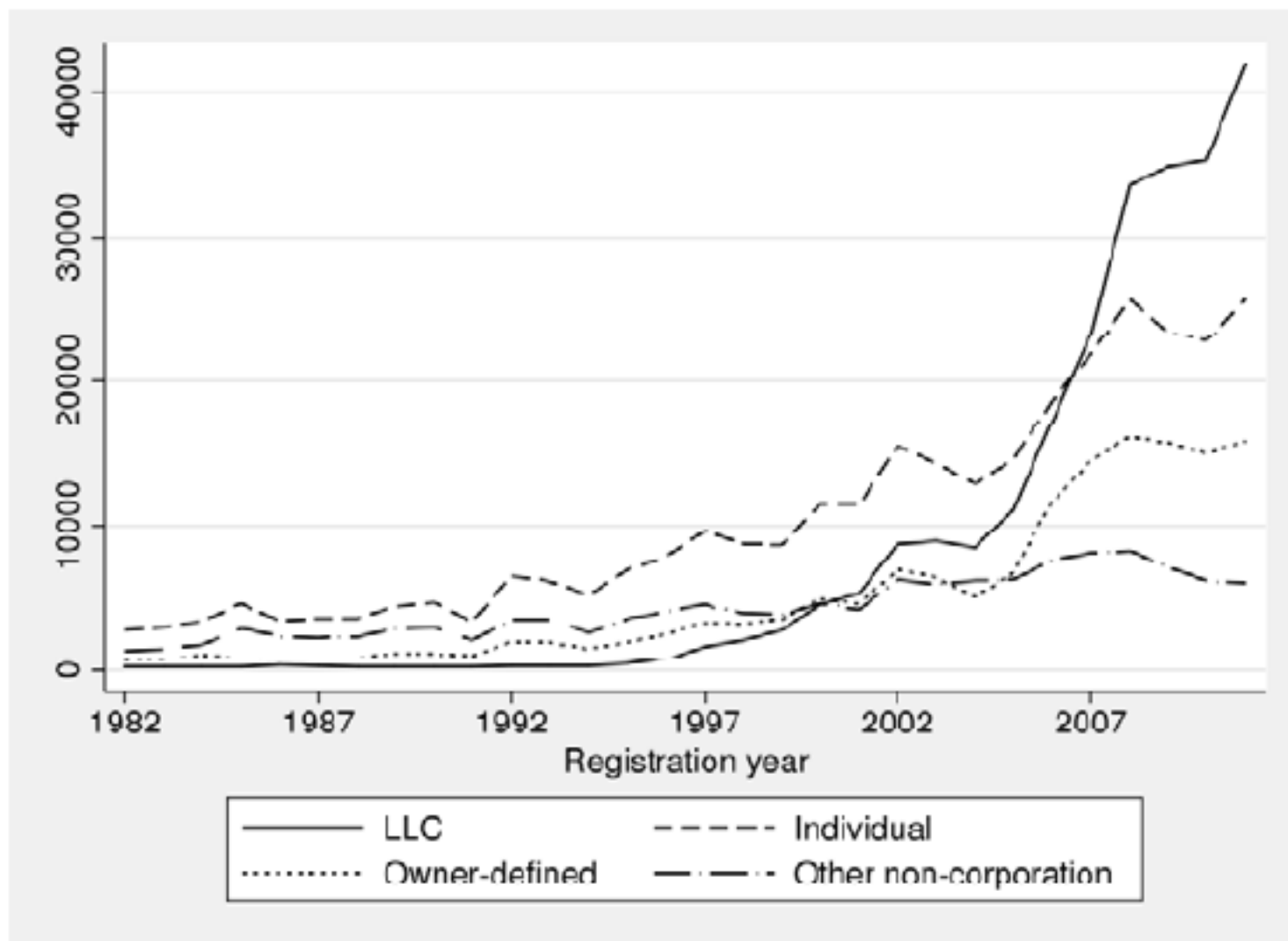
**Observations in design_search data file by *design search code* Category**

	(1)	(2)
	percent	
	design search codes	*serial numbers*
Category 26 Geometric figures and solids	0.470	0.663
Division 26.01 Circles	0.119	0.206
Division 26.11 Rectangles	0.070	0.206
Division 26.17 Lines, bands, bars	0.068	0.201
Category 2 Human beings	0.097	0.148
Category 3 Animals	0.073	0.116
Category 1 Celestial bodies, natural phenomena, geographical maps	0.069	0.159
Category 24 Heraldry, flags, crowns, crosses, arrows and symbols	0.057	0.130
Category 5 Plants	0.039	0.087
Category 27 Forms of writing	0.029	0.081
Category 7 Dwellings, buildings, monuments, stadiums, fountains, structural works and building materials	0.020	0.046
Category 6 Scenery	0.017	0.041
Category 4 Supernatural beings, mythological or legendary beings, fantastical beings or unidentifiable beings	0.013	0.032
Other	0.017	0.032
Observations	3,721,378	1,262,050

1) Calculated only for unique *serial numbers* observable in the design_search data file

Registrations age when cited
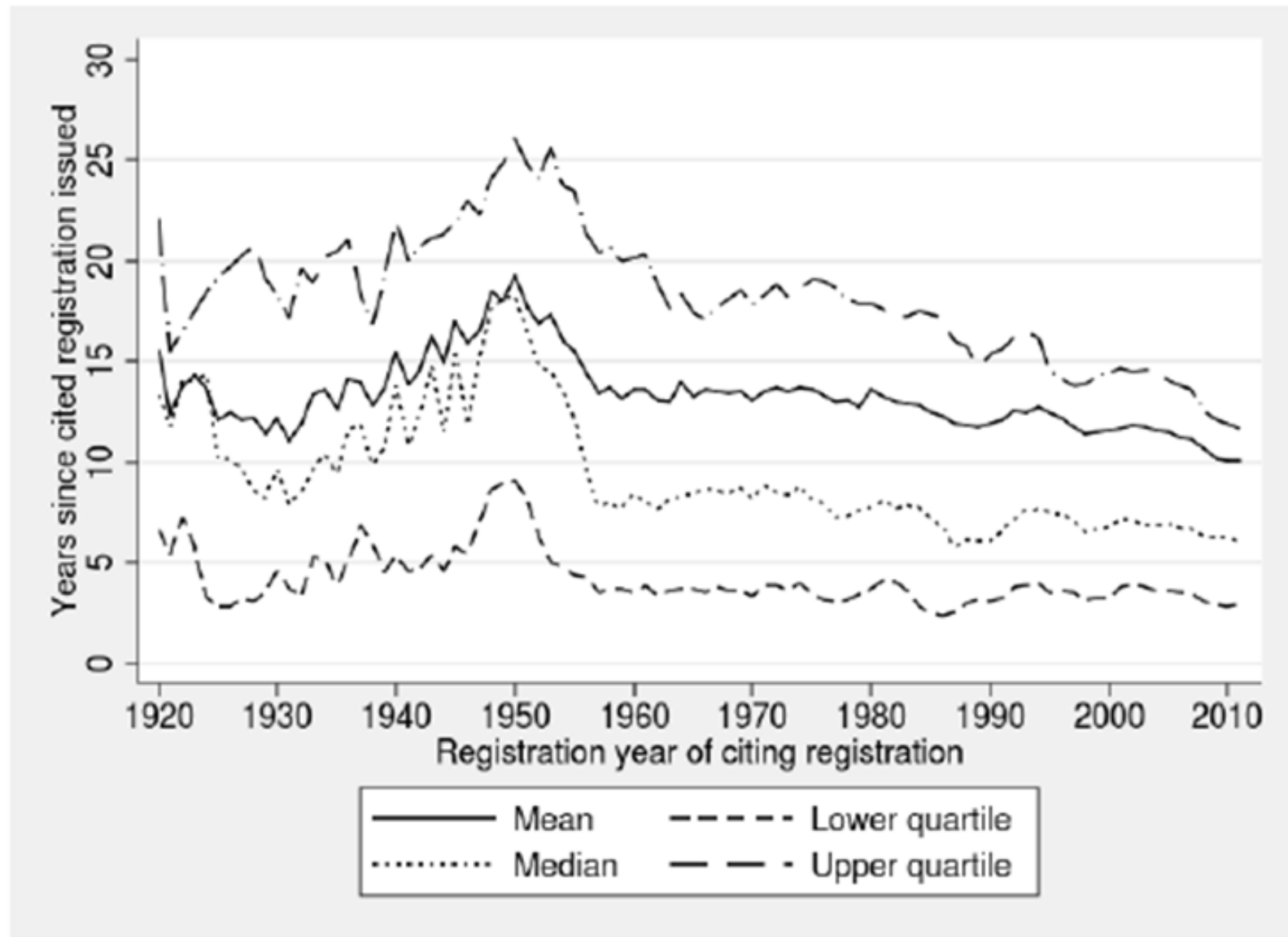
Live registration or pending application count by network size
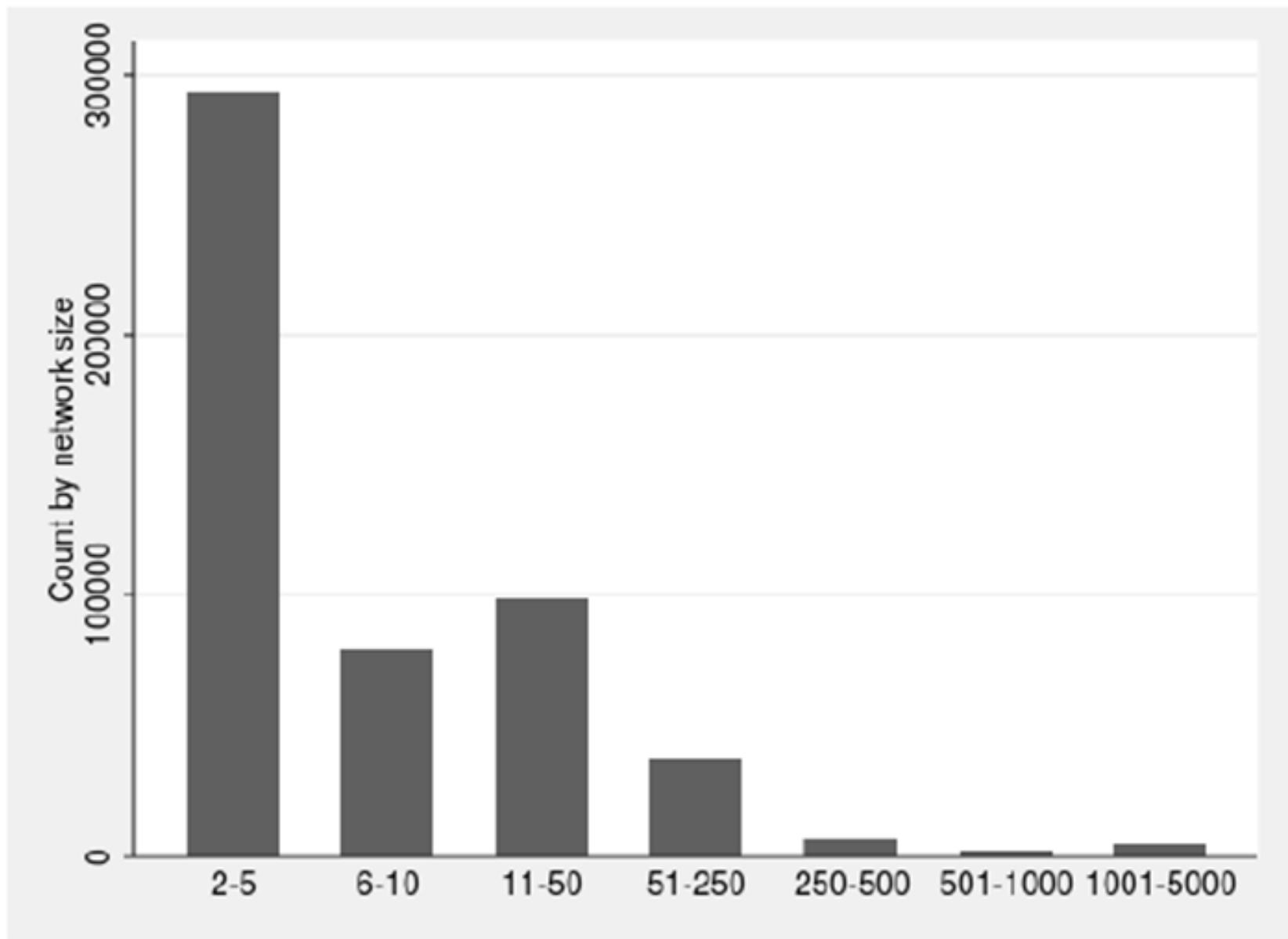
Next Steps

- Hard work to be done on validation
  - TM, unlike the patent literature, has not had decades of rigorous work
    - Patent data had serious minds tackling important questions prior to the arrival of "big data," e.g., Grilches, Pakkes, Scherer

- TM data is largely terra incognito
  - Payoffs likely in slow, molecular effort by researchers in the vein of "normal science"

Engelberg Center on Innovation Law and Policy, New York University School of Law
and the
United States Patent and Trademark Office

in cooperation with

Center for Law & Economics, ETH Zurich
Oxford Intellectual Property Research Centre
University of East Anglia

We are seeking paper proposals from economics, management, and legal scholars on the empirical study of trademark data for a one-and-a-half day workshop at the U.S. Patent and Trademark Office in Alexandria, Virginia to occur on September 26-27, 2013. This will constitute the companion workshop to our previous workshop at the University of Oxford in December 2012. Like its Oxford counterpart, this workshop is part of a broader project to support better scholarship in this embryonic area of research and lead to the publication of high quality and high impact studies. With the increased availability of trademark data, we are soliciting papers that open up new avenues of trademark research. Papers with empirical or experimental data and papers with a U.S., international, or foreign focus are welcome.

Proposals should be submitted by July 1, 2013, to saurabh.vishnubhakat@uspto.gov.

Authors will be notified of acceptance by July 15, 2013.

Please direct any questions to Barton Beebe (barton.beebe@nyu.edu) or Alan Marco (alan.marco@uspto.gov).