The Geography of Knowledge Production: Connecting Islands and Ideas

Andrew B. Bernard
Tuck@Dartmouth
CEP, CEPR and NBER

Andreas Moxnes
Princeton & U. Oslo
CEPR

Yukiko U. Saito
Waseda U.
RIETI

Preliminary

International Trade Online, April 17, 2020

Introduction

- The production of ideas:

$$y = f(i_1, i_2, ..)$$

- The quality/quantity of ideas might depend on

  1. The composition of teams (matching function)
     - ⋆ e.g. quality of team members and team size.
  2. The productivity of teams (match quality).

- Economic integration likely to affect both 1) and 2).

- We know little about the impact of economic integration on these margins.

What we do

- Use universe of geocoded patent data from Japanese Patent Office.
- Natural experiment: Connecting a Japanese island with bridges.
- Study the activities of inventors with large fall in travel time before/after the connections.
  - Inventor productivity.
  - Team characteristics: size, quality, distance to co-inventors.

Literature

- *Scientific production*
  - Catalini et al (2020), Waldinger (2011), Iaria et al (2018), Agrawal and Goldfarb (2007)
- *Distance & spread of knowledge* (citations):
  - Comin et al (2012), Head et al (2019), Jaffe et al (1993)
- *Teams and innovation*
  - Akcigit et al (2018)
- *Geography and innovation*
  - Railroads: Perlman (2016)
- *The bridge literature*
  - Akerman (2009), Armenter et al (2014), Arnarson (2016), Brooks and Donavan (2019).

Today

- Data, measurement and stylized facts.
- Research design.
- Results.

Data

- Patent data from JPO, 1981 to 2005.
- For each (Japanese) patent, we know
  - the applicant(s) and the set of inventors.
  - the work address (geocode) of each inventor and applicant.
  - the citations they receive from future JPO patents.

Descriptives

	Patent-level				
	# Patents	# Inventors		# Citations	
Year		Mean	Median	Mean	Median
1988	332,215	2.12	2	0.67	0
1998	381,138	2.19	2	1.02	0
	Inventor-level				
	# Inventors	# Patents		# Citations	
1988	311,846	2.16	1	1.83	0
1998	434,635	1.92	1	2.26	0
	Applicant-level				
	# Applicants	# Patents		# Citations	
1988	39,810	9.05	1	6.09	0
1998	55,643	7.74	1	8.17	0

Note: Citations refer to 10-year citation count. Citations added by patent examiner are not included.

Share Shikoku inventors: 0.7% - Share foreign inventors: 24% (1998)

Measurement : Knowledge productivity

Cumulative citation-weighted number of patents:

$$z_{it} = \sum_{s=1981}^{t} \sum_{p \in \mathbb{P}_{is}} c_p,$$

- $\mathbb{P}_{is}$ is the set of i's patents in year s
- c_p is patent p's citations over the 10 years after filing.

Measurement : Team quality

Average lagged z of the co-inventors of i in year t (leave-out mean):

$$\bar{z}_{it} = \frac{1}{\sum_{p \in \mathbb{P}_{it}} (\nu_p - 1)} \sum_{p \in \mathbb{P}_{it}} \sum_{j \in \mathbb{I}_p \backslash i} z_{jt-1}$$

- $\mathbb{I}_p$ is the set of inventors on patent p
- ν_p is the number of inventors (team size) on p.

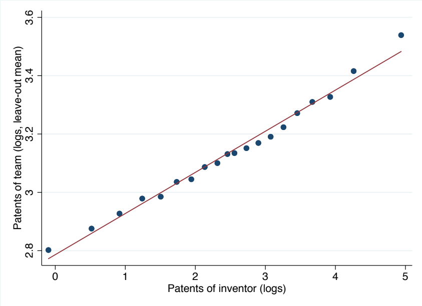
Measurement : Geography

Distance to co-inventors:

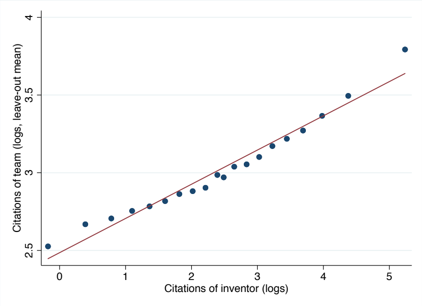
$$\bar{d}_{it} = \frac{1}{\sum_{p \in \mathbb{P}_{it}} (\nu_p - 1)} \sum_{p \in \mathbb{P}_{it}} \sum_{j \in \mathbb{I}_p \setminus i} \ln Dist_{ijt},$$

with $\ln Dist_{ijt} = 0$ if $Dist_{ijt} = 0$.

Fact 1: Positive assortative matching



(a) Quantity

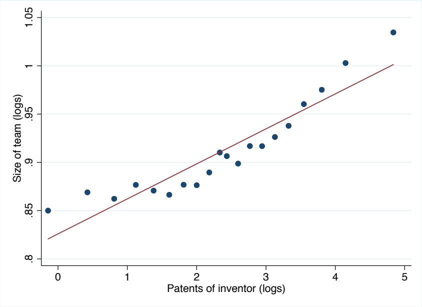


(b) Quality

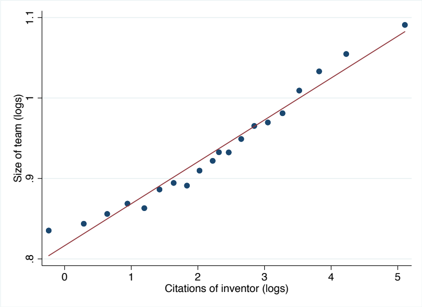
Note: All variables are demeaned by mesh averages. The OLS slope coefficients (solid lines) are .14 (left plot) and .21 (right plot). The sample includes all inventors filing a patent in 1998.

- Productive inventors work with each other.

Fact 2: Productive inventors work in larger teams



(c) Quantity

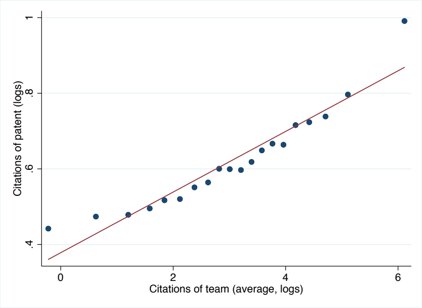


(d) Quality

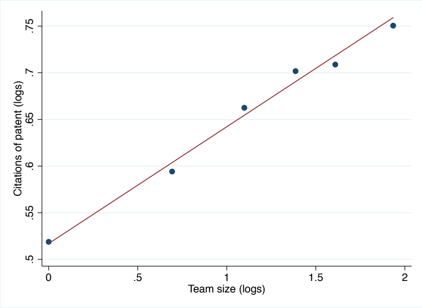
Note: All variables are demeaned by mesh averages. The OLS slope coefficients (solid lines) are .04 (left plot) and .05 (right plot). The sample includes all inventors filing a patent in 1998.

- Better inventors work on bigger teams.

Fact 3: Output rises with number and productivity of inventors



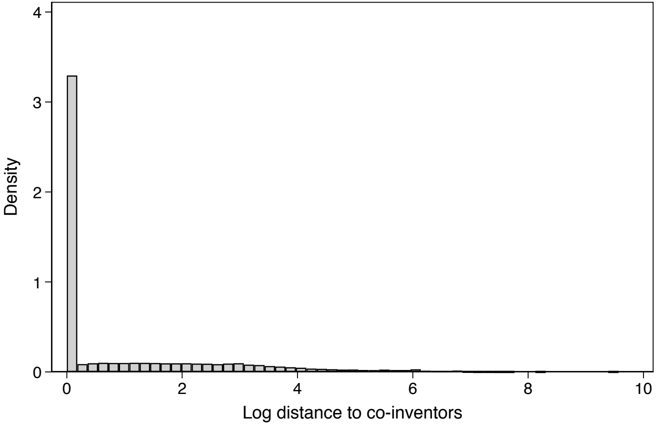
(e) Patent quality and team quality



(f) Patent quality and team size

Note: The OLS slope coefficients (solid lines) are .09 (left plot) and .14 (right plot). OLS coefficients with both team quality and team size included are .08 and .08. The sample includes all patents filed in 1998.

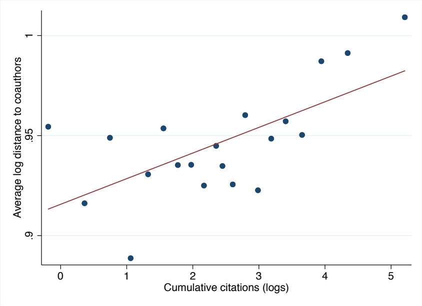
Fact 4: Most inventor teams are co-located.



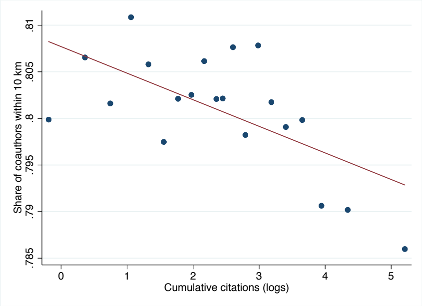
Note: The figure shows the histogram of average log distance to co-inventors across inventors. The sample includes all inventors filing a patent with at least one co-inventor in 1998.

- 59% of inventors have zero distance to co-inventors.
- Mean log distance = 1.01 (2.7km).

Fact 5: Productive inventors in more geographically dispersed teams



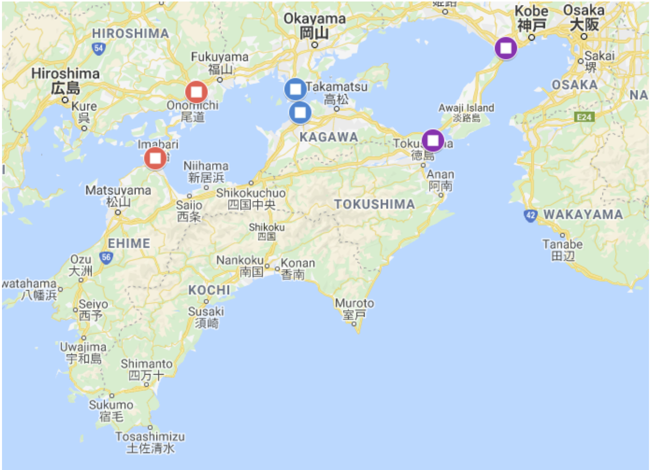
(g) Distance to coauthors



(h) Share of nearby coauthors

Note: All variables are demeaned by mesh averages. The OLS slope coefficients (solid lines) are .008 (left plot) and -.002 (right plot). The sample includes all inventors filing a patent in 1998.

Connecting islands and ideas



Great Seto Bridge (1988) Kobe-Awaji-Naruto Expressway (1998) Nishiseto Expressway (1999)

Great Seto Bridge

Nishiseto Expressway

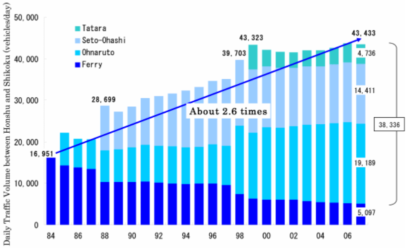
Kobe-Awaji-Naruto Expressway

Connecting islands and ideas

- Shikoku one of four main islands.
  - Population 4 million (3% of total) - relatively constant over time.
  - Economic activity concentrated in North-West.

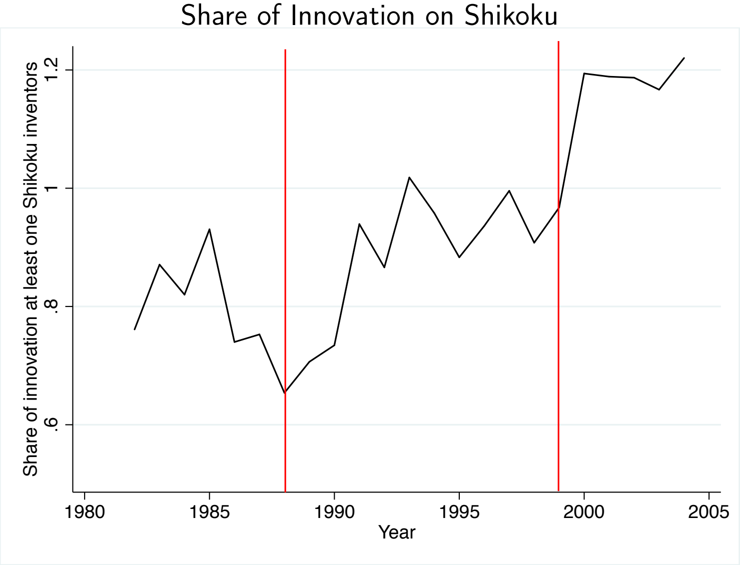
- Shikoku only accessible by ship/airplane until 1988.
  - Almost 3x vehicle traffic between 1984 and 2006.

Honshu-Shikoku Bridge Traffic 1984-2007



Source: Business Report of Honshu-Shikoku Bridges

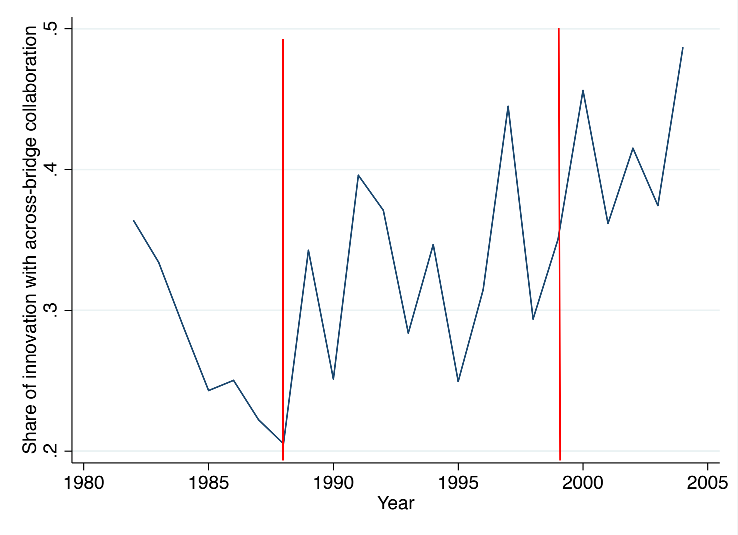
Innovation on Shikoku



Notes: The figure shows the share (in %) of citation-weighted patents with at least one inventor (>0) located on Shikoku in the application year. The population is all patents with at least one domestic inventor.

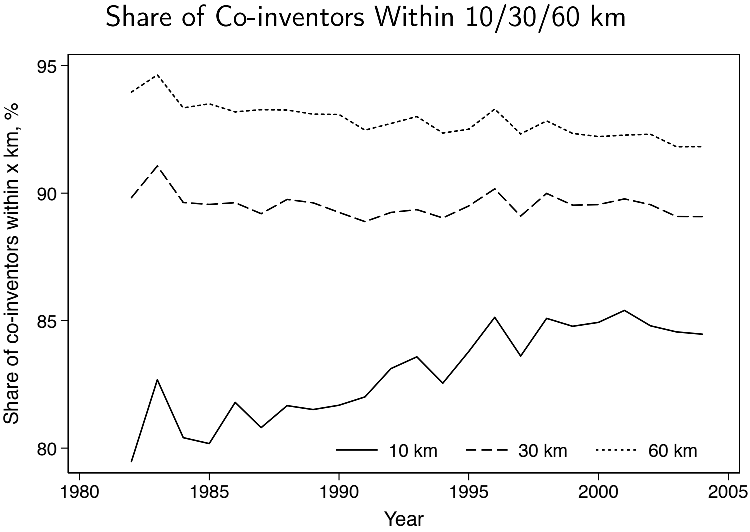
Shikoku collaboration

Share of Innovation with both Shikoku and non-Shikoku inventors



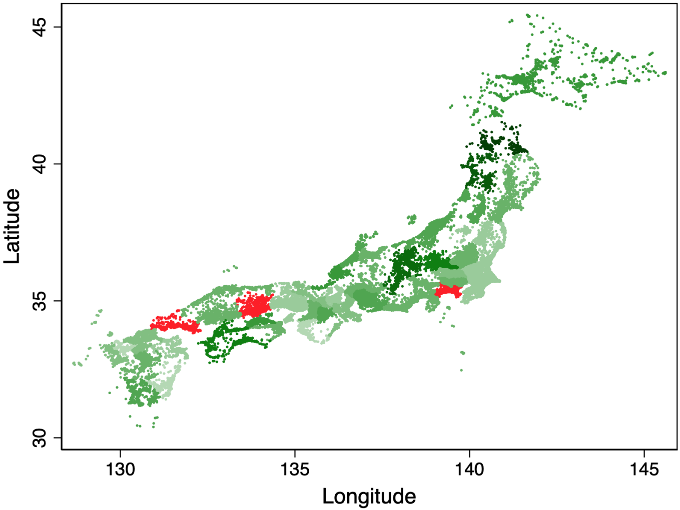
Notes: The figure shows the share (in %) of citation-weighted patents with at least one inventor located on Shikoku and one inventor not located on Shikoku. The population is all patents with at least two domestic inventors.

Collaboration distance



Notes: The figure is constructed by i) calculating the share of inventors within 10/30/60 km of each other, per patent, and then (ii) averaging these shares across patents using citations as weights. The population is all patents with at least two domestic inventors.

Innovation over time and space: 1994-1998 to 2000-2004



Notes: The figure show the change in innovation in each prefecture during from the period 1994-1998 and to 2000-2004. Innovation is measured by the citation-weighted number of patents across all inventors in a given cell. Darker green shades represent higher positive growth rates, whereas red shades represent negative growth rates.

Empirics: diff-in-diff

Change in outcomes for inventors with large vs small speed increase between Honshu/Shikoku, D_i.

- Based on last known location in 1998 or earlier.
- Regression
$$y_{it} = \alpha_i + \delta_t + \beta D_i \times Post_t + \gamma X_{it} + \varepsilon_{it}$$
  - $Post_t = t > 1999$.
- Controls, X_{it}:
  - ... interacted with $Post_t$:
    - ★ log distance from inventor i to Tokyo Station
    - ★ the quartile of i's pre-bridge productivity (z_{i1998})
    - ★ the first year i appears as an inventor (i.e., inventor age)
    - ★ the pre-bridge longitude and latitude of i
  - Shikoku trend ($t \times Shikoku_i$)

Getting to the other side : speed increase

Reduction in travel time between Honshu and Shikoku

	Ferry	Bridge	Speed increase
East route (Kobe city to Tokushima)	270	100	2.70
Central route (Kurashiki to Sakaide)	120	40	3.00
West route (Onomichi to Imabari)	160	60	2.67
Mean	183	67	2.75

Notes: Travel time in minutes across routes and modes (ferry/bridges). Source: Strait Crossings (2001).

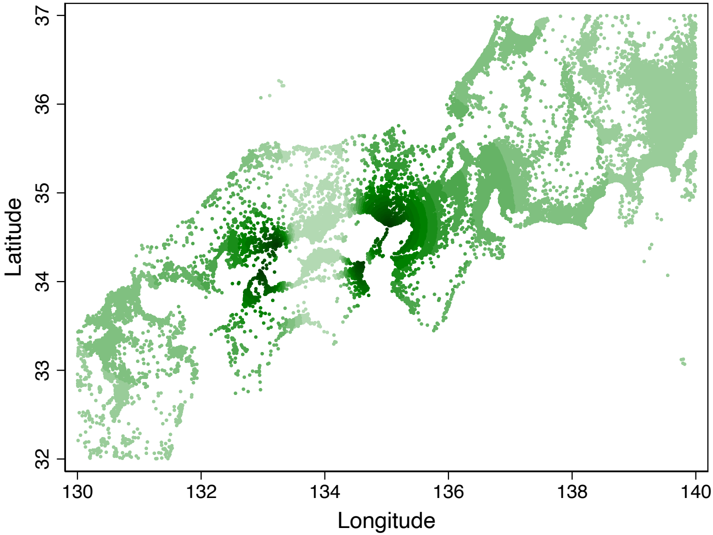
- Reliability: ferry suspended 280 times annually on average (Central route).

Getting to the other side : speed increase

- Bridges $k = \{1,2,3\}$ (central, east, west).
- Travel time to bridge k (minutes): $t_i^k = 60 d_i^k / \alpha$
  - d_i is geodesic distance to bridge k and $\alpha = 40$ km/h.
- $T^{Ferry} = 120$ min, $T^{Bridge} = 40$ min. Assume ferry links close to bridges.
- Let $t_i^* = \min_k \{t_i^k\}$ and $k_i^* = \arg\min_k \{t_i^k\}$. Speed increase in 98/99:

$$D_i = \begin{cases} 1 & \text{if} \quad k^* = 1 \\ \frac{t_i^* + T^{Ferry}}{t_i^* + T^{Bridge}} & \text{if} \quad k^* > 1 \text{ and } t_i^1 + T^{Bridge} > t_i^* + T^{Ferry} \\ \frac{t_i^1 + T^{Bridge}}{t_i^* + T^{Bridge}} & \text{if} \quad k^* > 1 \text{ and } t_i^1 + T^{Bridge} < t_i^* + T^{Ferry} \end{cases}$$

Getting to the other side : speed increase

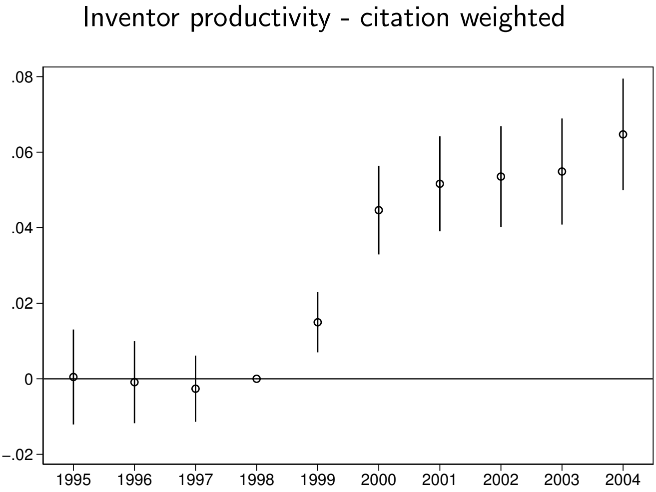


Notes: The figures show average D_i in each 10x10km cell for the 2nd/3rd bridge. Darker shades represent a greater speed increase to Honshu/Shikoku. The mean/median of D_i is 1.27/1.12.

Sample

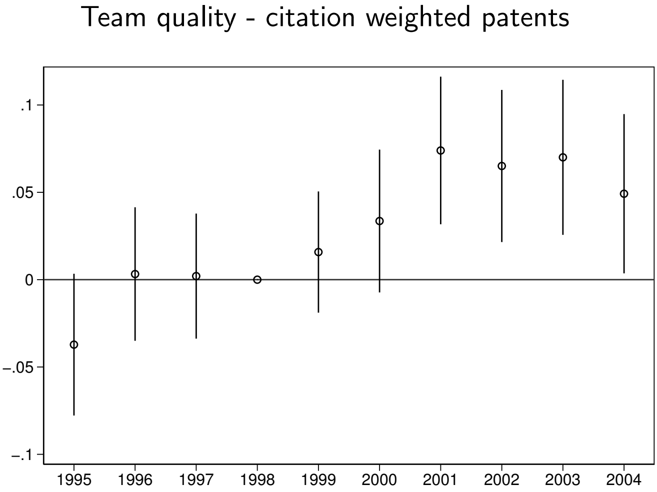
- Inventors with at least one patent in 1998 or earlier (to obtain last known geolocation).
- Time period 1995 to 2004.

Pre-trends : Productivity z_{it}



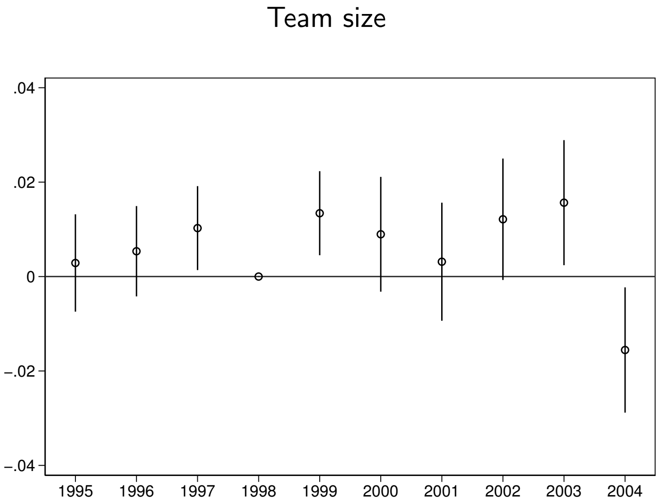
- Inventors with improved travel time increased productivity after the opening.

Pre-trends : Team quality $\bar{z}_{it}$



- Inventors with improved travel time increased team quality after the opening.

Pre-trends : Team Size



Results: Inventors and Teams

OLS Estimates: Inventor Productivity and Matching

Dep. var.	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i \times Post_t$	.036[a]	.067[a]	-.013	.050[a]	.072[a]	-.016
	(.005)	(.008)	(.010)	(.006)	(.011)	(.012)
Controls	No	No	No	Yes	Yes	Yes
Inventor FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	1,779,605	1,479,867	1,993,685	1,779,605	1,479,867	1,993,685

All inventors with a geocode in 1998 are included in the sample. The time period is 1995-2004. All specifications include inventor and year fixed effects. Robust standard errors clustered by inventor in parentheses. [a] $p< 0.01$, [b] $p< 0.05$, [c] $p< 0.1$.

- Higher inventor productivity and team quality for treated inventors.
- Max D_i is 3 vs median 1.12 $\longrightarrow$ 0.09 log points higher productivity.

Results: Mechanisms

OLS Estimates: Teams & Geography

	$\overline{\ln d_i}$	Share <10 km	Share Shikoku co-inventors		Share existing co-inventors	
$D_i \times Post_t$	.049[a]	-.003	.000	.000	.000	.000
	(.014)	(.003)	(.000)	(.001)	(.003)	(.003)
$Shikoku_i \times Post_t$				-.073[a]		-.020[b]
				(.008)		(.009)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Inventor FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	1,580,215	1,584,754	1,584,679	1,584,679	1,584,036	1,584,036

All inventors a team size >1 and with a geocode in 1998 are included in the sample. The time period is 1995-2004. The control variables are reported in the main text. Robust standard errors clustered by inventor in parentheses. [a] p< 0.01, [b] p< 0.05, [c] p< 0.1.

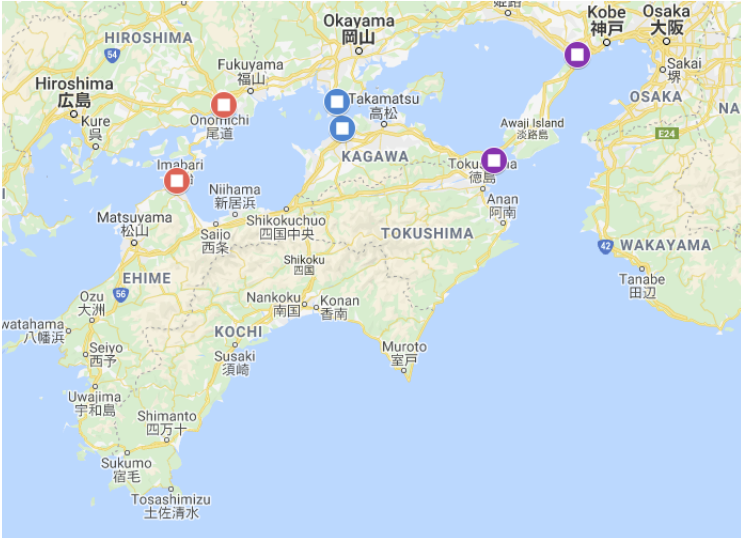
- Max D_i is 3 vs median 1.12 $\longrightarrow$ 10% higher distance to co-inventors.
- 7 percentage points fewer Shikoku co-inventors among Shikoku inventors.

Results: Summing up

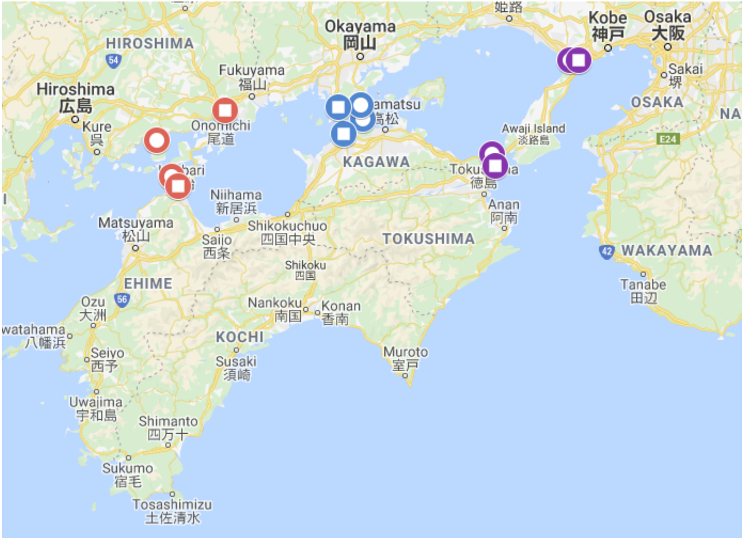
- More (quality-adjusted) output.
- Better matches.
- Greater distance to co-inventors.
- Fewer Shikoku co-inventors for Shikoku inventors.
- New team members for Shikoku inventors.

IV approach

- Identify 3 virtual bridges by minimizing distance over water.
- Similar in spirit to cost-based instruments for railroad (e.g. Banerjee et al, 2020) / highway (e.g. Duranton et al, 2014) network.
- Instrument D_i with log distance from inventor i to nearest 2nd/3rd *virtual* bridge.



Great Seto Bridge (1988) Kobe-Awaji-Naruto Expressway (1998) Nishiseto Expressway (1999)



Min distance - 2nd min distance - 3rd min distance

IV Results

2SLS Estimates: Inventor Productivity and Matching

	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i \times Post_t$	$.048^a$	$.067^a$	$-.022^b$
	(.003)	(.009)	(.010)
		First stage	
$D_i^{IV} \times Post_t$	1.015^a	1.015^a	1.015^a
	(.001)	(.001)	(.001)
Controls	Yes	Yes	Yes
Inventor FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	1,779,605	1,479,867	1,993,685

All inventors with a geocode in 1998 are included in the sample. The time period is 1995-2004. The control variables are reported in the main text. Robust standard errors clustered by inventor in parentheses. ^{a} $p < 0.01$, ^{b} $p < 0.05$, ^{c} $p < 0.1$.

Results: Heterogeneity

OLS Estimates: Heterogeneous Treatment Effects

	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i \times Post_t$	.043^a	.044^a	-.025
	(.009)	(.016)	(.016)
$D_i \times Post_t \times Highz_i$	.012	.047^a	.020
	(.010)	(.018)	(.019)
Controls	Yes	Yes	Yes
Inventor FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	1,779,605	1,479,867	1,993,685

All inventors with a geocode in 1998 are included in the sample. The time period is 1995-2004. The control variables are reported in the main text. Robust standard errors clustered by inventor in parentheses. ^{a} p$<$ 0.01, ^{b} p$<$ 0.05, ^{c} p$<$ 0.1.

- Define $Highz_i = 1$ if z_{i1998} is greater than median.
- Team quality/size increases more for highly productive inventors.

Robustness

1. Treatment indicators instead of continuous treatment.
2. The likelihood of inventor exit.
3. The 1988 bridge.
4. Quantity instead of quality results.

Robustness : Treatment indicators

OLS Estimates: Inventor Productivity and Matching

Dep. var.	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i^{Q2} \times Post_t$	$.021^a$	$.034^a$	$.116^a$
	(.004)	(.007)	(.008)
$D_i^{Q3} \times Post_t$	$.027^a$	$.055^a$	$.084^a$
	(.004)	(.009)	(.009)
$D_i^{Q4} \times Post_t$	$.050^a$	$.065^a$	$.080^a$
	(.005)	(.010)	(.011)
Controls	Yes	Yes	Yes
Inventor FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	1,779,605	1,479,867	1,993,685

All inventors with a geocode in 1998 are included in the sample. D_i^{Qk} is an indicator for whether the inventor is in the *kth* quartile of D_i. The 1st quartile is the omitted group. The time period is 1995-2004. The control variables are reported in the main text. Robust standard errors clustered by inventor in parentheses. ^{a} $p< 0.01$, ^{b} $p< 0.05$, ^{c} $p< 0.1$.

Robustness: Inventor Exit and Moves

OLS Estimates: Inventor exit and moves

	(1) Exit	(2) Exit	(3) Move	(4) Move
D_i	-.009[a]	-.014[a]	.011[a]	.001
	(.002)	(.003)	(.001)	(.001)
Controls	No	Yes	No	Yes
Inventor FE	No	No	No	No
Year FE	No	No	No	No
Obs	613,708	613,708	429,110	429,110

All inventors with a geocode in 1998 are included in the sample. The control variables are reported in the main text. Robust standard errors in parentheses. [a] $p< 0.01$, [b] $p< 0.05$, [c] $p< 0.1$.

- Exit = not in the patent data after 1999 (30%).
- Move = Shikoku $\leftrightarrow$ mainland by 2004 (24% among Shikoku inventors).
- Inventors with speed improvement less likely to stop inventing.

Robustness: 1988 Great Seto Bridge

Table: OLS Estimates: The Great Seto Bridge.

	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i \times Post_t$	$.071^a$	-.001	$-.118^a$
	(.018)	(.028)	(.026)
Controls	Yes	Yes	Yes
Inventor FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	1,046,681	945,436	1,390,313

All inventors with a geocode in 1988 are included in the sample. The time period is 1984-1993. The control variables are reported in the main text. Robust standard errors clustered by inventor in parentheses. [a] $p< 0.01$, [b] $p< 0.05$, [c] $p< 0.1$.

Robustness: Quantity not Quality

OLS Estimates: Inventor Productivity and Matching

Dep. var.	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\ln z_{it}$	$\ln \bar{z}_{it}$
$D_i \times Post_t$	$.031^a$	$.032^a$	$.022^a$	.010
	(.003)	(.007)	(.003)	(.009)
Controls	No	No	Yes	Yes
Inventor FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs	1,993,685	1,540,009	1,993,685	1,540,009

All inventors with a geocode in 1998 are included in the sample. The time period is 1995-2004. All specifications includes inventor and year fixed effects. Robust standard errors clustered by inventor in parentheses. All dependent variables are in logs. ^{a} $p< 0.01$, ^{b} $p< 0.05$, ^{c} $p< 0.1$.

Conclusions

- Work with people smarter than you.
- Bridges boost both productivity & matching.

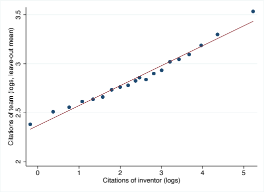
- Work in progress:
  - Mechanisms and competing hypotheses.
  - Aggregate effects - what about the left behind inventors?

Appendix : Team quality

- z_i and $\bar{z}_i$ mechanically related if inventors do not change teams.
- Instead calculate average z of co-inventors dropping all common patents:

$$\bar{z}_{it} = \frac{1}{\sum_{p \in \mathbb{P}_{it}} (\nu_p - 1)} \sum_{p \in \mathbb{P}_{it}} \sum_{j \in \mathbb{I}_p \setminus i} z_{jt-1}^{-i}$$

where z_{jt}^{-i} is the cumulative citations of j excluding patents in collaboration with i.



Note: All variables are demeaned by mesh averages. The OLS slope coefficient (solid lines) are .20. The sample includes all inventors filing a patent in 1998.

Appendix: Inventor name measurement error

- Inventors with identical names are treated as one.
- Solution: Drop inventors observed in on average ≥ 2 locations in the same year.
- 5% of inventors dropped.

OLS Estimates: Inventor Productivity and Matching

Dep. var.	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i \times Post_t$	$.055^a$	$.073^a$	-.014
	(.006)	(.011)	(.013)
Controls	Yes	No	Yes
Inventor FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	1,665,620	1,372,956	1,878,808

All inventors with a geocode in 1998 are included in the sample. The time period is 1995-2004. All specifica include inventor and year fixed effects. Robust standard errors clustered by inventor in parentheses. [a] p< 0.01, 0.05, [c] p< 0.1.

Appendix: Applicant regressions

- Change in outcomes for applicants with large vs small speed increase getting to the Honshu/Shikoku, D_i
- Based on last known location in 1998 or earlier.

OLS Estimates: Applicant Productivity and Matching

Dep. var.	$\ln z_{it}$	$\ln \bar{z}_{it}$	$\bar{v}_{it}$
$D_i \times Post_t$	$.046^b$	.004	$-.044^c$
	(.019)	(.031)	(.027)
Controls	Yes	No	Yes
Inventor FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	127,320	116,514	150,978

All applicants with a geocode in 1998 are included in the sample. The time period is 1995-2004. All specifica include applicant and year fixed effects. Robust standard errors clustered by applicant in parentheses. ^{a} p< 0.01, 0.05, ^{c} p< 0.1.