# Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age? - Evidence from Patent Citations

Yao Li
University of Western Ontario

April 7, 2009

## Research Questions

- How localized is intranational and international knowledge flow? How do:

  - national and subnational borders affect diffusion
  - distance and internal distance affect diffusion

- Does the pattern of knowledge diffusion change?

  - Time trend?
  - Age profile?

- What are the sources of border effects in knowledge flows?

  - (Assignee) Self-citation
  - Aggregation bias

## Motivation

- Contentious debate about localization of intranational knowledge flows.

  *- Jaffe, Trajtenberge and Henderson (1993, QJE), HJT (2005, AER), Thompson and Fox-Kean (2005, AER).*

- Black box of localization of knowledge flows.

  - Most studies only examine the localization effect, e.g., *JTH (1993, 2005), TFK (2005), Thompson (2006, REStat), Griffth, Lee and Reenen (2007, NBER)*.

  - Do not explicitly decompose the contribution from distance and borders.

  - New and old knowledge may be different.

## What I Do

- Look at cross-patent citation database from NBER.

  - Patents embody ideas/knowledge.
  - Region *i*'s patents cite region *j*'s patents $=$ knowledge flows from region *j* to *i*.
  - Use patent citations to track knowledge flows.

- Assign patents to MSA (Metropolitan Statistical Areas), state and national level.
- Characterize age distribution of knowledge diffusion.
- Estimate border and distance effects.
- Analyze the changing pattern (age profile and time trend) of knowledge diffusion.

# Overview



## Main Findings

- Borders and distance matter for knowledge flows:
  - Excluding self-citations:
    halving distance $\uparrow$ citations by 5.5%;
    85% (of initial) knowledge lost crossing national border;
    78% lost crossing MSA border; 12% lost crossing state border.
  - Including self-citations $\uparrow$ border and distance effects.
    (Existing literature did not look at self-citations.)

- On average, national borders effect larger than subnational.
  Size of border and distance effects $\downarrow$ with patent age.
  Size of border and distance effects $\uparrow$ over time.

- Self-citation accounts for 50% border effects.
  Disaggregated data $\downarrow$ border effects.

# Overview



## Contribution

- Novel age profiles for border and distance effects. Consistent with knowledge diffusion process.
- New findings on time trend of border and distance effects. (not extensively studied in literature)
- Newly constructed finer data (matched at MSA level) helps to explore subnational localization and sources of border effect.
- Part of the resolutions proposed to border puzzle in knowledge flows might be extended and linked to trade flows in future.

# Empirical Specification and Data



- NBER Patent Citation Database:
  More than 3 million patents granted by US patent office.
  All citations (more than 16 million) made by each patent since 1975.
  More than 40% from foreigners (outside of the U.S.)
  Use patent citations between 1980 and 1997.
- Sample contains 357 regions:
  270 MSAs in the U.S. + 49 phantom MSAs (non-metro area for each state)
  Outside of the U.S.: 38 countries (main patent cited nations)
- Cover more than 99.9% patents and citations in NBER data.
  More than 93% can be matched to MSA.
- Sample size: $357 \times 357 \times 18 = 2294082$ region pairs.

# Empirical Specification and Data



- Empirical gravity equation motivated by theoretical gravity equation of knowledge flows (See Appendix of paper for derivation)

- Fixed effects: to control for unobserved multilateral resistance terms.

**Empirical Gravity Equation (Baseline Regression)**

$$\ln(\frac{c_{ij}}{y_i y_j}) = k + \alpha ln d_{ij} + \beta_1 B_{ij}^{sn} + \beta_2 B_{ij}^{n} + r_1^i CI^i + r_2^j CE^j + (1-\sigma)\varepsilon_{ij}$$

$c_{ij}$: how many citations region *j* receives from region *i* (i.e., region *i* cites region *j*'s knowledge; knowledge flows from *j* to *i*).

$y_j$: total number of citations region *j* receives.

$CI^i$: 1 if *i* is the citing region, 0 o.w.;
$CE^j$: 1 if *j* is the cited region, 0 o.w.

## Question 1

- **How localized is knowledge diffusion?**
  - Halving distance $\uparrow$ knowledge flows by 6.5% (5.5% if without self-citations).
  - Excluding self-citations, aggregate knowledge flows: 85% (of initial) knowledge lost crossing national border; 78% lost crossing MSA border; 12% lost crossing state border.
  - National border effect always larger than subnational.
  - Self-citations (SC) partly exaggerate border and distance effects.

# Main Results



| Specification: | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | With self-citation | | | Without self-citation | | |
| $lnd_{ij}$ | -0.131** | -0.211** | -0.154** | **-0.116**** | -0.167** | -0.128** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| $B_{ij}^{m}$ | -2.134** | | -2.245** | **-1.509**** | | -1.573** |
| | (0.014) | | (0.013) | (0.015) | | (0.014) |
| $B_{ij}^{s}$ | -0.224** | -0.655** | | **-0.124**** | -0.433** | |
| | (0.009) | (0.009) | | (0.009) | (0.009) | |
| $B_{ij}^{n}$ | -2.589** | -0.858** | -2.433** | **-1.903**** | -0.695** | -1.821** |
| | (0.018) | (0.015) | (0.017) | (0.019) | (0.015) | (0.018) |
| $B_{ij}^{m}$ effect | 8.449** | | 9.440** | **4.524**** | | 4.823** |
| | (0.119) | | (0.126) | (0.067) | | (0.067) |
| $B_{ij}^{s}$ effect | 1.252** | 1.925** | | **1.132**** | 1.542** | |
| | (0.011) | (0.017) | | (0.011) | (0.014) | |
| $B_{ij}^{n}$ effect | 13.316** | 2.360** | 11.390** | **6.707**** | 2.003** | 6.178** |
| | (0.243) | (0.034) | (0.196) | (0.126) | (0.029) | (0.109) |
| Citing-region effect | yes | yes | yes | yes | yes | yes |
| Cited-region effect | yes | yes | yes | yes | yes | yes |
| Year dummies | yes | yes | yes | yes | yes | yes |
| F-statistics | 1826 | 1714 | 1825 | 1721 | 1672 | 1723 |
| Adjusted $R^2$ | 0.74 | 0.73 | 0.74 | 0.73 | 0.72 | 0.73 |

Notes: ** Significant at 1% level.

## Question 2

- **Does the pattern of knowledge diffusion change?**

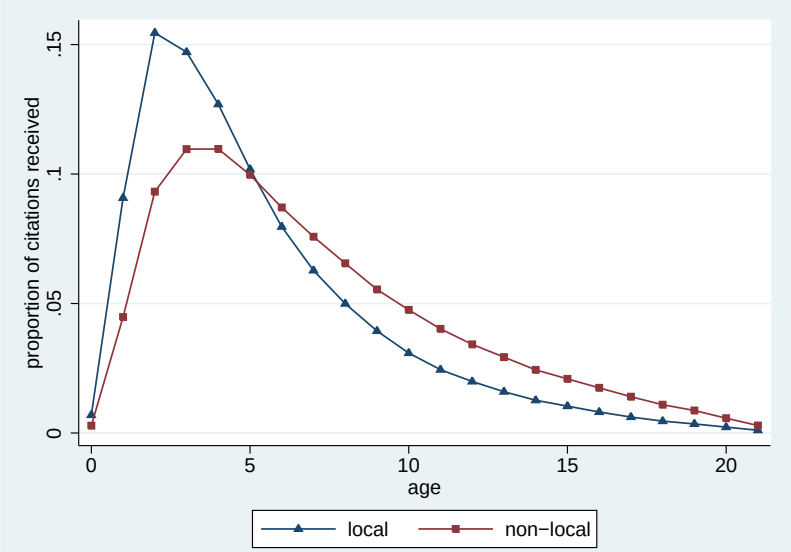
  *Conjecture in literature: "... given that we know that localization effects are likely to fade over time ... " (HJT, 2005, AER)*

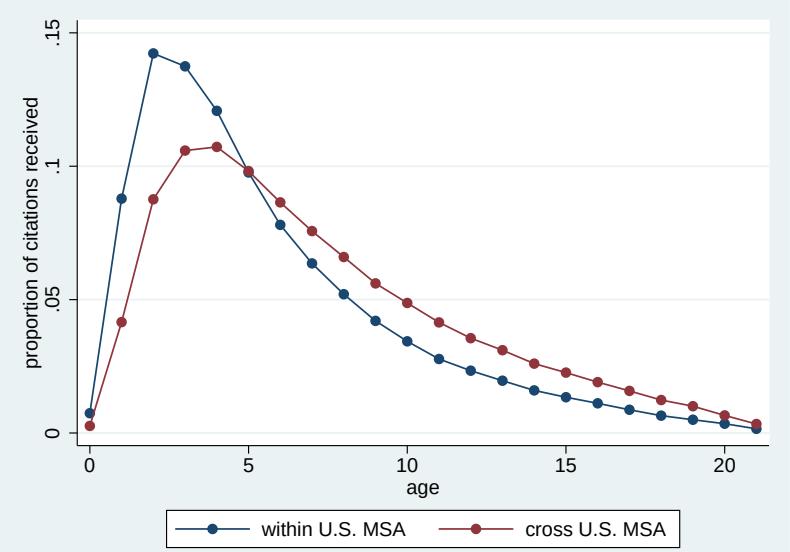
- Age is defined as a citation lag between cited and citing patent.
- Use proportion of citation received in total citation to characterize age distribution of knowledge diffusion.

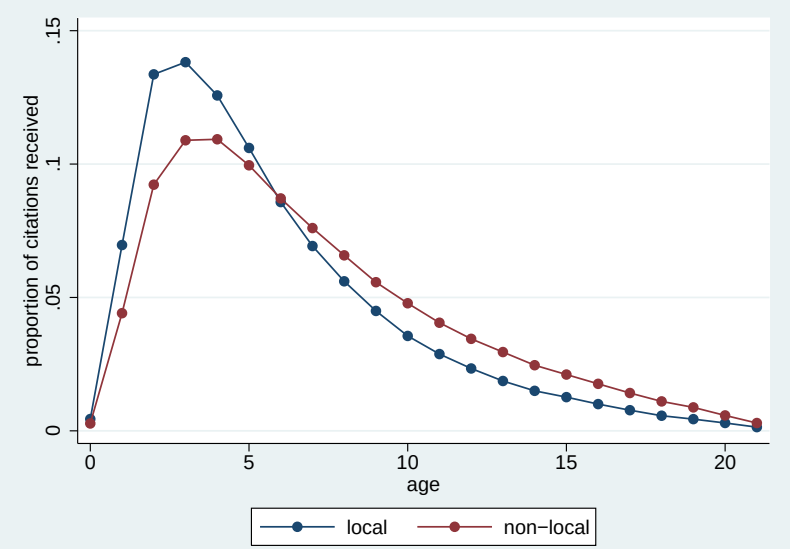
# Main Results

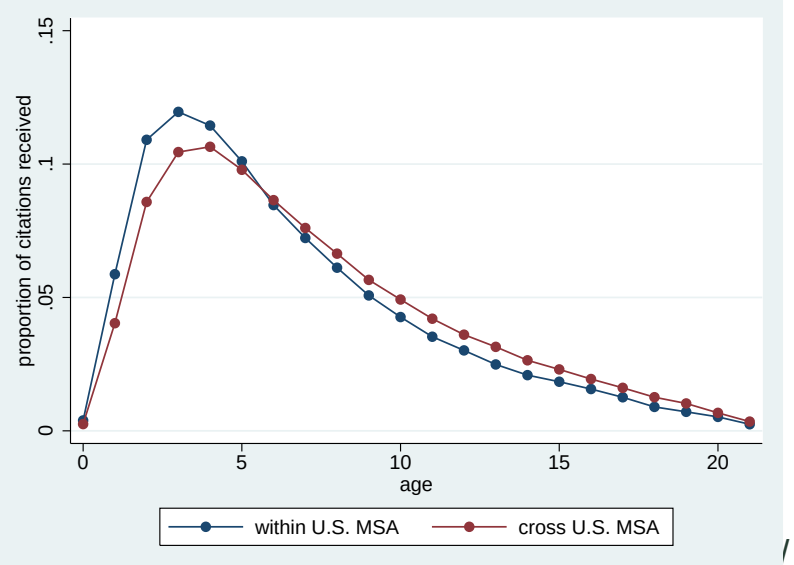


## **Age distribution of knowledge flows** (above: w/ SC; below: w/o SC)



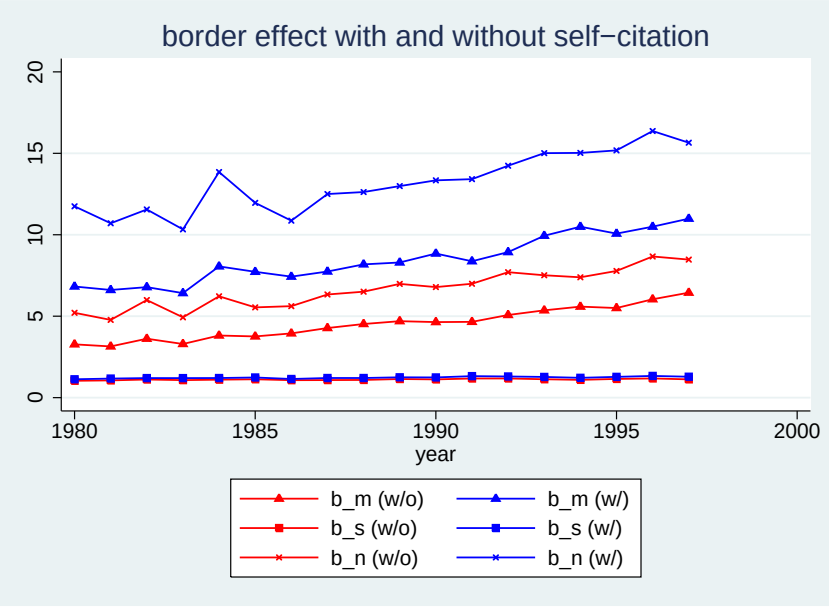
## Distance and border effects decrease with age of knowledge.

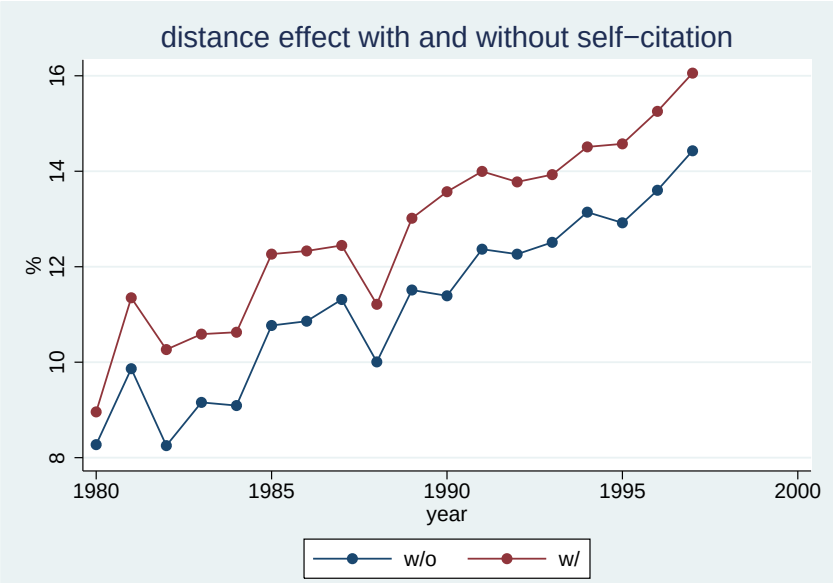
**Estimates by Age of Knowledge (without Self-citation)**

| Specification: | age [0,5) | age [5,10) | age [10,15) | age [15,20) | age [20,more) |
|---|---|---|---|---|---|
| $lnd_{ij}$ | -0.092** | -0.091** | -0.079** | -0.065** | -0.059* |
| | (0.003) | (0.004) | (0.005) | (0.008) | (0.028) |
| MSA border effect | 3.713** | 3.214** | 2.694** | 2.379** | 1.996** |
| | (0.074) | (0.077) | (0.092) | (0.136) | (0.382) |
| state border effect | 1.114** | 1.099** | 1.071** | 1.074† | 1.068 |
| | (0.016) | (0.018) | (0.025) | (0.041) | (0.148) |
| national border effect | 5.863** | 4.618** | 3.726** | 3.289** | 2.492** |
| | (0.151) | (0.140) | (0.159) | (0.229) | (0.570) |
| Citing-region effect | yes | yes | yes | yes | yes |
| Cited-region effect | yes | yes | yes | yes | yes |
| Year dummies | yes | yes | yes | yes | yes |
| F-statistics | 824 | 710 | 399 | 232 | 46 |
| Adjusted $R^2$ | 0.68 | 0.65 | 0.63 | 0.66 | 0.69 |

Notes: ** Significant at 1% level. * Significant at 5% level. † Significant at 10% level.

## Time trend



## Question 3: Sources of border effect?

- 11% of citations are self-citation.
  They account for approximately 50% MSA and national border effects.

- Aggregation bias:

  - Overestimate aggregate border effect.

  - Evidence:
  - 1) State to MSA level decomposing $\downarrow$ border effects.
  - 2) By age group decomposing $\downarrow$ border effects.
  - 3) By category decomposing $\downarrow$ border effects.

# Main Results



**Border and Distance Effects by Category (without Self-citation)**

| | Whole sample | Cat 1 Chemical | Cat 2 C.&C. | Cat 3 D.&M. | Cat 4 E.&E. | Cat 5 Mechanical | Cat 6 Others |
|---|---|---|---|---|---|---|---|
| $lnd_{ij}$ | -0.116** | -0.055* | -0.056** | -0.020 | -0.047** | -0.057** | -0.084** |
| | (0.002) | (0.009) | (0.015) | (0.014) | (0.009) | (0.006) | (0.004) |
| $B_{ij}^{m}$ Effect | 4.524** | 2.323** | 1.530** | 1.693** | 2.162** | 2.467** | 2.461** |
| | (0.067) | (0.131) | (0.153) | (0.149) | (0.132) | (0.086) | (0.068) |
| $B_{ij}^{s}$ Effect | 1.132** | 1.154** | 1.048 | 1.101 | 1.096* | 1.098** | 1.094** |
| | (0.011) | (0.050) | (0.077) | (0.077) | (0.050) | (0.028) | (0.021) |
| $B_{ij}^{n}$ Effect | 6.707** | 3.753** | 2.338** | 2.402** | 3.164** | 4.142** | 3.781** |
| | (0.126) | (0.263) | (0.273) | (0.247) | (0.231) | (0.184) | (0.132) |
| Citing effect | yes | yes | yes | yes | yes | yes | yes |
| Cited effect | yes | yes | yes | yes | yes | yes | yes |
| Year effect | yes | yes | yes | yes | yes | yes | yes |
| F-statistics | 1721 | 214 | 177 | 174 | 233 | 413 | 554 |
| Adjusted $R^2$ | 0.73 | 0.55 | 0.59 | 0.57 | 0.58 | 0.64 | 0.65 |

At finer category/industry level, distance less important; border effects $\downarrow$ but still significant.

National border effects always larger than subnational.

## Robustness to Alternative Specifications

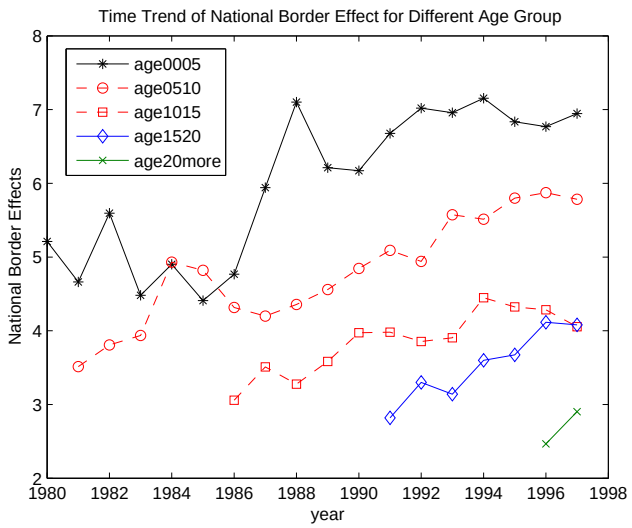
Test of interaction between time effect and age effect.

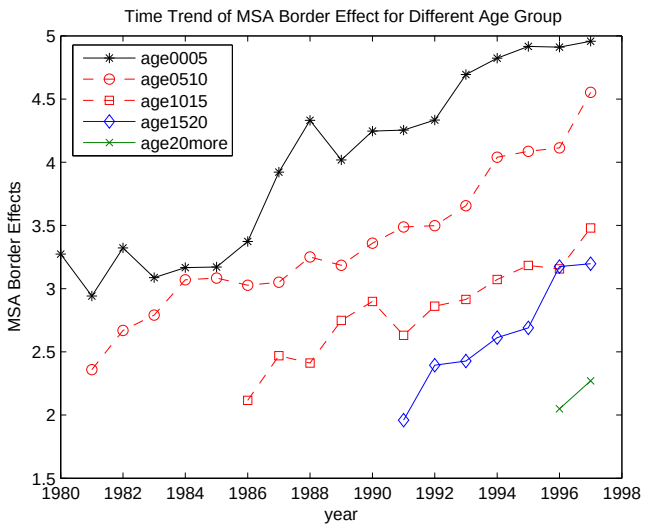
Time trend for border effects holds even for a subsample with similar age.

Among each category, age profiles are robust.

Among each category and each age group, time trend for border and distance effects ↑.

Exceptions of distance effect for very old patents.



# Conclusion



- Intranational and international knowledge diffusion are both strongly localized.
  Subnational border, national border and distance contribute to localization.
  Most subnational border effect comes from MSA level, rather than state.
- New knowledge faces the largest impediment to the diffusion.
  Time trend of border and distance effects is increasing.
- Sorting out self-citation substantially reduces border effects.
  Decomposing data contributes to the reduction of border effects.

# Ongoing Work



## Trade, Technology Diffusion and Welfare

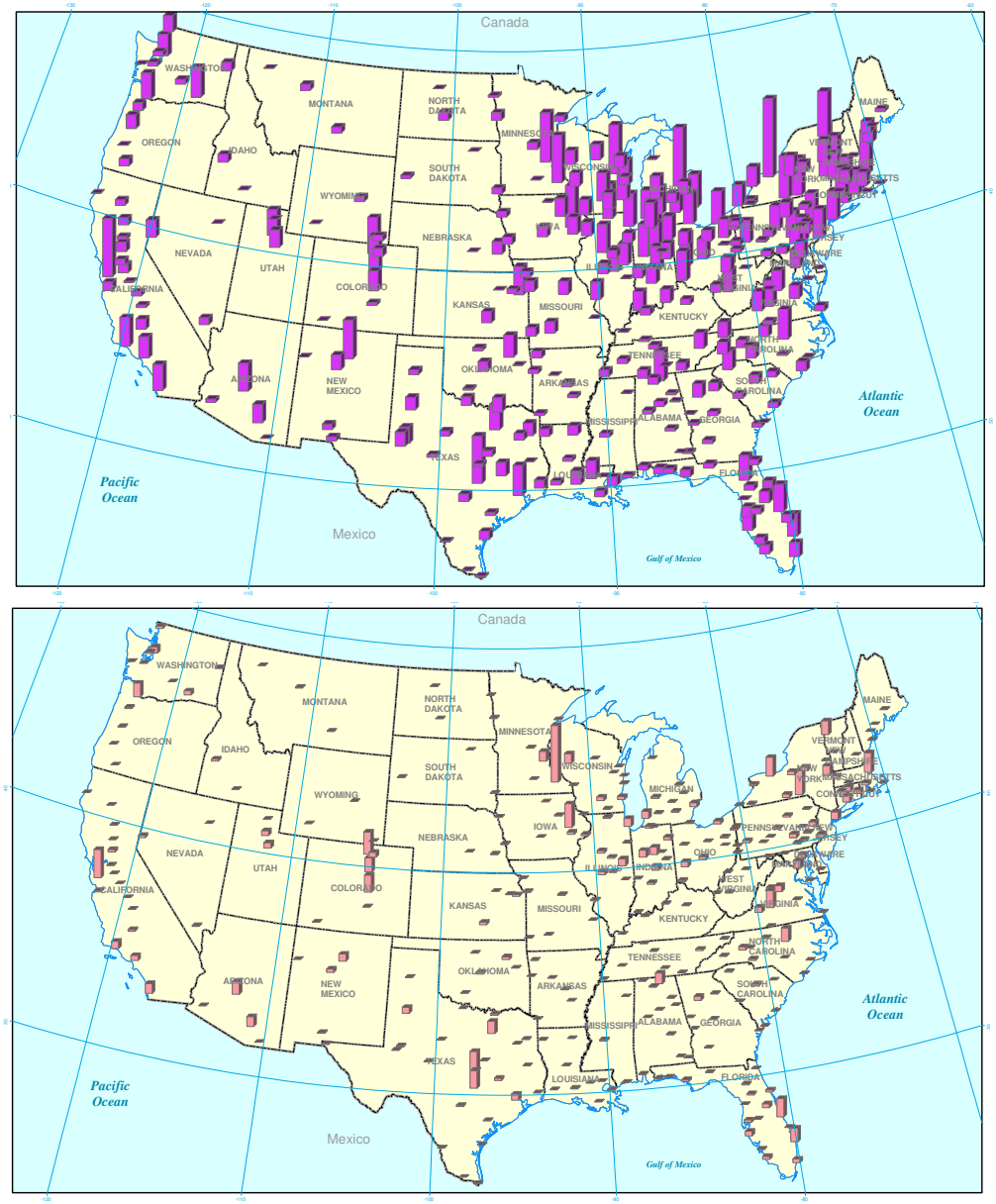
- Introducing country-pair specific barriers to bilateral technology diffusion in an Eaton-Kortum (Alvarez and Lucas) model of trade
- To investigate and quantify the impact of technology diffusion on trade pattern and welfare gains evaluation

## Knowledge Agglomeration and Border Effects

- To explain why the size of border effect is inversely related to the degree of concentration in knowledge flows.
- Consistent with stylized facts in trade flows.

# Ongoing Work





# Ongoing Work



***Thank you !***