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)
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
● Constructs a new bridge between USPTO patent data and administrative firm-level data
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)
• 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
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 with 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
    - standardized name $\rightarrow$ INT BUSINESS MACHINES CORP
    - stem name $\rightarrow$ 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 BU NI ES MAC J NES CORP fuzzy matching $\rightarrow$ INT BUSINESS MACHINES CORP

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

  - e.g. INT BUSINESS MACHINES CORP (in USPTO)
  - name, address matching $\rightarrow$ INT BUSIESS MACHINES CORP (in BR) (in Armonk, NY)
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)

\[ \text{e.g. International Business Machines Corporation} \rightarrow \text{INT BUSINESS MACHINES CORP} \]

\[ \text{stem name} \rightarrow \text{INT BUSINESS MACHINES} \]
**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 → INT BUSINESS MACHINES CORP
  
  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 BUΝΙESS MΑΧΙΝΕS CORP → 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
**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 → INT BUSINESS MACHINES CORP
  - 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 BU\text{NESS MACJINES CORP} → 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

⇒ ★ e.g. INT BUSINESS MACHINES CORP (in USPTO) (in Armonk, NY) → INT BU\text{NESS 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)
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)
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 $\geq 2$ 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”:**

<table>
<thead>
<tr>
<th>target</th>
<th>matched using internet search results</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
</tr>
<tr>
<td>IBM CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
</tr>
<tr>
<td>IBM CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
</tr>
<tr>
<td>IBM CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
</tr>
<tr>
<td>IBM CORP</td>
<td>INT BUSINESS MACHINES CORPORATION</td>
</tr>
</tbody>
</table>
• Combine this with the previously constructed patent-firm crosswalk
Combine this with the previously constructed patent-firm crosswalk

e.g. IBM CORP → 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 $\xleftarrow{\text{Patent-firm crosswalk (step 1-3)}}$ INT BUSINESS MACHINES CORP (in BR)
Combine this with the previously constructed patent-firm crosswalk

- IBM CORP → INT BUSINESS MACHINES CORP
  - Internet search-aided
  - (new)

- INT BUSINESS MACHINES CORP (in BR) → INT BUSINESS MACHINES CORP (reference)
  - Patent-firm crosswalk (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 (new) \[\text{Internet search-aided} \rightarrow \text{INT BUSINESS MACHINES CORP (reference)}\] \[\text{Patent-firm crosswalk (step 1-3)} \rightarrow \text{INT BUSINESS MACHINES CORP (in BR)}\]
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) \(\rightarrow\) INT BUSINESS MACHINES CORP (reference) \(\leftarrow\) INT BUSINESS MACHINES CORP (in BR)

- Use the search-aided bridge for those unmatched and dropped matches (from step 1-3)
Combine this with the previously constructed patent-firm crosswalk: “search-aided bridge”

- 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) \(\rightarrow\) INT BUSINESS MACHINES CORP (reference) \(\leftarrow\) 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)

  ⇒ ★ e.g. IBM CORP (new)
    (in Armonk, NY) \(\rightarrow\) INT BUSINESS MACHINES CORP (in BR)
    (in Armonk, NY)
Step 5: Stem Name Matching & 2nd Round Search-Aided

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
1. Data

2. Matching Methodology

3. Match Results

4. Benefits and Practical Applications of the Bridge

5. Concluding Remarks
## Match Rates by Model Types

<table>
<thead>
<tr>
<th>Model</th>
<th>Patent Level</th>
<th>Assignee Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Name Matching (D+)</td>
<td>62</td>
<td>55.5</td>
</tr>
<tr>
<td>Stem Name Matching (D+)</td>
<td>14.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Search-Aided Algorithm</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>Std./Stem Name Matching D- (No Search-Aided)</td>
<td>4.5</td>
<td>8.9</td>
</tr>
<tr>
<td>Overall</td>
<td>88.2</td>
<td>80.1</td>
</tr>
</tbody>
</table>

**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)

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Patent Level</th>
<th>Assignee Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Name Matching (D+)</td>
<td>62</td>
<td>55.5</td>
</tr>
<tr>
<td>Stem Name Matching (D+)</td>
<td>14.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Search-Aided Algorithm</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>Std./Stem Name Matching D- (No Search-Aided)</td>
<td>4.5</td>
<td>8.9</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>88.2</strong></td>
<td><strong>80.1</strong></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Model</th>
<th>Patent Level</th>
<th>Assignee Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Name Matching (D+)</td>
<td>62</td>
<td>55.5</td>
</tr>
<tr>
<td>Stem Name Matching (D+)</td>
<td>14.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Search-Aided Algorithm</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>Std./Stem Name Matching D- (No Search-Aided)</td>
<td>4.5</td>
<td>8.9</td>
</tr>
<tr>
<td>Overall</td>
<td>88.2</td>
<td>80.1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Patent Level</th>
<th>Assignee Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Name Matching (D+)</td>
<td>62</td>
<td>55.5</td>
</tr>
<tr>
<td>Stem Name Matching (D+)</td>
<td>14.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Search-Aided Algorithm</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>Std./Stem Name Matching D- (No Search-Aided)</td>
<td>4.5</td>
<td>8.9</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>88.2</strong></td>
<td><strong>80.1</strong></td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Model</th>
<th>Patent Level</th>
<th>Assignee Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Name Matching (D+)</td>
<td>62</td>
<td>55.5</td>
</tr>
<tr>
<td>Stem Name Matching (D+)</td>
<td>14.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Search-Aided Algorithm</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>Std./Stem Name Matching D- (No Search-Aided)</td>
<td>4.5</td>
<td>8.9</td>
</tr>
<tr>
<td>Overall</td>
<td>88.2</td>
<td>80.1</td>
</tr>
</tbody>
</table>

**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 faction
Benefits of the Bridge

1. **Match rate has been improved**
   - Internet search-aided algorithm enhances the rate of matches by a non-negligible faction

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 faction

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)

\[ \text{NTR Gap}_j = \frac{\text{Non-NTR Rate}_j}{\text{WTO members, avg. 4%}} - \frac{\text{NTR Rate}_j}{\text{non-mkt econ., avg. 37%}} \]
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)
**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)

\[
\text{NTR Gap}_j = \frac{\text{Non-NTR Rate}_j}{\text{NTR Rate}_j}
\]

* Use **NTR Gap** 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)

\[ \Delta y_{ijp} = \beta_1 \text{Post}_p \times NTR \text{Gap}_{ijp0} + \beta_2 NTR \text{Gap}_{ijp0} + X_{ijp0} \gamma_1 + X_{ijp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp}. \]
• Diff-in-Diff regression to test the impact of China’s competition (Pierce and Schott 2016)

\[ \Delta y_{ijp} = \beta_1 Post_p \times NTR_{gap_{ijp0}} + \beta_2 NTR_{gap_{ijp0}} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \epsilon_{ijp}. \]

• \( i \): 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)

\[ \Delta y_{ijp} = \beta_1 \text{Post}_p \times NTR \text{ Gap}_{ijp0} + \beta_2 NTR \text{ Gap}_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \epsilon_{ijp} . \]

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

- \( \Delta y_{ijp} \): 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio

Y. Ding (WB), K. Jo (KDI), S. Kim (Princeton) Improving Patent Assignee-Firm Bridge with Web Search Results
**Real-World Application of the Bridge**

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

\[
\Delta y_{ijp} = \beta_1 Post_p \times NTR \text{ Gap}_{ijp0} + \beta_2 NTR \text{ Gap}_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp}.
\]

- \(i\): firm, \(j\): main 6-digit NAICS industry firm \(i\) belongs to, \(p\): time period (1992–1999, 2000–2007)
- \(\Delta y_{ijp}\): 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- \(NTR \text{ Gap}_{ijp0}\): emp-wgtd 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.)
Diff-in-Diff regression to test the impact of China’s competition (Pierce and Schott 2016)

\[ \Delta y_{ijp} = \beta_1 Post_p \times NTR \text{ Gap}_{ijp0} + \beta_2 NTR \text{ Gap}_{ijp0} + X_{ijp0} \gamma_1 + X_{jp0} \gamma_2 + \delta_j + \delta_p + \alpha + \epsilon_{ijp}. \]

- \( i \): firm, \( j \): main 6-digit NAICS industry firm \( i \) belongs to, \( p \): time period (1992–1999, 2000–2007)
- \( \Delta y_{ijp} \): 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- \( NTR \text{ Gap}_{ijp0} \): emp.-wgtd 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; \( \delta_j, \delta_p \): 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)**

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

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

- **\(\Delta y_{ijp}\)**: 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio

- **NTR \ Gap_{ijp0}**: emp.-wgtd 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; \(\delta_j, \delta_p\): industry, time fixed effects

⇒ See the impact of **China’s competition** on **different types of firm innovation**
**Real-World Application of the Bridge**

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

\[
\Delta y_{ijp} = \beta_1 \text{Post}_p \times \text{NTR Gap}_{ijp0} \times \text{InnovIntens}_{ijp0} + \beta_2 \text{Post}_p \times \text{NTR Gap}_{ijp0} + \beta_3 \text{Post}_p \times \text{InnovIntens}_{ijp0} \\
+ \beta_4 \text{NTR Gap}_{ijp0} \times \text{InnovIntens}_{ijp0} + \beta_5 \text{NTR Gap}_{ijp0} + \beta_6 \text{InnovIntens}_{ijp0} + X_{ijp0} \gamma_1 + X_{ijp0} \gamma_2 + \delta_j + \delta_p + \alpha + \varepsilon_{ijp}.
\]

- \( i \): firm, \( j \): main 6-digit NAICS industry firm \( i \) belongs to, \( p \): time period (1992–1999, 2000–2007)
- \( \Delta y_{ijp} \): 7yr DHS growth of i) No. patents firm applied each year, ii) avg. self-citation ratio
- \( \text{NTR Gap}_{ijp0} \): emp.-wghted avg. of 1999 industry-level NTR gaps across all the industries firm operates, measured in the start year for each period \( p0 \) (first diff.)
- \( \text{Post}_p \): Dummy equal to one for \( p = 2000–2007 \), post treatment period (second diff.)
- \( \text{InnovIntens}_{ijp0} \): avg. of past 5yr innovation intensity (nb. of patents / emp) measured in \( p0 \)
- \( X_{ijp0}, X_{ijp0} \): covariates for firm size, tech. class-level past patenting trend, trade status, NTR rate; \( \delta_j, \delta_p \): industry, time fixed effects

\[\Rightarrow\] Can interact w/ innovation intensity \( \text{InnovIntens}_{ijp0} \) to see how the impact depends on accumulated technological advantages of firms.
**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)
- \( \Delta y_{ijp} \): 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.)
- \( YoungFirm_{ijp0} \): young firm indicator measured in \( p_0 \)
- \( X_{ijp0}, X_{jp0} \): covariates for firm size, tech. class-level past patenting trend, trade status, NTR rate; \( \delta_j, \delta_p \): industry, time fixed effects

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

- 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!

Y. Ding (WB), K. Jo (KDI), S. Kim (Princeton) Improving Patent Assignee-Firm Bridge with Web Search Results
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

Y. Ding (WB), K. Jo (KDI), S. Kim (Princeton) Improving Patent Assignee-Firm Bridge with Web Search Results
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

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

**Table 6: Models to Match Patent Assignee and BR Establishment**
### Table 6: Models to Match Patent Assignee and BR Establishment

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

- For assignees matched with multiple LBD firms, pick the one with the highest Jaro-Winkler score
  - Calculate Jaro-Winkler similarly between 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 between 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
  - Patent-level concordance between patent assignees and LBD firms
# Sorting Order for the Patent-level Match

<table>
<thead>
<tr>
<th>Year Window</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>appyear</td>
<td>1</td>
</tr>
<tr>
<td>gyear</td>
<td>2</td>
</tr>
<tr>
<td>appyear-1</td>
<td>3</td>
</tr>
<tr>
<td>gyear-1</td>
<td>4</td>
</tr>
<tr>
<td>appyear-2</td>
<td>5</td>
</tr>
<tr>
<td>gyear-2</td>
<td>6</td>
</tr>
<tr>
<td>appyear-3</td>
<td>7</td>
</tr>
<tr>
<td>gyear-3</td>
<td>8</td>
</tr>
<tr>
<td>appyear+1</td>
<td>9</td>
</tr>
<tr>
<td>gyear+1</td>
<td>10</td>
</tr>
<tr>
<td>appyear+2</td>
<td>11</td>
</tr>
<tr>
<td>gyear+2</td>
<td>12</td>
</tr>
<tr>
<td>appyear+3</td>
<td>13</td>
</tr>
<tr>
<td>gyear+3</td>
<td>14</td>
</tr>
</tbody>
</table>

# Table 7: Preference Ordering of the Patent-level Match
## Sequence of Name Matching (STEM Name)

<table>
<thead>
<tr>
<th>Model</th>
<th>State</th>
<th>City</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA1</td>
<td>Strict State</td>
<td>Strict City</td>
<td>11</td>
</tr>
<tr>
<td>AA2</td>
<td>Strict State</td>
<td>Fuzzy City</td>
<td>10</td>
</tr>
<tr>
<td>BB1</td>
<td>Missing State</td>
<td>Strict City</td>
<td>9</td>
</tr>
<tr>
<td>BB2</td>
<td>Missing State</td>
<td>Fuzzy City</td>
<td>8</td>
</tr>
<tr>
<td>CC</td>
<td>Strict State</td>
<td>Missing City</td>
<td>7</td>
</tr>
<tr>
<td>DD</td>
<td>Strict State</td>
<td>Different City</td>
<td>6</td>
</tr>
<tr>
<td>EE</td>
<td>Missing State</td>
<td>Missing City</td>
<td>5</td>
</tr>
<tr>
<td>FF1</td>
<td>Different States</td>
<td>Same City (Strict or Fuzzy)</td>
<td>4</td>
</tr>
<tr>
<td>FF2</td>
<td>Different States</td>
<td>Missing City</td>
<td>3</td>
</tr>
<tr>
<td>FF3</td>
<td>Missing State</td>
<td>Different City</td>
<td>2</td>
</tr>
<tr>
<td>FF4</td>
<td>Different States</td>
<td>Different City</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 8: Models to Evaluate Stem Name Matches**
### Sequence of Name Matching (STEM Name)

<table>
<thead>
<tr>
<th>Model</th>
<th>State</th>
<th>City</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA1</td>
<td>Strict State</td>
<td>Strict City</td>
<td>11</td>
</tr>
<tr>
<td>AA2</td>
<td>Strict State</td>
<td>Fuzzy City</td>
<td>10</td>
</tr>
<tr>
<td>BB1</td>
<td>Missing State</td>
<td>Strict City</td>
<td>9</td>
</tr>
<tr>
<td>BB2</td>
<td>Missing State</td>
<td>Fuzzy City</td>
<td>8</td>
</tr>
<tr>
<td>CC</td>
<td>Strict State</td>
<td>Missing City</td>
<td>7</td>
</tr>
<tr>
<td>DD</td>
<td>Strict State</td>
<td>Different City</td>
<td>6</td>
</tr>
<tr>
<td>EE</td>
<td>Missing State</td>
<td>Missing City</td>
<td>5</td>
</tr>
<tr>
<td>FF1</td>
<td>Different States</td>
<td>Same City (Strict or Fuzzy)</td>
<td>4</td>
</tr>
<tr>
<td>FF2</td>
<td>Different States</td>
<td>Missing City</td>
<td>3</td>
</tr>
<tr>
<td>FF3</td>
<td>Missing State</td>
<td>Different City</td>
<td>2</td>
</tr>
<tr>
<td>FF4</td>
<td>Different States</td>
<td>Different City</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8: Models to Evaluate Stem Name Matches
USPTO Patent Assignees

<table>
<thead>
<tr>
<th>assignee</th>
<th>organization</th>
<th>uspto_std</th>
<th>uspto_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>2</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NJ</td>
</tr>
<tr>
<td>3</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>4</td>
<td>IBM Corp</td>
<td>IBM CORP</td>
<td>IBM</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>5</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>

LBD-BR Target Firm

<table>
<thead>
<tr>
<th>firm_id</th>
<th>BR_name</th>
<th>BR_std</th>
<th>BR_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>yyyy</td>
<td>International Business Machines Corporation</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>
### (Pseudo) Match Result Example: IBM

#### USPTO Patent Assignees

<table>
<thead>
<tr>
<th>matching method</th>
<th>organization</th>
<th>uspto_std</th>
<th>uspto_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strict name &amp; addr. (city,state) (A1)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>Strict name &amp; addr. (city only) (F1)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NJ</td>
</tr>
<tr>
<td>Fuzzy name &amp; addr. (city,state) (A3)</td>
<td>International Business Machines</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td></td>
<td>IBM Corp</td>
<td>IBM CORP</td>
<td>IBM</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td></td>
<td>International Business Machines</td>
<td>INT BUSINESS MACHINES</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>

#### LBD-BR Target Firm

<table>
<thead>
<tr>
<th>firm_id</th>
<th>BR_name</th>
<th>BR_std</th>
<th>BR_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>yyyyyy</td>
<td>International Business Machines Corporation</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>
### USPTO Patent Assignees

<table>
<thead>
<tr>
<th>matching method</th>
<th>organization</th>
<th>uspto_std</th>
<th>uspto_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strict name &amp; addr. (city,state) (A1)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>Strict name &amp; addr. (city only) (F1)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NJ</td>
</tr>
<tr>
<td>Fuzzy name &amp; addr. (city,state) (A3)</td>
<td>International Business Machines Corporation</td>
<td>INT BUNIESS MACHINES CORP</td>
<td>INT BUNIESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>Internet search aided matching</td>
<td>IBM Corp</td>
<td>IBM CORP</td>
<td>IBM</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td></td>
<td>International Business Machines</td>
<td>INT BUSINESS MACHINES</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONG</td>
<td>NY</td>
</tr>
</tbody>
</table>

### LBD-BR Target Firm

<table>
<thead>
<tr>
<th>firm_id</th>
<th>BR_name</th>
<th>BR_std</th>
<th>BR_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>yyyyyyy</td>
<td>International Business Machines Corporation</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>
### USPTO Patent Assignees

<table>
<thead>
<tr>
<th>matching method</th>
<th>organization</th>
<th>uspto_std</th>
<th>uspto_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strict name &amp; addr. (city,state) (A1)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>Strict name &amp; addr. (city only) (F1)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>ARMONK</td>
<td>NJ</td>
</tr>
<tr>
<td>Fuzzy name &amp; addr. (city,state) (A3)</td>
<td>International Business Machines Corp</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>Internet search aided matching</td>
<td>IBM Corp</td>
<td>IBM CORP</td>
<td>IBM CORP</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
<tr>
<td>Stem name &amp; fuzzy addr. (city,state) (AA2)</td>
<td>International Business Machines</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>

### LBD-BR Target Firm

<table>
<thead>
<tr>
<th>firm_id</th>
<th>BR_name</th>
<th>BR_std</th>
<th>BR_stm</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>yyyyyyy</td>
<td>International Business Machines Corporation</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>INT BUSINESS MACHINES CORP</td>
<td>ARMONK</td>
<td>NY</td>
</tr>
</tbody>
</table>
**Main Result: Overall Impact on Firm Innovation**

<table>
<thead>
<tr>
<th></th>
<th>( \Delta \text{Patents} )</th>
<th>( \Delta \text{Self-cite} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{NTR gap} \times \text{Post} )</td>
<td>0.049 (0.279)</td>
<td>0.052 (0.291)</td>
</tr>
<tr>
<td>Observations</td>
<td>6,500</td>
<td>6,500</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>( j, \ p )</td>
<td>( j, \ p )</td>
</tr>
<tr>
<td>Controls</td>
<td>full</td>
<td>full</td>
</tr>
</tbody>
</table>

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

**BUT**, firms with tech. advantage increase internal innovation under foreign competitive pressure.