Language barriers and the speed of knowledge diffusion

Kyle Higham[1] **and Sadao Nagaoka[2,3]**

RIETI October 2022, Online

[1] Institute of Innovation Research, Hitotsubashi University
[2] Faculty of Economics, Tokyo Keizai University
[3] Research Institute of Economy, Trade and Industry (RIETI)

Hitotsubashi University Institute of Innovation Research

HITOTSUBASHI UNIVERSITY

1. Introduction
2. Background and Hypotheses
3. Empirical Setting and Data
4. Results
5. Conclusions

1. Introduction
2. Background and Hypotheses
3. Empirical Setting and Data
4. Results
5. Conclusions

Introduction

- Language barriers exist in global knowledge diffusion processes
- Previous research has confirmed this existence but gone no further
- There are many potential mediators of the impact of language barriers on knowledge diffusion in the context of innovation
- Research relating diffusion speed and language barriers is lacking

➢ We fill this gap by considering two questions:
  - ➢ Can we see accelerated diffusion upon the removal of a language barrier using patent citations?
  - ➢ If so, how do the properties of follow-on innovators and the inherent properties of the knowledge itself mediate this acceleration?

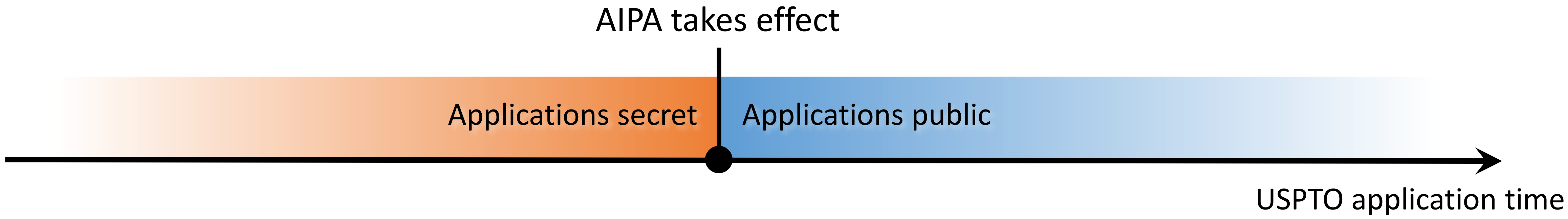
Introduction

- The Japanese language is an ideal setting for asking questions about the relationship between language and technical knowledge flow
  - There are no large concentrations of Japanese speakers outside of Japan
  - Japan is technologically advanced and produces a lot of new technical knowledge (much of which is only published in Japanese)
- We leverage a US patent policy change to examine the effect of the language barrier on knowledge flows into the US innovation ecosystem from Japan.

1. Introduction
2. Background and Hypotheses
3. Empirical Setting and Data
4. Results
5. Conclusions

Background and Hypotheses

- The American Inventor's Protection Act (1999) (AIPA) brought in many significant changes to the US patent system
- Notably, for us, this included the introduction of pre-grant publication 18 months after priority, which was already in effect in Japan and many other countries*



*Exemptions were only given when applications were not filed anywhere else

Background and Hypotheses

These policy goals of AIPA were quite explicit with respect to language barriers:

Notwithstanding how diligent a patent applicant might be, there are circumstances where such an applicant can nonetheless lose years of effective patent term due to delays in the PTO and other circumstances beyond her control. While our foreign competitors are able to see the latest U.S. patent technology in their native languages barely six months after a U.S. inventor files a patent application in their country, the reverse is not true. U.S. researchers and investors are denied the opportunity to learn what their foreign competitors are working on until a U.S. patent issues. This causes duplicative research and wasted developmental expenditures, putting U.S. inventors at a serious disadvantage *vis-a-vis* their foreign counterparts and competitors.

(US House Report 106-287 Part 1)

Background and Hypotheses

Hypotheses: Observation of knowledge flow acceleration

H1: The introduction of pre-grant publications to the US patent system **significantly reduced the citation lag for citations from US inventors, relative to Japanese inventors**, reflecting accelerated knowledge flow to the US inventors

H2: We observe the above effects even controlling for the preferences of citing local prior art, which we identify using the Japanese-U.S.-European triadic applications.

H3: Because examiner citations to recent prior art are unlikely to capture knowledge flows between inventors, **we will not observe a strong acceleration of examiner citations** from US-based applications relative to Japan-originating ones.

Background and Hypotheses

Hypotheses: Mediation of knowledge flow acceleration

The frequency of translations would increase with a firm's size and its experience in the Japanese market, due to appropriation advantage. Thus:

H4: The AIPA had the **largest impact on firms that were small** and had few resources to access Japan-originating technical knowledge before the policy change, or were **less likely to be directly involved in or otherwise familiar with the Japanese innovation ecosystem.**

Background and Hypotheses

Hypotheses: Mediation of knowledge flow acceleration

The effects of AIPA declines with the extent of translation before AIPA. Assuming that the translation is imperfect and ex-ante patent quality is difficult to assess, and that perceived patent quality is more heterogeneous in fast-moving technical fields:

H5: The AIPA had the largest effect on citation timing for high-quality patents and those in fast-moving technical fields.

1. Introduction
2. Background and Hypotheses
3. Empirical Setting and Data
4. Results
5. Conclusions

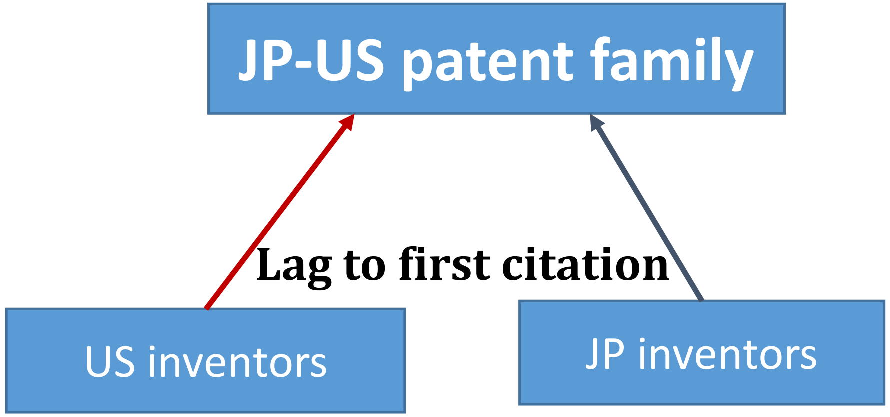
Empirical Setting and Data

- Take set of patent families with equivalents only filed in both US and Japan, filed by firms with Japan-based inventors (the "twin cohort")
- If application at USPTO was before November 29th 2000, applications would be available 18 months after priority date **in Japanese only**
- If application at USPTO was on or after November 29th 2000, applications would be available 18 months after priority date **in both Japanese and English**

→ Measure acceleration in knowledge diffusion caused by this change

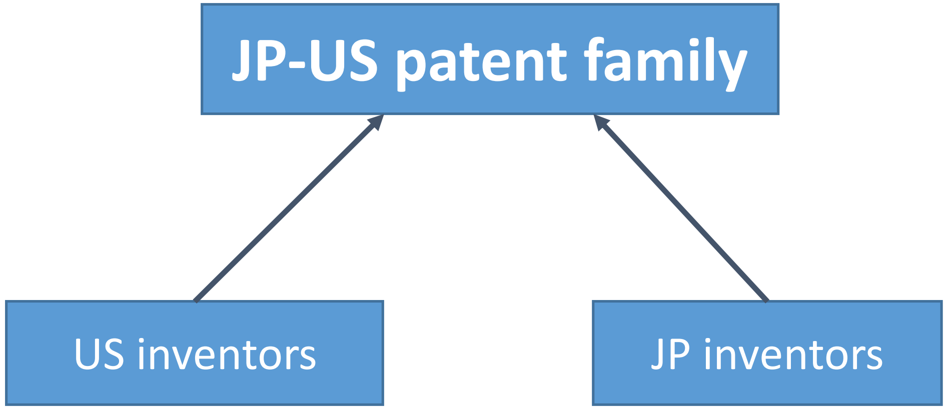
Empirical setting (H1-H3)

Pre-AIPA
(published 18 months after priority in Japanese only)



Pre-treatment

Post-AIPA
(published 18 months after priority in both English and Japanese)



Treated (can now access information without language barrier)

Control (information being made available in English does not change accessibility)

Difference-in-differences set-up:

$$Lag_{j\to i} = \alpha + \beta_1\, US_j + \beta_2\, Time_i + \beta_3\, US_j \times postAIPA_i$$
$$+\, \boldsymbol{\gamma} \cdot \boldsymbol{u}_i + \boldsymbol{\lambda} \cdot \boldsymbol{v}_j + \epsilon$$

With $\boldsymbol{u}$ and $\boldsymbol{v}$ as controls on cited and citing patents i and j.

Note: all citing patents are granted US patents.

Empirical setting (H4-H5)

- Incorporate mediating factors using two methods:
  - Triple difference
  - Split-sample (run previous DID on high/low-M groups)

Triple difference-in-differences set-up:

$$\begin{aligned} Lag_{j\to i} = \alpha &+ \beta_1\, US_j + \beta_2\, Time_i + \beta_3\, US_j \times postAIPA_i \\ &+ \beta_4\, M + \beta_5\, M \times postAIPA + \beta_6\, M \times US \\ &+ \beta_7\, M \times US \times postAIPA \\ &+ \boldsymbol{\gamma} \cdot \boldsymbol{u_i} + \boldsymbol{\lambda} \cdot \boldsymbol{v_j} + \epsilon \end{aligned}$$

With $\boldsymbol{u}$ and $\boldsymbol{v}$ as controls on cited and citing patents i and j, M is mediating factor.

Data

Sources:

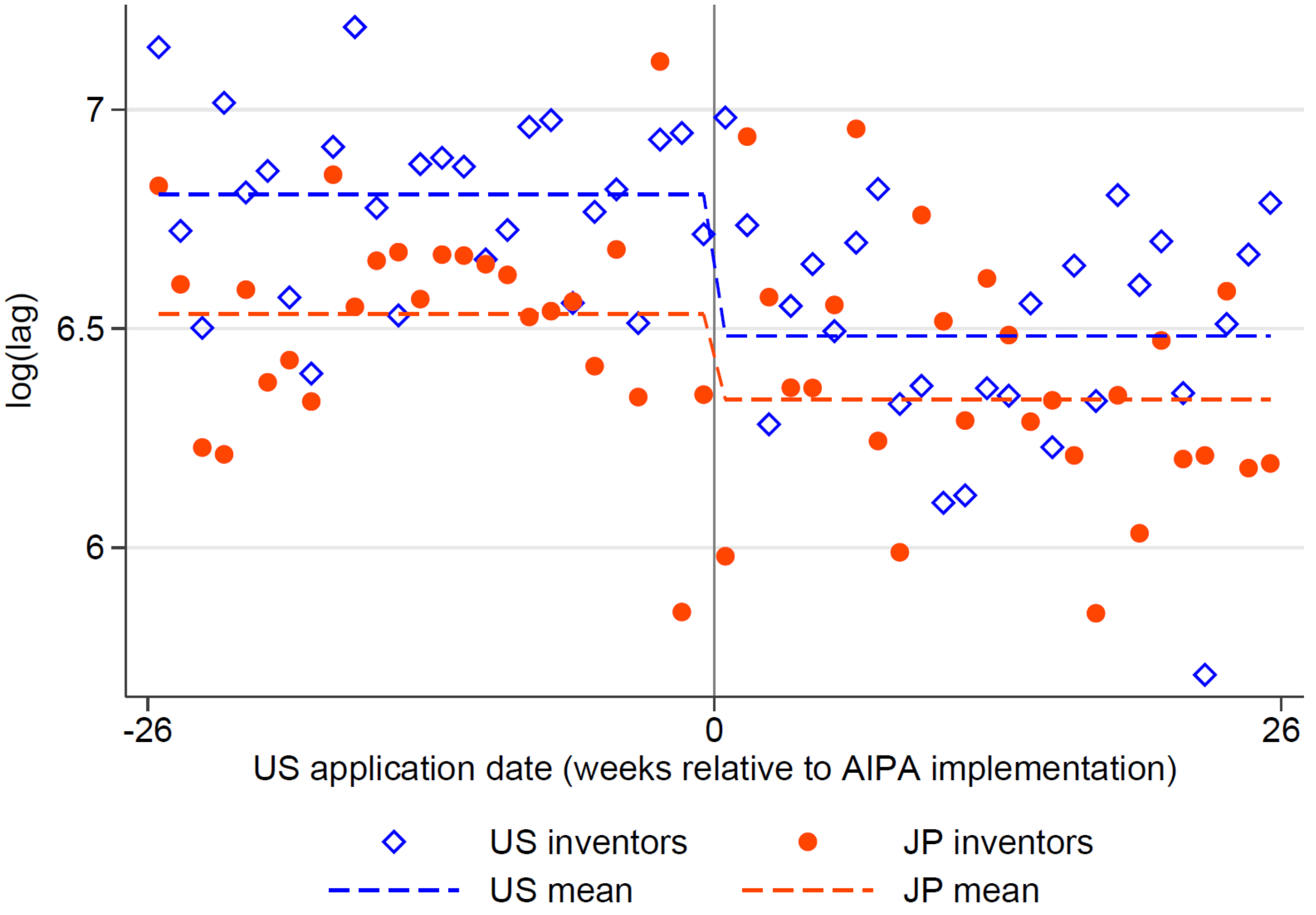
- PATSTAT: Twin application data and citation data
- USPTO: Harmonized inventor and firm ids with locations

Restrictions:

- 6-month window either side of AIPA implementation
- Time to first citation where each twin has at least one citation from each source (JP and US), thus bias towards high-quality
- Maximum 10 years citation lag, minimum 18 months
- Only consider first citation for each family-family citation pair
- No assignee self-citations
- Citing applicant must have at least one patent before policy change

1. Introduction
2. Background and Hypotheses
3. Empirical Setting and Data
4. Results
5. Conclusions

Results: Raw first-citation lags (applicant, twins)



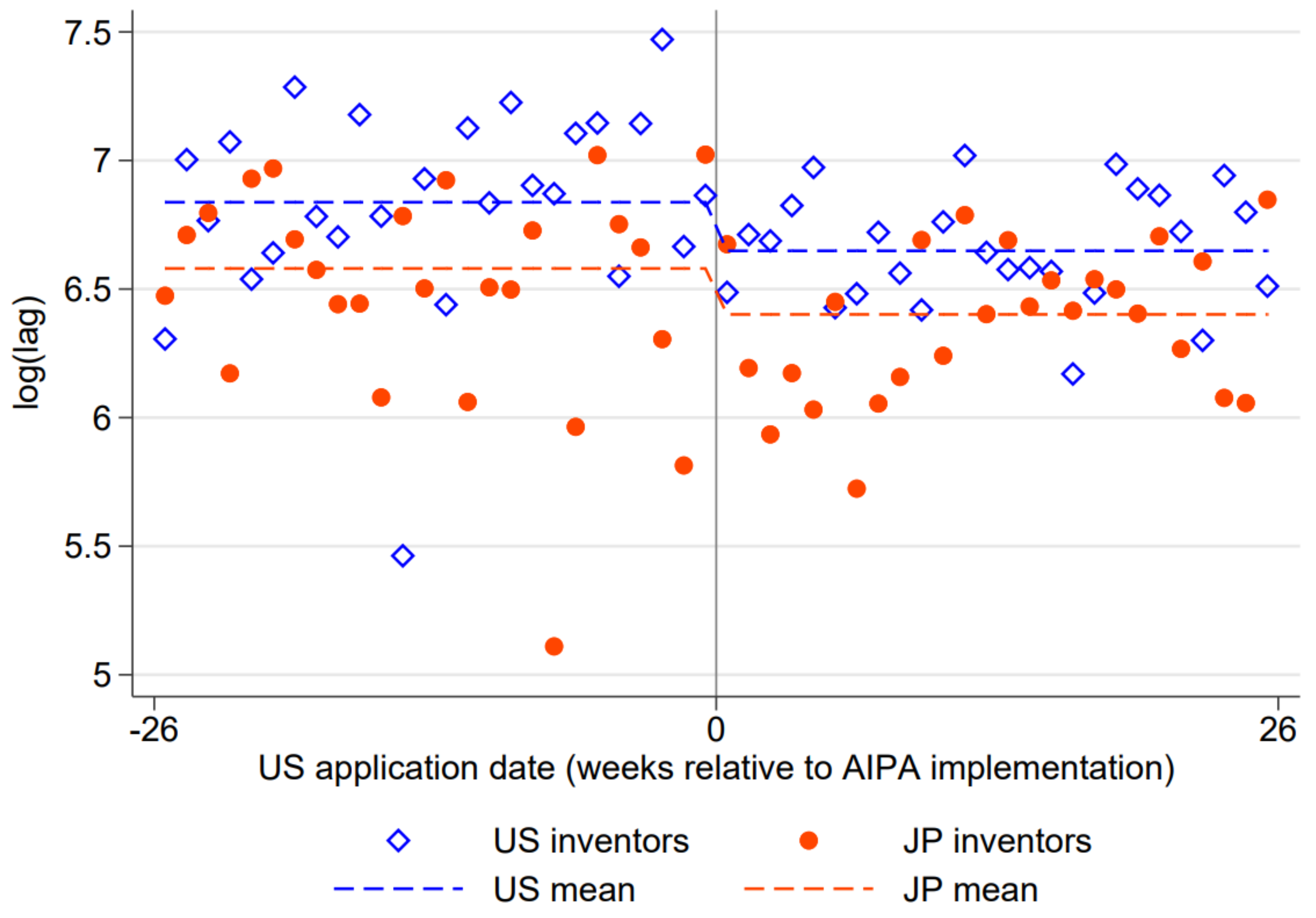
Results: Twin and triadic cohorts (H1)

H1: AIPA Significantly reduced the citation lag for citations from US inventors, relative to Japanese inventors

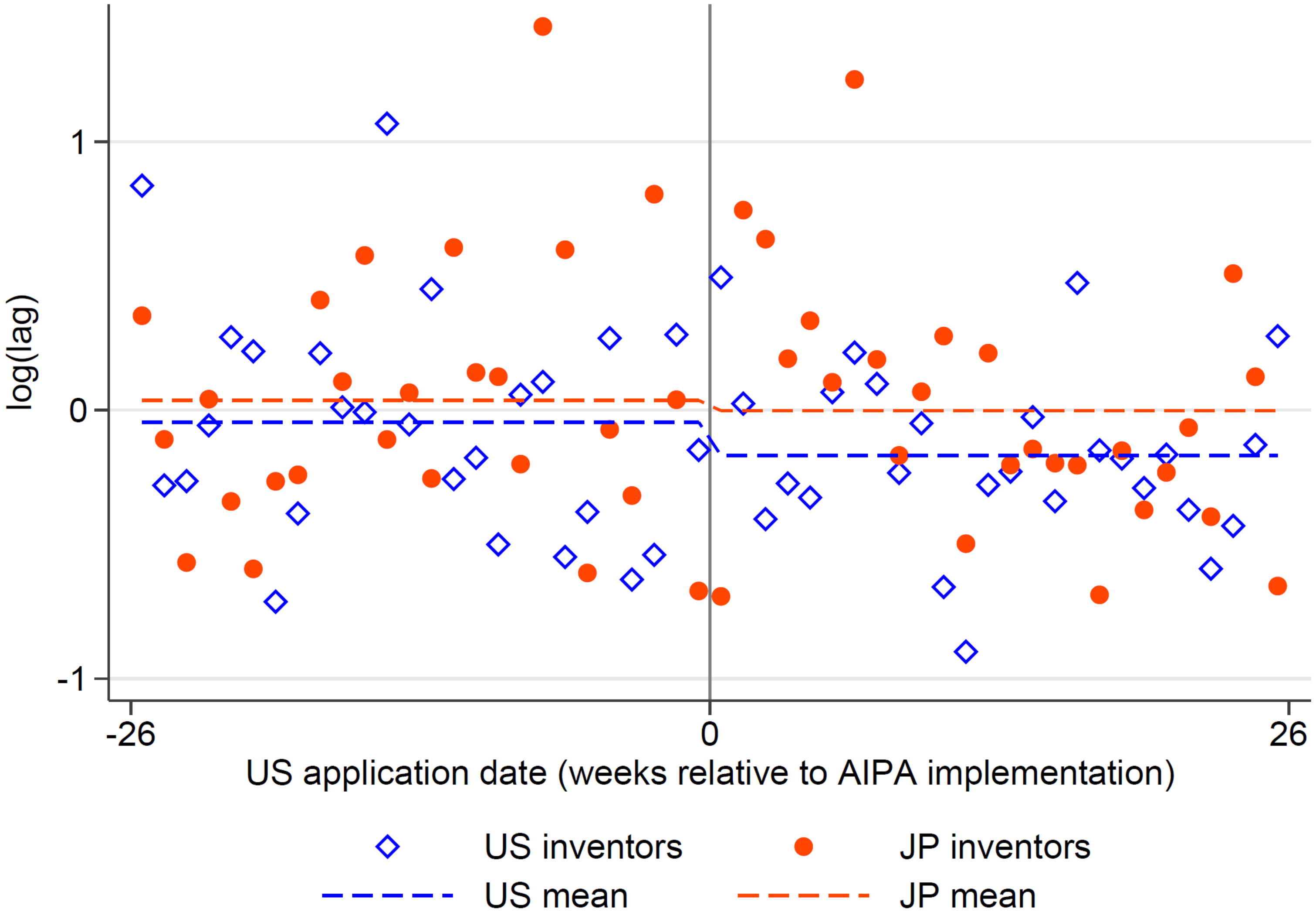
	(1) Twin Set	(2) Triadic Set	(3) Differenced
US × post-AIPA	-0.129**	-0.006	-0.016
	(0.013)	(0.386)	(0.235)
US × post-AIPA × Twin			-0.116**
			(0.050)
US assignee (citing)	0.277***	0.257***	0.262**
	(0.006)	(0.004)	(0.013)
Science dependence	0.006	-0.004	-0.006
	(0.914)	(0.881)	(0.760)
Originality	-0.013	0.017	-0.008
	(0.681)	(0.460)	(0.733)
Citing inventors	-0.057	-0.012	-0.038
	(0.564)	(0.850)	(0.638)
Constant	6.790***	6.120**	6.579**
	(0.007)	(0.022)	(0.022)
Observations	2770	1542	4312

Notes: Threshold on the number of backward citations made by citing patents is set to 100. p-values displayed in round brackets. *p<0.1; **p<0.05; ***p<0.01

Results: Raw first-citation lags (applicant, triadic)



Results: Differenced first-citation lags (app-triadic)



Results: Twin and triadic cohorts (H2)

H2: We observe the effects of H1 even controlling for the preferences of citing local prior art.

	(1) Twin Set	(2) Triadic Set	(3) Differenced
US × post-AIPA	-0.129**	-0.006	-0.016
	(0.013)	(0.386)	(0.235)
US × post-AIPA × Twin			-0.116**
			(0.050)
US assignee (citing)	0.277***	0.257***	0.262**
	(0.006)	(0.004)	(0.013)
Science dependence	0.006	-0.004	-0.006
	(0.914)	(0.881)	(0.760)
Originality	-0.013	0.017	-0.008
	(0.681)	(0.460)	(0.733)
Citing inventors	-0.057	-0.012	-0.038
	(0.564)	(0.850)	(0.638)
Constant	6.790***	6.120**	6.579**
	(0.007)	(0.022)	(0.022)
Observations	2770	1542	4312

Notes: Threshold on the number of backward citations made by citing patents is set to 100. p-values displayed in round brackets. *p<0.1; **p<0.05; ***p<0.01

Results: Examiner citations to twin cohort (H3)

H3: We will not observe a strong acceleration for examiner citations

	(1)
US × post-AIPA	0.022***
	(0.003)
US assignee (citing)	0.058
	(0.105)
Science dependence	-0.002
	(0.965)
Originality	0.022
	(0.544)
Citing inventors	0.009
	(0.920)
Constant	6.329**
	(0.016)
Observations	4360
Parallel trends p-value	0.092*

Results: Appropriability mediators, triple DID (H4)

H4: Largest impact on firms that were small and those less likely to be directly involved in, or otherwise familiar, with the Japanese innovation ecosystem.

M:	(1) R&D scale
$M \times$ US $\times$ post-AIPA	0.063*
	(0.094)
US $\times$ post-AIPA	-0.151**
	(0.015)
$M \times$ post-AIPA	-0.045*
	(0.058)
$M \times$ US	0.104*
	(0.057)
R&D scale (citing)	-0.153**
	(0.017)
US assignee (citing)	0.224**
	(0.019)
Observations	2770

Notes: Each mediator M in this table corresponds to a property of the citing patent, where M is described at the top of each column. Threshold on the number of backward citations made by citing patents is set to 100. p-values displayed in round brackets. *p<0.1; **p<0.05; ***p<0.01

Results: Appropriability mediators, split sample (H4)

H4: Largest impact on firms that were small and those less likely to be directly involved in, or otherwise familiar, with the Japanese innovation ecosystem.

	(1)	(2)
M:	R&D scale	Prior JPO app.
	Subset below median	***No prior JPO app.***
US × post-AIPA	-0.318**	-0.177***
	(0.020)	(0.007)
US assignee (citing)	0.232*	0.505***
	(0.063)	(0.002)
Observations	1020	720
	Subset above median	***Prior JPO app.***
US × post-AIPA	-0.035	-0.111**
	(0.199)	(0.045)
US assignee (citing)	0.356*	0.314**
	(0.057)	(0.024)
Observations	1102	2428

Results: Invention-specific mediators, triple DID (H5)

H5: The AIPA had the largest effect on citation timing for high-quality patents and those in fast-moving technical fields.

M:	(1) JPO citations	(2) Field Pace
$M \times$ US $\times$ post-AIPA	-0.130***	0.027
	(0.004)	(0.136)
US $\times$ post-AIPA	-0.133**	-0.145**
	(0.012)	(0.028)
$M \times$ post-AIPA	0.120*	-0.000
	(0.056)	(0.965)
$M \times$ US	0.174***	0.066**
	(0.010)	(0.011)
US assignee (citing)	0.282***	0.288**
	(0.008)	(0.020)
JPO citations (cited)	-0.182***	
	(0.003)	
Field pace		0.051**
		(0.015)
Observations	2770	2728

Notes: Each mediator M in this table corresponds to a property of the cited patent, where M is described at the top of each column. Threshold on the number of backward citations made by citing patents is set to 100. p-values displayed in round brackets. *p<0.1; **p<0.05; ***p<0.01

Results: Invention-specific mediators, split sample (H5)

H5: The AIPA had the largest effect on citation timing for high-quality patents and those in fast-moving technical fields.

M:	(1) JPO citations	(2) Field Pace
	Subset below median	
US × post-AIPA	-0.005	-0.244**
	(0.366)	(0.011)
US assignee (citing)	0.161**	0.307***
	(0.012)	(0.004)
Observations	1318	1350
	Subset above median	
US × post-AIPA	-0.228**	-0.045*
	(0.025)	(0.071)
US assignee (citing)	0.366***	0.261**
	(0.009)	(0.019)
Observations	1452	1378

1. Introduction
2. Background and Hypotheses
3. Empirical Setting and Data
4. Results
5. Conclusions

Conclusions

- We provide one of the first quantitative assessments of the role of language barriers on knowledge diffusion and the factors that mediate their impact
- We implement a DID framework to take advantage of a patent policy change that effectively removed a language barrier between the US and Japan

➢ For applicant citations, we observe a raw acceleration of about 13%, or about 6 months at the median, due to language barrier removal. This amounts to about half of the diffusion delay for the US inventors.

➢ Acceleration is much larger for knowledge contained in the highest quality patents, and for those existing within fast-paced technological ecosystems

➢ Acceleration is mostly confined to firms with low R&D intensities, and to those that were unlikely to be involved in the Japanese market before the policy change.

Limitations and Extensions

- **We do not attempt to measure the direct economic consequences of the AIPA through this channel**. Future work can assess, in detail, how the improved access to foreign knowledge helped smaller firms in the US via their growth or innovation outcomes
- The **sample is relatively small**. This is a result of our being highly restrictive in our sample construction to ensure any effects we find were interpretable and meaningful; personally, we do not see this as a limitation *per se*
- Boilerplate forward citations caveat: **citations are a noisy indicator of knowledge flow**. However, we note that many of our results (e.g., lack of effect on large firms) are consistent with the existence of a signal.

Parting words

- Pre-grant publication provides a **significant public good for cumulative innovation** through earlier translations of foreign patents
- The **impact of language barriers are heterogeneous** for different kinds of follow-on innovators, as well as for different kinds of knowledge
- We provide an **original framework** for studying language barriers in the context of knowledge diffusion. For example, one could use a similar framework to study knowledge spilling *out* of the US into countries where English is not widely spoken
- Machine translation has made significant progress in the past two decades. This work, alongside others, suggests that **ensuring this technology is useful and accessible to small firms is vital** for equal access to public knowledge in foreign languages.

Thank you for your attention!

“Language Barriers and the Speed of Knowledge Diffusion”,

RIETI Discussion Paper Series 22-E-074

tinyurl.com/HighamRIETI