

IMPROVING PATENT ASSIGNEE-FIRM BRIDGE WITH WEB SEARCH RESULTS

Yuheng Ding
(World Bank)

Karam Jo
(Korea Development Institute)

Seula Kim
(Princeton University)

NBER Innovation Information Initiative Technical Working Group Meeting - Fall 2023

December 2, 2023

DISCLAIMER: Any views expressed are those of the authors and not those of the Korea Development Institute or the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2095. (CBDRB-FY22-P2095-R9872)

MOTIVATION

- Firm innovation is a major source of creative destruction and economic growth

MOTIVATION

- Firm innovation is a major source of creative destruction and economic growth
- Patents have long been acknowledged as a rich source of data for firm innovation

MOTIVATION

- Firm innovation is a major source of creative destruction and economic growth
- Patents have long been acknowledged as a rich source of data for firm innovation
- ↪ However, linking the granted patents to the owning firms is nontrivial
 - No unique assignee identifiers
 - No consistent format for assignee names and addresses

MOTIVATION

- Firm innovation is a major source of creative destruction and economic growth
- Patents have long been acknowledged as a rich source of data for firm innovation
- ↳ However, linking the granted patents to the owning firms is nontrivial
 - No unique assignee identifiers
 - No consistent format for assignee names and addresses
- Prior efforts to overcome this issue by linking USPTO patent data to firm-level data
(Hall et al. 2001; Kerr and Fu 2008; Balasubramanian and Sivadasan 2010, 2011; Graham et al. 2018; Dreisigmeyer et al. 2018; Autor et al. 2020; Arora et al. 2021)

MOTIVATION

- Firm innovation is a major source of creative destruction and economic growth
- Patents have long been acknowledged as a rich source of data for firm innovation
- ↪ However, linking the granted patents to the owning firms is nontrivial
 - No unique assignee identifiers
 - No consistent format for assignee names and addresses
- Prior efforts to overcome this issue by linking USPTO patent data to firm-level data
(Hall et al. 2001; Kerr and Fu 2008; Balasubramanian and Sivadasan 2010, 2011; Graham et al. 2018; Dreisigmeyer et al. 2018; Autor et al. 2020; Arora et al. 2021)
- ↪ However, existing crosswalks still contain pitfalls
 - Discontinuity in sample period: either only years before or after 2000
 - Publicly listed firms only: miss firms not reported in publicly available data

THIS PAPER

- Constructs **a new bridge** between USPTO patent data and administrative firm-level data

THIS PAPER

- Constructs a **new bridge** between USPTO patent data and administrative firm-level data, which
 - 1 Obtains longitudinal consistency and covers the **longest** period (amongst USPTO-LBD, 1976-2016)
 - 2 Contains the **population** of U.S. patenting firms
 - 3 Improves **match rate** (**7pp** and **2.9pp** at the patent and assignee level)

THIS PAPER

- Constructs a **new bridge** between USPTO patent data and administrative firm-level data, which
 - 1 Obtains longitudinal consistency and covers the **longest** period (amongst USPTO-LBD, 1976-2016)
 - 2 Contains the **population** of U.S. patenting firms
 - 3 Improves **match rate** (7pp and 2.9pp at the patent and assignee level)
- Builds on earlier approaches by introducing **internet search-aided algorithm** in Autor et al. 2020

THIS PAPER

- Constructs a **new bridge** between USPTO patent data and administrative firm-level data, which
 - 1 Obtains longitudinal consistency and covers the **longest** period (amongst USPTO-LBD, 1976-2016)
 - 2 Contains the **population** of U.S. patenting firms
 - 3 Improves **match rate** (7pp and 2.9pp at the patent and assignee level)
- Builds on earlier approaches by introducing **internet search-aided algorithm** in Autor et al. 2020
- Brings in potential benefits to researchers
 - Allow analysis of firm innovation over a long period of time based on consistent linking algorithms
 - Useful for studies on firm innovation activities by small or young firms (or entrepreneurship)

TABLE OF CONTENTS

- 1 Data
- 2 Matching Methodology
- 3 Match Results
- 4 Benefits and Practical Applications of the Bridge
- 5 Concluding Remarks

DATA SOURCES

- 1 USPTO PatentsView database
- 2 Business Register (**BR**)
- 3 Longitudinal Business Database (**LBD**)

DATA SOURCES

1 USPTO PatentsView database

- Tracks all patents granted by the USPTO from 1976 onward
- Detailed information for patenting activities
(application/grant dates, technology class, patent citation, and the name and address of patent assignees, etc.)

2 Business Register (BR)

3 Longitudinal Business Database (LBD)

DATA SOURCES

1 USPTO PatentsView database

- Tracks all patents granted by the USPTO from 1976 onward
- Detailed information for patenting activities
(application/grant dates, technology class, patent citation, and the name and address of patent assignees, etc.)

2 Business Register (BR)

- A comprehensive database of the U.S. business establishments with paid employees
- Longitudinal business demographics and characteristics about establishment
(establishment/parent firm identifiers, name, address, and single/multi-unit indicator, etc.)

3 Longitudinal Business Database (LBD)

DATA SOURCES

1 USPTO PatentsView database

- Tracks all patents granted by the USPTO from 1976 onward
- Detailed information for patenting activities
(application/grant dates, technology class, patent citation, and the name and address of patent assignees, etc.)

2 Business Register (BR)

- A comprehensive database of the U.S. business establishments with paid employees
- Longitudinal business demographics and characteristics about establishment
(establishment/parent firm identifiers, name, address, and single/multi-unit indicator, etc.)

3 Longitudinal Business Database (LBD)

- The universe of private non-farm establishments and firms with at least one paid employee in the U.S.
- Detailed firm-level information
(employment, payroll, industry codes, establishment and firm identifiers, etc.)

DATA SOURCES

1 USPTO PatentsView database

- Tracks all patents granted by the USPTO from 1976 onward
- Detailed information for patenting activities
(application/grant dates, technology class, patent citation, and the name and address of patent assignees, etc.)

2 Business Register (BR)

- A comprehensive database of the U.S. business establishments with paid employees
- Longitudinal business demographics and characteristics about establishment
(establishment/parent firm identifiers, name, address, and single/multi-unit indicator, etc.)

3 Longitudinal Business Database (LBD)

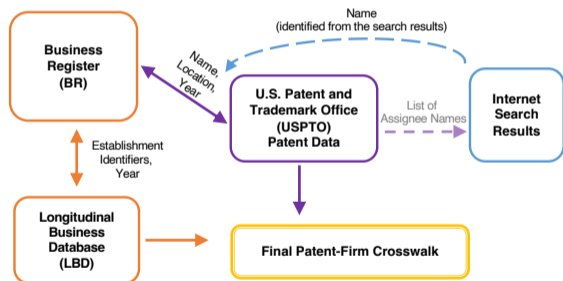
- The universe of private non-farm establishments and firms with at least one paid employee in the U.S.
- Detailed firm-level information
(employment, payroll, industry codes, establishment and firm identifiers, etc.)

⇒ **USPTO+BR**: matching algorithms; **BR+LBD**: firm-establishment identifiers

TABLE OF CONTENTS

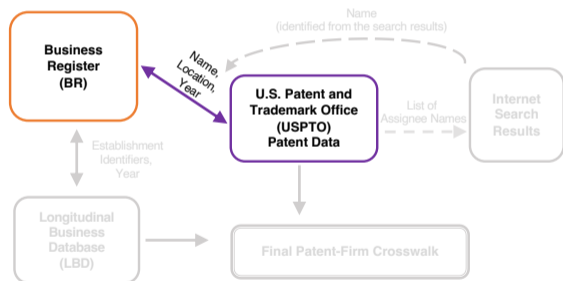
- 1 Data
- 2 Matching Methodology
- 3 Match Results
- 4 Benefits and Practical Applications of the Bridge
- 5 Concluding Remarks

OVERVIEW OF MATCHING PROCESS



- 1 Name standardization (NBER PDP)
- 2 Strict and fuzzy name matching
- 3 Identify LBD firms (via identifiers)
- 4 Internet search-aided algorithm (Autor et al. 2020)
- 5 Stem name matching
& 2nd-round internet search-aided

STEP 1,2: NAME STANDARDIZATION AND MATCHING



- 1 Name standardization (NBER PDP)
- 2 Strict and fuzzy name matching
- 3 Identify LBD firms (via identifiers)
- 4 Internet search-aided algorithm (Autor et al. 2020)
- 5 Stem name matching
& 2nd-round internet search-aided

STEP 1,2: NAME STANDARDIZATION AND MATCHING

- Assignees/estab. are same entities w/ identical names and location (city, state) in a given year

STEP 1,2: NAME STANDARDIZATION AND MATCHING

- Assignees/estab. are same entities w/ identical names and location (city, state) in a given year
- Name standardization generates std. and stem names (NBER PDP)

e.g. International Business Machines Corporation $\xrightarrow{\text{standardized name}}$ INT BUSINESS MACHINES CORP
 $\xrightarrow{\text{stem name}}$ INT BUSINESS MACHINES

- Strict and fuzzy name matching by assignee/estab. name and address in a given year
 - Use the SAS DQMatch fuzzy matching procedure

e.g. INT BUNIESS MACJINES CORP $\xrightarrow{\text{fuzzy matching}}$ INT BUSINESS MACHINES CORP

- Year is application or grant year; implement a +3/-3 year window
- Start w/ the most restrictive criteria and gradually allow "fuzziness": only keep reliable matches

▶ More

STEP 1,2: NAME STANDARDIZATION AND MATCHING

- Assignees/estab. are same entities w/ identical names and location (city, state) in a given year
- Name standardization generates std. and stem names (NBER PDP)

e.g. International Business Machines Corporation $\xrightarrow{\text{standardized name}}$ INT BUSINESS MACHINES CORP
 $\xrightarrow{\text{stem name}}$ INT BUSINESS MACHINES

- Strict and fuzzy name matching by assignee/estab. name and address in a given year
 - Use the SAS DQMatch fuzzy matching procedure

e.g. INT BUNIESS MACJINES CORP $\xrightarrow{\text{fuzzy matching}}$ INT BUSINESS MACHINES CORP

- Year is application or grant year; implement a +3/-3 year window
- Start w/ the most restrictive criteria and gradually allow "fuzziness": only keep reliable matches

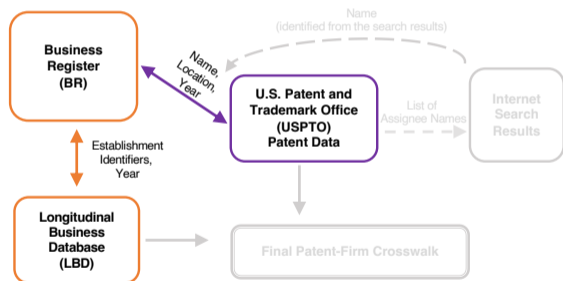
[More](#)

⇒ ★ e.g. INT BUSINESS MACHINES CORP (in USPTO)
(in Armonk, NY)

$\xleftrightarrow{\text{name, address matching}}$

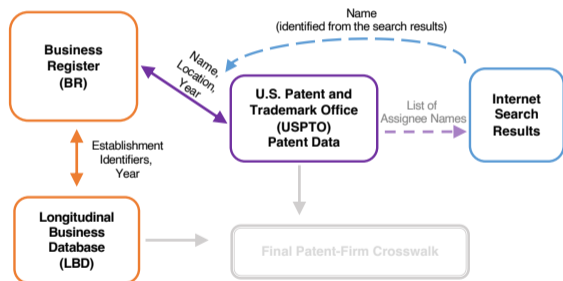
INT BUSIESS MACHINES CORP (in BR)
(in Armonk, NY)

STEP 3: IDENTIFYING LBD FIRMS



- 1 Name standardization (NBER PDP)
- 2 Strict and fuzzy name matching
- 3 Identify LBD firms (via identifiers) [▶ More](#)
- 4 Internet search-aided algorithm (Autor et al. 2020)
- 5 Stem name matching
& 2nd-round internet search-aided

STEP 4: INTERNET SEARCH-AIDED ALGORITHM



- 1 Name standardization (NBER PDP)
- 2 Strict and fuzzy name matching
- 3 Identify LBD firms (via identifiers) [▶ More](#)
- 4 Internet search-aided algorithm (Autor et al. 2020)
- 5 Stem name matching
& 2nd-round internet search-aided

STEP 4: INTERNET SEARCH-AIDED ALGORITHM

- Following Autor et al. (2020), use the machine-learning capacities of internet search engine
 - Overcome unresolved abbreviation/misspelling issues and improve match rate
 - Apply it to USPTO assignees only

STEP 4: INTERNET SEARCH-AIDED ALGORITHM

- Following Autor et al. (2020), use the machine-learning capacities of internet search engine
 - Overcome unresolved abbreviation/misspelling issues and improve match rate
 - Apply it to USPTO assignees only
- Extract top 5 search results for patent assignee names → identify pairs sharing ≥ 2 results
 - Put every patent assignee name into the Google.com & collect the URLs of the top five search results

STEP 4: INTERNET SEARCH-AIDED ALGORITHM

- Following Autor et al. (2020), use the machine-learning capacities of internet search engine
 - Overcome unresolved abbreviation/misspelling issues and improve match rate
 - Apply it to USPTO assignees only
- Extract top 5 search results for patent assignee names → identify pairs sharing ≥ 2 results
 - Put every patent assignee name into the Google.com & collect the URLs of the top five search results

e.g. “IBM CORP”:

target	matched using internet search results
IBM CORP	INT BUSINESS MACHINES CORP
IBM CORP	INT BUSINESS MAHINES CORP
IBM CORP	INT BUSINSS MACHINES CORP
IBM CORP	INT BUSINESS MACHINES CORPORARION

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk

e.g. IBM CORP $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk

e.g. IBM CORP $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP $\xleftrightarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk

e.g. IBM CORP $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP $\xleftrightarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)

(new) (reference)

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk: “search-aided bridge”

e.g. IBM CORP $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP $\xleftarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)
(new) (reference)

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk: “search-aided bridge”

e.g. IBM CORP $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP $\xleftarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)
(new) (reference)

- Use the search-aided bridge for those unmatched and dropped matches (from step 1-3)

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk: “search-aided bridge”

e.g. IBM CORP $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP $\xleftarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)
(new) (reference)

- Use the search-aided bridge for those unmatched and dropped matches (from step 1-3)
 - Rank new matches and keep reliable matches as before
(use the name of the reference firm, and the location, year of the new firm; and compare w/ BR establishment)

STEP 4: INTERNET SEARCH-AIDED ALGORITHM (CONT'D)

- Combine this with the previously constructed patent-firm crosswalk: “search-aided bridge”

e.g. IBM CORP (new) $\xrightarrow{\text{Internet search-aided}}$ INT BUSINESS MACHINES CORP (reference) $\xleftarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)

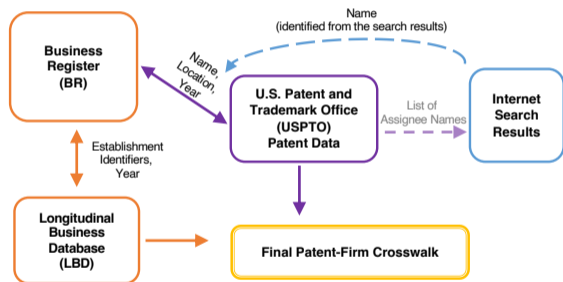
- Use the search-aided bridge for those unmatched and dropped matches (from step 1-3)
 - Rank new matches and keep reliable matches as before
(use the name of the reference firm, and the location, year of the new firm; and compare w/ BR establishment)

\Rightarrow ★ e.g. IBM CORP (new) (in Armonk, NY) $\xleftrightarrow{\text{Internet search-aided matching}}$ INT BUSINESS MACHINES CORP (in BR) (in Armonk, NY)

▶ Example

STEP 5: STEM NAME MATCHING & 2ND ROUND SEARCH-AIDED

▶ EXAMPLE



- 1 Name standardization (NBER PDP)
- 2 Strict and fuzzy name matching
- 3 Identify LBD firms (via identifiers) [▶ More](#)
- 4 Internet search-aided algorithm (Autor et al. 2020)
- 5 Stem name matching [▶ More](#)
& 2nd-round internet search-aided

TABLE OF CONTENTS

- 1 Data
- 2 Matching Methodology
- 3 Match Results
- 4 Benefits and Practical Applications of the Bridge
- 5 Concluding Remarks

MATCH RATES BY MODEL TYPES

Model	Patent Level	Assignee Level
Std. Name Matching (D+)	62	55.5
Stem Name Matching (D+)	14.8	12.8
Search-Aided Algorithm	7	2.9
Std./Stem Name Matching D- (No Search-Aided)	4.5	8.9
Overall	88.2	80.1

Table 1: Match Rates by Aggregate Model Types (%)

MATCH RATES BY MODEL TYPES

- Overall match rates: **88.2%** (patent level) and **80.1%** (assignee level)

Model	Patent Level	Assignee Level
Std. Name Matching (D+)	62	55.5
Stem Name Matching (D+)	14.8	12.8
Search-Aided Algorithm	7	2.9
Std./Stem Name Matching D- (No Search-Aided)	4.5	8.9
Overall	88.2	80.1

Table 2: Match Rates by Aggregate Model Types (%)

MATCH RATES BY MODEL TYPES

- Overall match rates: **88.2%** (patent level) and **80.1%** (assignee level)
 - Half of the matches are based on std. name matching

Model	Patent Level	Assignee Level
Std. Name Matching (D+)	62	55.5
Stem Name Matching (D+)	14.8	12.8
Search-Aided Algorithm	7	2.9
Std./Stem Name Matching D- (No Search-Aided)	4.5	8.9
Overall	88.2	80.1

Table 3: Match Rates by Aggregate Model Types (%)

MATCH RATES BY MODEL TYPES

- Overall match rates: **88.2%** (patent level) and **80.1%** (assignee level)
 - Half of the matches are based on std. name matching
 - Still, stem name matching and search-aided algorithm significantly improve the match rate

Model	Patent Level	Assignee Level
Std. Name Matching (D+)	62	55.5
Stem Name Matching (D+)	14.8	12.8
Search-Aided Algorithm	7	2.9
Std./Stem Name Matching D- (No Search-Aided)	4.5	8.9
Overall	88.2	80.1

Table 4: Match Rates by Aggregate Model Types (%)

MATCH RATES BY MODEL TYPES

- Overall match rates: **88.2%** (patent level) and **80.1%** (assignee level)
 - Half of the matches are based on std. name matching
 - Still, stem name matching and search-aided algorithm significantly improve the match rate
 - Improvement by search-aided accounts for **8.5%** and **4.1%** of the total patent and assignee level matches

Model	Patent Level	Assignee Level
Std. Name Matching (D+)	62	55.5
Stem Name Matching (D+)	14.8	12.8
Search-Aided Algorithm	7	2.9
Std./Stem Name Matching D- (No Search-Aided)	4.5	8.9
Overall	88.2	80.1

Table 5: Match Rates by Aggregate Model Types (%)

TABLE OF CONTENTS

- 1 Data
- 2 Matching Methodology
- 3 Match Results
- 4 Benefits and Practical Applications of the Bridge**
- 5 Concluding Remarks

BENEFITS OF THE BRIDGE

① **Match rate** has been improved

↪ Internet search-aided algorithm enhances the rate of matches by a non-negligible fraction

BENEFITS OF THE BRIDGE

- ① **Match rate** has been improved
 - ↪ Internet search-aided algorithm enhances the rate of matches by a non-negligible fraction
- ② **The longest longitudinal** patent assignee-firm bridge for administrative data (1976-2016)
 - ↪ Provides a stable bridge (based on consistent matching algorithms) over a longer horizon
 - ↪ Tracing firm innovation over an extended period is possible

BENEFITS OF THE BRIDGE

① **Match rate** has been improved

↳ Internet search-aided algorithm enhances the rate of matches by a non-negligible fraction

② **The longest longitudinal** patent assignee-firm bridge for administrative data (1976-2016)

↳ Provides a stable bridge (based on consistent matching algorithms) over a longer horizon

↳ Tracing firm innovation over an extended period is possible

③ Inclusion of **non-public** firms

↳ Studying firm innovation for small or young firms, not perfectly covered in public data, is available

REAL-WORLD APPLICATION OF THE BRIDGE

- E.g. **The impact of Chinese competition on firm innovation and business dynamism in the U.S.**
(Jo 2019; Jo and Kim 2021)

REAL-WORLD APPLICATION OF THE BRIDGE

- E.g. **The impact of Chinese competition on firm innovation and business dynamism in the U.S.**
(Jo 2019; Jo and Kim 2021)
 - Firm innovation
 - Citation-adjusted number of patent applications as a proxy for firm innovation
 - Self-citation ratio as the internal-ness of firm innovation (Akcigit and Kerr 2018)

REAL-WORLD APPLICATION OF THE BRIDGE

- E.g. **The impact of Chinese competition on firm innovation and business dynamism in the U.S.**
(Jo 2019; Jo and Kim 2021)
 - Firm innovation
 - Citation-adjusted number of patent applications as a proxy for firm innovation
 - Self-citation ratio as the internal-ness of firm innovation (Akcigit and Kerr 2018)
 - Foreign competition shock
 - Removal of uncertainty about U.S. trade policy toward China on imposed tariff rates after China's WTO accession in 2001 (Pierce and Schott 2016, Handley and Limão 2017)

$$\text{NTR Gap}_j = \underbrace{\text{Non-NTR Rate}_j}_{\text{for non-mkt econ., avg. 37\%}} - \underbrace{\text{NTR Rate}_j}_{\text{for WTO members, avg. 4\%}}$$

- * Use **NTR Gap_j** measured in 1999 (a year before the US gov. granted Permanent-NTR status to China)

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR Gap_{ijp0} + \beta_2 NTR Gap_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR\ Gap_{ijp0} + \beta_2 NTR\ Gap_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp}.$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR Gap_{ijp0} + \beta_2 NTR Gap_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)
- Δy_{ijp} : 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR\ Gap_{ijp0} + \beta_2 NTR\ Gap_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)
- Δy_{ijp} : 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- $NTR\ Gap_{ijp0}$: emp.-wgted avg. of 1999 industry-level NTR gaps across all the industries firm operates, measured in the start year for each period $p0$ (first diff.)
- $Post_p$: Dummy equal to one for $p = 2000-2007$, post treatment period (second diff.)

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR\ Gap_{ijp0} + \beta_2 NTR\ Gap_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)
- Δy_{ijp} : 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- $NTR\ Gap_{ijp0}$: emp.-wgted avg. of 1999 industry-level NTR gaps across all the industries firm operates, measured in the start year for each period $p0$ (first diff.)
- $Post_p$: Dummy equal to one for $p = 2000-2007$, post treatment period (second diff.)
- X_{ijp0}, X_{jp0} : covariates for firm size, age, tech. class-level past patenting trend, trade status, NTR rate; δ_j, δ_p : industry, time fixed effects

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR\ Gap_{ijp0} + \beta_2 NTR\ Gap_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)
- Δy_{ijp} : 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- $NTR\ Gap_{ijp0}$: emp.-wgted avg. of 1999 industry-level NTR gaps across all the industries firm operates, measured in the start year for each period $p0$ (first diff.)
- $Post_p$: Dummy equal to one for $p = 2000-2007$, post treatment period (second diff.)
- X_{ijp0}, X_{jp0} : covariates for firm size, age, tech. class-level past patenting trend, trade status, NTR rate; δ_j, δ_p : industry, time fixed effects

⇒ See the impact of **China's competition** on **different types of firm innovation**

▶ Table

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR\ Gap_{ijp0} \times InnovIntens_{ijp0} + \beta_2 Post_p \times NTR\ Gap_{ijp0} + \beta_3 Post_p \times InnovIntens_{ijp0} + \beta_4 NTR\ Gap_{ijp0} \times InnovIntens_{ijp0} + \beta_5 NTR\ Gap_{ijp0} + \beta_6 InnovIntens_{ijp0} + \mathbf{X}_{ijp0} \gamma_1 + \mathbf{X}_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)
- Δy_{ijp} : 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- $NTR\ Gap_{ijp0}$: emp.-wgted avg. of 1999 industry-level NTR gaps across all the industries firm operates, measured in the start year for each period p_0 (first diff.)
- $Post_p$: Dummy equal to one for $p = 2000-2007$, post treatment period (second diff.)
- $InnovIntens_{ijp0}$: avg. of past 5yr innovation intensity (nb. of patents / emp) measured in p_0
- $\mathbf{X}_{ijp0}, \mathbf{X}_{jp0}$: covariates for firm size, tech. class-level past patenting trend, trade status, NTR rate; δ_j, δ_p : industry, time fixed effects

⇒ Can interact w/ innovation intensity $InnovIntens_{ijp0}$ to see how the impact depends on

accumulated technological advantages of firms

Table

REAL-WORLD APPLICATION OF THE BRIDGE

- Diff-in-Diff regression to test the impact of China's competition (Pierce and Schott 2016)

$$\Delta y_{ijp} = \beta_1 Post_p \times NTR\ Gap_{ijp0} \times YoungFirm_{ijp0} + \beta_2 Post_p \times NTR\ Gap_{ijp0} + \beta_3 Post_p \times YoungFirm_{ijp0} + \beta_4 NTR\ Gap_{ijp0} \times YoungFirm_{ijp0} + \beta_5 NTR\ Gap_{ijp0} + \beta_6 YoungFirm_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp} .$$

- i : firm, j : main 6-digit NAICS industry firm i belongs to, p : time period (1992–1999, 2000–2007)
- Δy_{ijp} : 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- $NTR\ Gap_{ijp0}$: emp.-wgted avg. of 1999 industry-level NTR gaps across all the industries firm operates, measured in the start year for each period $p0$ (first diff.)
- $Post_p$: Dummy equal to one for $p = 2000–2007$, post treatment period (second diff.)
- $YoungFirm_{ijp0}$: young firm indicator measured in $p0$
- X_{ijp0}, X_{jp0} : covariates for firm size, tech. class-level past patenting trend, trade status, NTR rate; δ_j, δ_p : industry, time fixed effects

⇒ Can interact w/ young firm indicator $YoungFirm_{ijp0}$ to see the impact on young firm activity

REAL-WORLD APPLICATION OF THE BRIDGE

- Our bridge is applicable to studying this idea by allowing:
 - ① the identification of the causal effect of **the Chinese competition**
(the coverage of both pre- and post-2000 periods)
 - ② the **Diff-in-Diff (DD) specification** to identify the Chinese competition shock
(the coverage of pre-1990s helps test the parallel pre-trends assumption)
 - ③ to study the effect on innovation activities of **young firms and business dynamism**
(the coverage of non-public firms)

TABLE OF CONTENTS

- 1 Data
- 2 Matching Methodology
- 3 Match Results
- 4 Benefits and Practical Applications of the Bridge
- 5 Concluding Remarks

CONCLUSION AND FUTURE WORK

- We construct a longitudinally consistent linkage b/w US patent assignees and firms (1976-2016)
 - by introducing an internet search-aided algorithm

CONCLUSION AND FUTURE WORK

- We construct a longitudinally consistent linkage b/w US patent assignees and firms (1976-2016)
 - by introducing an internet search-aided algorithm
- We improve matching rate and extend the time horizon of existing crosswalks
- This can provide benefits to researchers studying firm innovation and business dynamism

CONCLUSION AND FUTURE WORK

- We construct a longitudinally consistent linkage b/w US patent assignees and firms (1976-2016)
 - by introducing an internet search-aided algorithm
- We improve matching rate and extend the time horizon of existing crosswalks
- This can provide benefits to researchers studying firm innovation and business dynamism
- Still, there is room for improvement of the current bridge
 - The current matching procedures do not include manual matching
 - False positive results might still exist; can further report false positive rates by each model type

CONCLUSION AND FUTURE WORK

- We construct a longitudinally consistent linkage b/w US patent assignees and firms (1976-2016)
 - by introducing an internet search-aided algorithm
- We improve matching rate and extend the time horizon of existing crosswalks
- This can provide benefits to researchers studying firm innovation and business dynamism
- Still, there is room for improvement of the current bridge
 - The current matching procedures do not include manual matching
 - False positive results might still exist; can further report false positive rates by each model type
- Working on technical notes presenting more details on match info/statistics – please stay tuned!

THANK YOU! 😊

yding4@worldbank.org

karamjo@gmail.com

sk6285@princeton.edu

Appendix

SEQUENCE OF NAME MATCHING

Model	Assignee Name	State	City	Sequence
A1	Strict Name	Strict State	Strict City	1
A2	Strict Name	Strict State	Fuzzy City	2
A3	Fuzzy Name	Strict State	Strict City	3
A4	Fuzzy Name	Strict State	Fuzzy City	4
B1	Strict Name	Missing State	Strict City	5
B2	Strict Name	Missing State	Fuzzy City	6
B3	Fuzzy Name	Missing State	Strict City	7
B4	Fuzzy Name	Missing State	Fuzzy City	8
C1	Strict Name	Strict State	Missing City	9
C2	Fuzzy Name	Strict State	Missing City	10
D1	Strict Name	Strict State	Different City	11
D2	Fuzzy Name	Strict State	Different City	12
E1	Strict Name	Missing State	Missing City	13
E2	Fuzzy Name	Missing State	Missing City	14
F1	Strict Name	Different States	Same City (Strict or Fuzzy)	15
F2	Strict Name	Different States	Missing City	16
F3	Strict Name	Missing State	Different City	17
F4	Strict Name	Different States	Different City	18
G1	Fuzzy Name	Different States	Same City (Strict or Fuzzy)	19
G2	Fuzzy Name	Different States	Missing City	20
G3	Fuzzy Name	Missing State	Different City	21
G4	Fuzzy Name	Different States	Different City	22

Table 6: Models to Match Patent Assignee and BR Establishment

◀ Main

SEQUENCE OF NAME MATCHING

Model	Assignee Name	State	City	Sequence
A1	Strict Name	Strict State	Strict City	1
A2	Strict Name	Strict State	Fuzzy City	2
A3	Fuzzy Name	Strict State	Strict City	3
A4	Fuzzy Name	Strict State	Fuzzy City	4
B1	Strict Name	Missing State	Strict City	5
B2	Strict Name	Missing State	Fuzzy City	6
B3	Fuzzy Name	Missing State	Strict City	7
B4	Fuzzy Name	Missing State	Fuzzy City	8
C1	Strict Name	Strict State	Missing City	9
C2	Fuzzy Name	Strict State	Missing City	10
D1	Strict Name	Strict State	Different City	11
D2	Fuzzy Name	Strict State	Different City	12
E1	Strict Name	Missing State	Missing City	13
E2	Fuzzy Name	Missing State	Missing City	14
F1	Strict Name	Different States	Same City (Strict or Fuzzy)	15
F2	Strict Name	Different States	Missing City	16
F3	Strict Name	Missing State	Different City	17
F4	Strict Name	Different States	Different City	18
G1	Fuzzy Name	Different States	Same City (Strict or Fuzzy)	19
G2	Fuzzy Name	Different States	Missing City	20
G3	Fuzzy Name	Missing State	Different City	21
G4	Fuzzy Name	Different States	Different City	22

Table 6: Models to Match Patent Assignee and BR Establishment

◀ Main

- For assignees matched w/ multiple LBD firms, pick the one w/ the highest Jaro-Winkler score
 - Calculate Jaro-Winkler similarly b/w patent assignee name and all the linked BR establishment names
 - A patent assignee is matched to a unique firm in a given reference year
 - Firm-level concordance b/w patent assignees and LBD firms
- Link patents of assignees to the matched LBD firms
 - Could be at most two matches for a given patent (one by application year and the other by grant year)
 - Use the same criteria as before, and then compare the year gaps [▶ More](#)
 - Patent-level concordance b/w patent assignees and LBD firms

SORTING ORDER FOR THE PATENT-LEVEL MATCH

Year Window	Sequence
appyear	1
gyear	2
appyear-1	3
gyear-1	4
appyear-2	5
gyear-2	6
appyear-3	7
gyear-3	8
appyear+1	9
gyear+1	10
appyear+2	11
gyear+2	12
appyear+3	13
gyear+3	14

Table 7: Preference Ordering of the Patent-level Match

[← Main](#)

SEQUENCE OF NAME MATCHING (STEM NAME)

Model	State	City	Score
AA1	Strict State	Strict City	11
AA2	Strict State	Fuzzy City	10
BB1	Missing State	Strict City	9
BB2	Missing State	Fuzzy City	8
CC	Strict State	Missing City	7
DD	Strict State	Different City	6
EE	Missing State	Missing City	5
FF1	Different States	Same City (Strict or Fuzzy)	4
FF2	Different States	Missing City	3
FF3	Missing State	Different City	2
FF4	Different States	Different City	1

Table 8: Models to Evaluate Stem Name Matches

[← Main](#)

SEQUENCE OF NAME MATCHING (STEM NAME)

Model	State	City	Score
AA1	Strict State	Strict City	11
AA2	Strict State	Fuzzy City	10
BB1	Missing State	Strict City	9
BB2	Missing State	Fuzzy City	8
CC	Strict State	Missing City	7
DD	Strict State	Different City	6
EE	Missing State	Missing City	5
FF1	Different States	Same City (Strict or Fuzzy)	4
FF2	Different States	Missing City	3
FF3	Missing State	Different City	2
FF4	Different States	Different City	1

Table 8: Models to Evaluate Stem Name Matches

[← Main](#)

(PSEUDO) MATCH RESULT EXAMPLE: IBM

USPTO Patent Assignees

assignee	organization	uspto_std	uspto_stm	city	state
1	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY
2	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NJ
3	International Buniess Macjines Corporation	INT BUNIESS MACJINES CORP	INT BUNIESS MACHINES	ARMONK	NY
4	IBM Corp	IBM CORP	IBM	ARMONK	NY
5	International Business Machines	INT BUSINESS MACHINES	INT BUSINESS MACHINES	ARMONG	NY

LBD-BR Target Firm

firm_id	BR_name	BR_std	BR_stm	city	state
yyyyyyy	International Business Machines Corporation	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY

[← Main](#)

(PSEUDO) MATCH RESULT EXAMPLE: IBM

USPTO Patent Assignees

matching method	organization	uspto_std	uspto_stm	city	state
Strict name & addr. (city,state) (A1)	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY
Strict name & addr. (city only) (F1)	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NJ
Fuzzy name & addr. (city,state) (A3)	International Buniess Macjines Corporation	INT BUNIESS MACJINES CORP	INT BUNIESS MACHINES	ARMONK	NY
	IBM Corp	IBM CORP	IBM	ARMONK	NY
	International Business Machines	INT BUSINESS MACHINES	INT BUSINESS MACHINES	ARMONG	NY

LBD-BR Target Firm

firm_id	BR_name	BR_std	BR_stm	city	state
yyyyyy	International Business Machines Corporation	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY

[◀ Main](#)

(PSEUDO) MATCH RESULT EXAMPLE: IBM

USPTO Patent Assignees

matching method	organization	uspto_std	uspto_stm	city	state
Strict name & addr. (city,state) (A1)	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY
Strict name & addr. (city only) (F1)	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NJ
Fuzzy name & addr. (city,state) (A3)	International Buniess Macjines Corporation	INT BUNIESS MACJINES CORP	INT BUNIESS MACHINES	ARMONK	NY
Internet search aided matching	IBM Corp	IBM CORP	IBM	ARMONK	NY
	International Business Machines	INT BUSINESS MACHINES	INT BUSINESS MACHINES	ARMONG	NY

LBD-BR Target Firm

firm_id	BR_name	BR_std	BR_stm	city	state
yyyyyy	International Business Machines Corporation	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY

[◀ Main](#)

(PSEUDO) MATCH RESULT EXAMPLE: IBM

USPTO Patent Assignees

matching method	organization	uspto_std	uspto_stm	city	state
Strict name & addr. (city,state) (A1)	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY
Strict name & addr. (city only) (F1)	International Business Machines Corp	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NJ
Fuzzy name & addr. (city,state) (A3)	International Buniess Macjines Corporation	INT BUNIESS MACJINES CORP	INT BUNIESS MACHINES	ARMONK	NY
Internet search aided matching	IBM Corp	IBM CORP	IBM	ARMONK	NY
Stem name & fuzzy addr. (city,state) (AA2)	International Business Machines	INT BUSINESS MACHINES	INT BUSINESS MACHINES	ARMONG	NY

LBD-BR Target Firm

firm_id	BR_name	BR_std	BR_stm	city	state
yyyyyy	International Business Machines Corporation	INT BUSINESS MACHINES CORP	INT BUSINESS MACHINES	ARMONK	NY

[◀ Main](#)

MAIN RESULT: OVERALL IMPACT ON FIRM INNOVATION

	Δ Patents	Δ Self-cite
	(1)	(2)
NTR gap \times Post	0.049	0.052
	(0.279)	(0.291)
Observations	6,500	6,500
Fixed effects	j, p	j, p
Controls	full	full

- No evidence that foreign competition shock affects overall innovation intensity and composition

◀ Main

MAIN RESULT: ESCAPE-COMPETITION EFFECT

	Δ Patents	Δ Self-cite	Δ Patents	Δ Self-cite
	(1)	(2)	(3)	(4)
NTR gap \times Post	0.049	0.052	0.054	-0.051
	(0.279)	(0.291)	(0.287)	(0.295)
\times Innovation intensity			-0.017	0.784***
			(0.233)	(0.268)
Observations	6,500	6,500	6,500	6,500
Fixed effects	j, p	j, p	j, p	j, p
Controls	full	full	full	full

- No evidence that foreign competition shock affects overall innovation intensity and composition
- **BUT**, firms with tech. advantage increase internal innovation under foreign competitive pressure

◀ Main