# The Contribution of High-Skilled Immigrants to Innovation in the United States

Shai Bernstein [1] Rebecca Diamond [2] Abhisit Jiranaphawiboon [2]
Tim McQuade [3] Beatriz Pousada [2]

[1]Harvard Business School [2]Stanford University

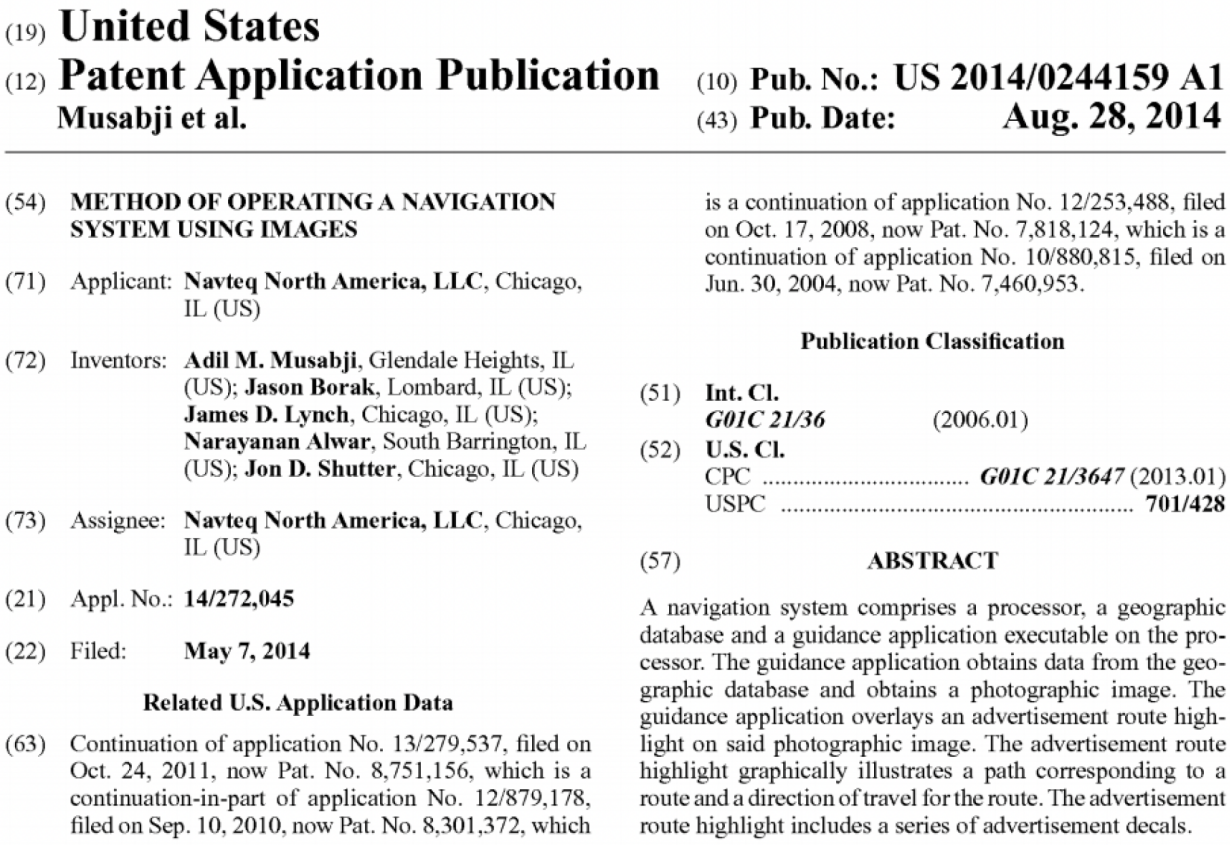
[3]Berkeley

November 1, 2024

# How important are immigrants to US innovation?

- Innovation, technology progress: key drivers of growth (Romer (1990); Aghion and Howitt (1992))
- Suggestive evidence immigrants large contributor to US innovation:
  - 23% of STEM occupations (ACS, 2016)
  - 26% of US-based Nobel winnners
- **What is the contribution of Immigrants to Innovation in the U.S.?**
  - Despite large interest, data constraints on individual immigration status limits ability to answer this question
  - Little is known about immigrants' **indirect effects** on US innovation:
    - Collaborations between immigrants and US-born inventors may allow knowledge sharing, jointly increasing innovation of both groups (Moser et al. (2014))

# Empirical Challenges



(19) **United States**

(12) **Patent Application Publication**

**Musabji et al.**

(10) **Pub. No.: US 2014/0244159 A1**

(43) **Pub. Date: Aug. 28, 2014**

(54) **METHOD OF OPERATING A NAVIGATION SYSTEM USING IMAGES**

(71) Applicant: **Navteq North America, LLC**, Chicago, IL (US)

(72) Inventors: **Adil M. Musabji**, Glendale Heights, IL (US); **Jason Borak**, Lombard, IL (US); **James D. Lynch**, Chicago, IL (US); **Narayanan Alwar**, South Barrington, IL (US); **Jon D. Shutter**, Chicago, IL (US)

(73) Assignee: **Navteq North America, LLC**, Chicago, IL (US)

(21) Appl. No.: **14/272,045**

(22) Filed: **May 7, 2014**

**Related U.S. Application Data**

(63) Continuation of application No. 13/279,537, filed on Oct. 24, 2011, now Pat. No. 8,751,156, which is a continuation-in-part of application No. 12/879,178, filed on Sep. 10, 2010, now Pat. No. 8,301,372, which is a continuation of application No. 12/253,488, filed on Oct. 17, 2008, now Pat. No. 7,818,124, which is a continuation of application No. 10/880,815, filed on Jun. 30, 2004, now Pat. No. 7,460,953.

**Publication Classification**

(51) **Int. Cl.**
***G01C 21/36*** (2006.01)

(52) **U.S. Cl.**
CPC ..... ***G01C 21/3647*** (2013.01)
USPC ..... **701/428**

(57) **ABSTRACT**

A navigation system comprises a processor, a geographic database and a guidance application executable on the processor. The guidance application obtains data from the geographic database and obtains a photographic image. The guidance application overlays an advertisement route highlight on said photographic image. The advertisement route highlight graphically illustrates a path corresponding to a route and a direction of travel for the route. The advertisement route highlight includes a series of advertisement decals.

# We bring new data to these questions

**Utilize novel data and a new approach to identify individuals' immigrant status:**

- Infutor: address aggregator covering 300 million adults over past 30 years (Diamond et al. 2018).
  - Provides information on names, year of birth, address history, SSN.
- A combination of SSN and year of birth identifies immigrants:
  - Infer SSN assignment year from first 5-digits of SSN
  - Natives assigned SSN during their youth
  - Identify immigrants who received SSN after 19th birthday
  - Cross-sectional variation in immigrant shares across US counties has $R^2$ of around 90%.
- Merge immigrant status with the universe of patents (1990-2016) using name/address: link 70% of inventors

# Research Questions

**Our goal is to "take stock" of current role of immigrants in US innovation**

1. Describe productivity differences between immigrants and natives
   - Do immigrants/natives have different life-cyle innovation output?
   - Do immigrants bring to the U.S. unique knowledge and skills?

2. Quantify indirect effects of immigrants/natives through collaboration externalities
   - Estimate causal effects of immigrant/natives on immigrant native co-authors
   - Use early death of prior collaborator, estimate impact on co-authors subsequent patent production (Jaravel et al. (2018))

3. Quantify aggregate contribution of immigrants to US innovation, accounting for external effects on native innovation
   - Estimate innovation production function using collaborator death estimates

# Data

1. **Innovation**
   - NBER Data Project and Google Patents
   - Roughly 800 thousand inventors from 1990-2016
2. **Infutor Database**
   - Aggregator of address history data using many sources, including phonebooks, magazine subscriptions, and credit header files.
   - Complete address history for more than 300 million US residents (78% of US pop)
   - Includes complete names, age, gender, and SSN.

Patent data were linked to infutor data by inventor name, city, and state at time of patent application, with 70% match rate.

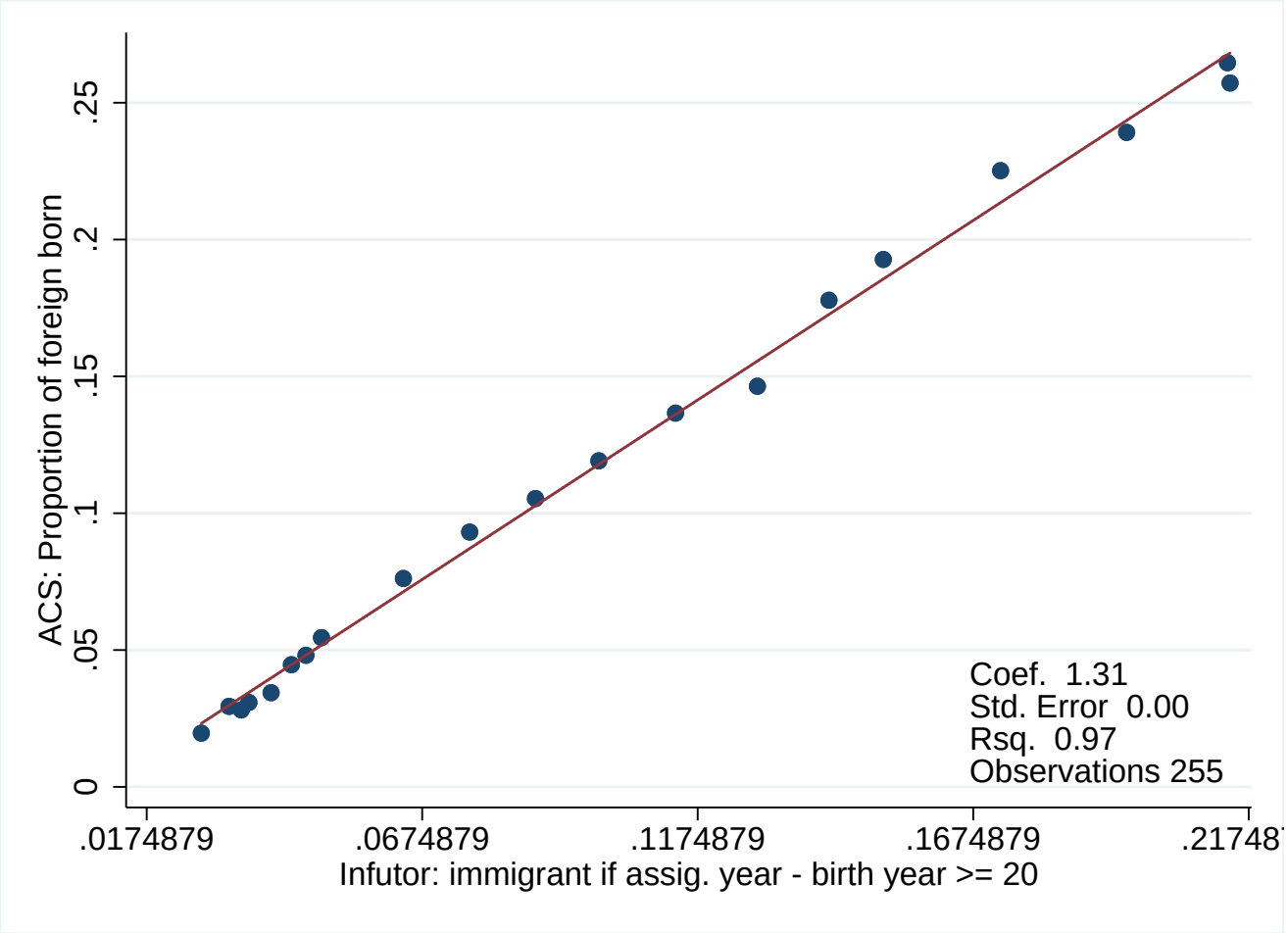
# Identifying Immigrants

- The SSN structure (between 1937-2011):

$$\underbrace{XXX}_{\text{area number}} - \underbrace{XX}_{\text{group number}} - \underbrace{XXXX}_{\text{serial number}}$$

- SSA provides the month and year in which group numbers began to be issued ("High Group List")
  - Designed to validate issued SSNs and prevent fraud
- We verify accuracy and consistency of group numbers based on latest year of birth within a group number
- Classify individuals as immigrants if received SSN after age 20 (or higher)

# ACS Validation: State-level immigrant proportions (45-50 years old)



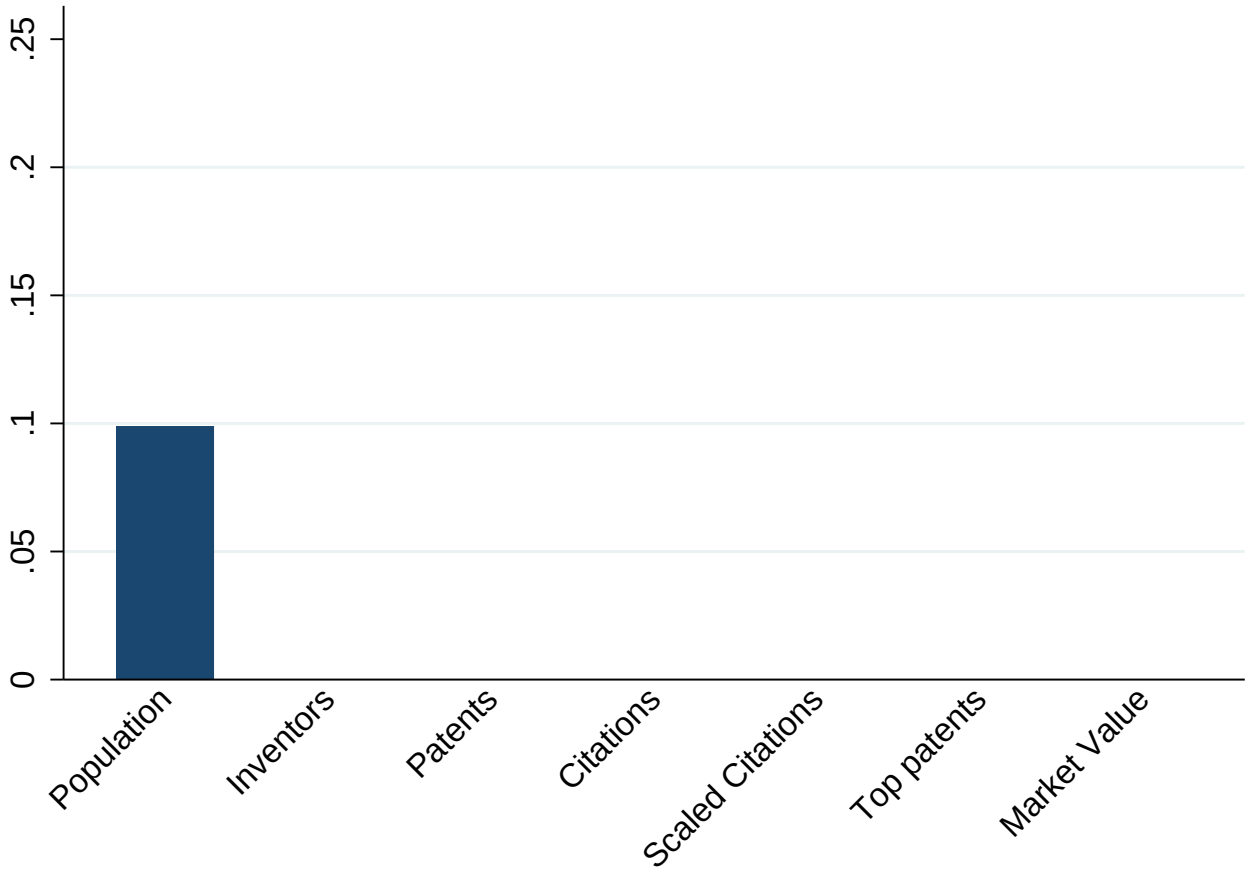

Note: $ACS_{i,2000} = \alpha + \beta Infutor_{i,2000} + \varepsilon_{i,2000}$, immigrants are classified for assignment age>=20. ACS has roughly 31% more immigrants, when focusing on 45-50 age group. Department of Homeland Security estimates that 34% of immigrants were illegal in 2014.

# Measuring Innovation

## Widely used measures in the literature

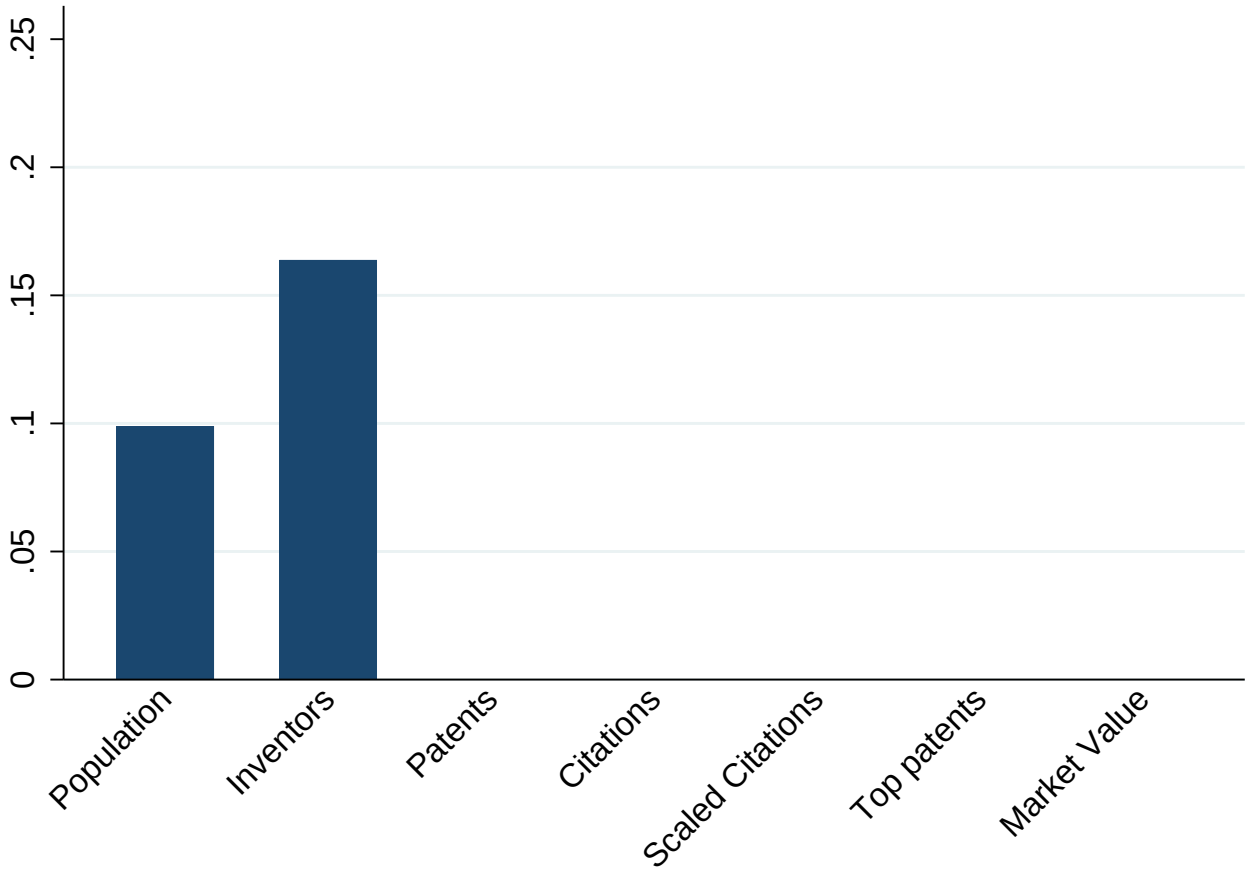
- **Innovative scale** - number of granted patents
- **Patent quality**
  1. Citation counts per patent - proxy for technological contribution (Hall et al. 2005, Kogan et al. 2017)
     - Rely on 3-year citation horizon (or 5-year)
     - Scale citation counts by the average number of cites within a given technology class and year.
  2. Top patents - top 10% of citations within tech class and year
- **Economic importance of patents** - stock market reaction (Kogan et al. 2017), imputed for private firms

# Immigrants Contribution to Innovation 1976-2012



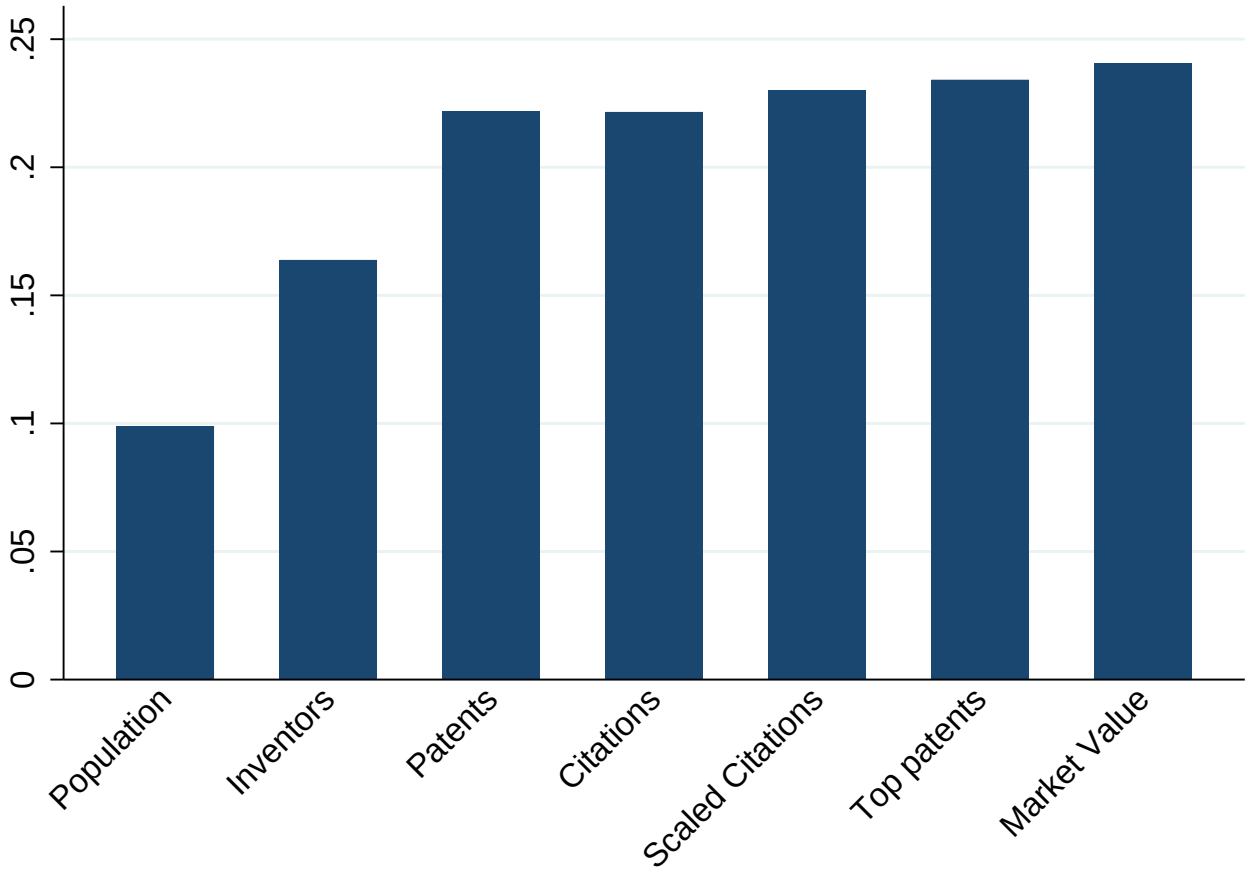

Note: Innovation shares are apportioned by team size. Relative to immigrant inventors share in the population, they contribute 40% more in terms of scaled citations, and 47% more in terms of economic value of innovation (based on Kogan et al. 2017).

# Immigrants Contribution to Innovation 1976-2012




Note: Innovation shares are apportioned by team size. Relative to immigrant inventors share in the population, they contribute 40% more in terms of scaled citations, and 47% more in terms of economic value of innovation (based on Kogan et al. 2017).

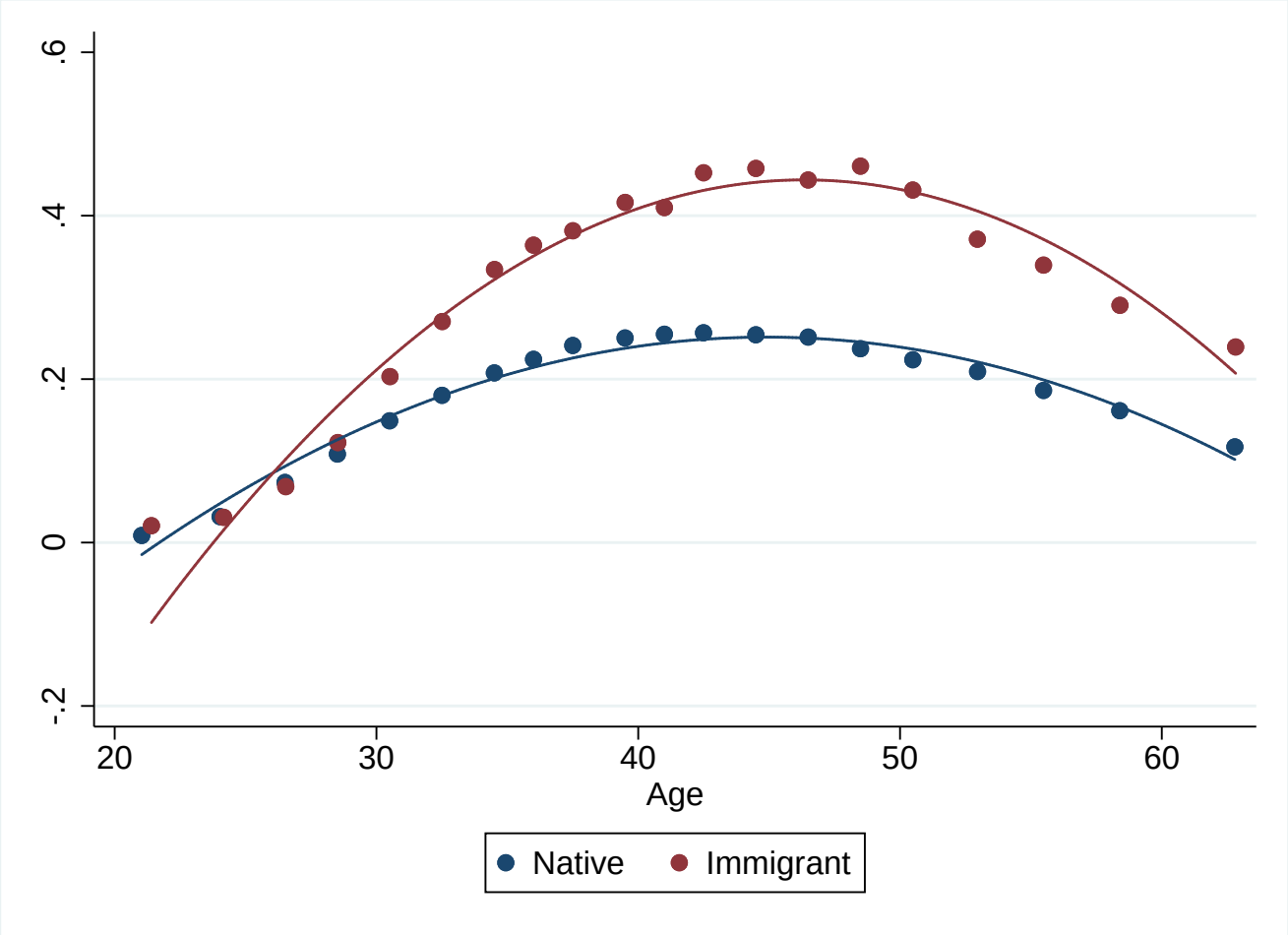
# Immigrants Contribution to Innovation 1976-2012




Note: Innovation shares are apportioned by team size. Relative to immigrant inventors share in the population, they contribute 40% more in terms of scaled citations, and 47% more in terms of economic value of innovation (based on Kogan et al. 2017).

# Productivity over the Life-Cycle

## Annualized Adjusted Citations Weighted Number of Patents




Note: Binscatters plot relationship between innovative productivity measure and inventors age

# Attempting to Explain Productivity Gaps

## Annualized Number of Patents

**Panel A: Annual Number of Patents**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Immigrant | 0.093*** | 0.092*** | 0.071*** | 0.062*** | 0.071*** | 0.063*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Observations | 15,714,917 | 15,714,917 | 15,714,906 | 15,192,932 | 15,709,593 | 15,187,669 |
| Year FE | yes | yes | yes | yes | yes | no |
| YOB FE | no | yes | yes | yes | yes | yes |
| County FE | no | no | yes | no | no | no |
| County × Tech FE | no | no | no | yes | no | no |
| County × Year FE | no | no | no | no | yes | yes |
| Tech × Year FE | no | no | no | no | no | yes |

Note: gap remains even after the addition of granular controls

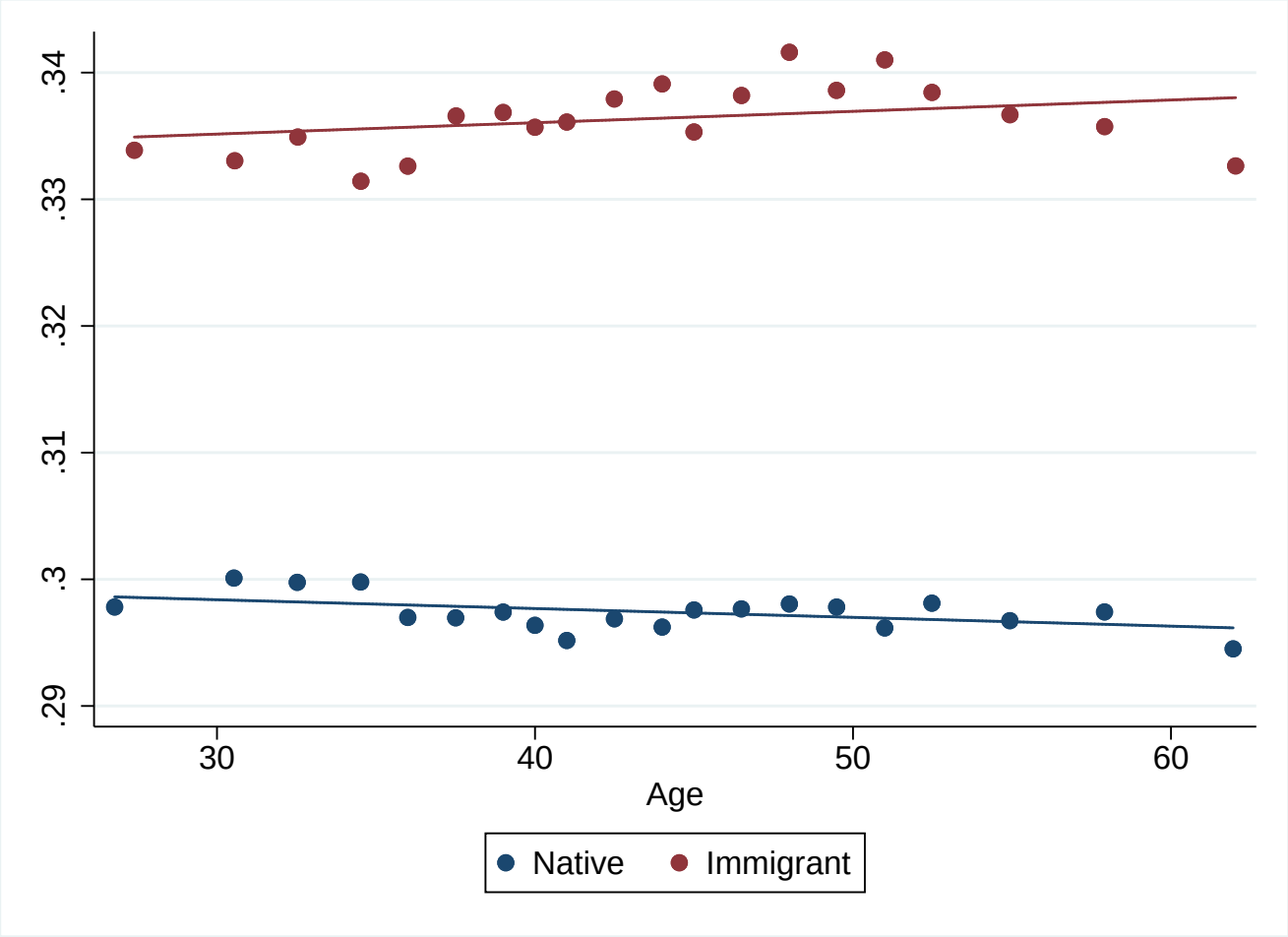
# Global Knowledge Markets

**Question: Do immigrants bring unique knowledge to US innovation markets?**

- Immigrants may be well-positioned to import global knowledge and integrate foreign ideas with US-based knowledge
- Theories of human capital accumulation suggest that new ideas can be generated as novel combinations of existing material
  - Usher, 1954; Becker, 1982;Weitzman, 1998

# Global Knowledge Markets

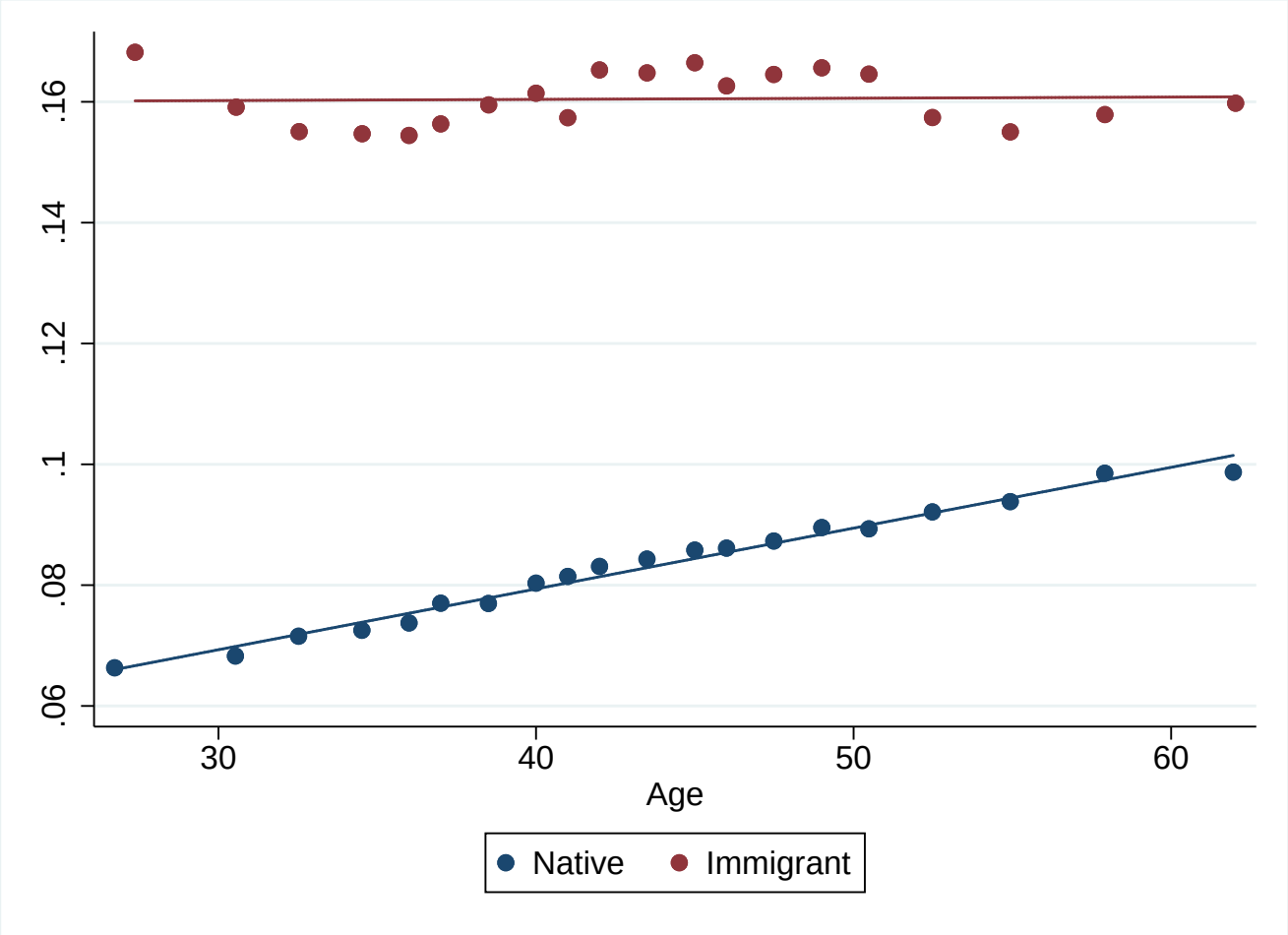
## Share of Foreign Backward Citations




Note: Reliance on foreign technologies in own patent production

# Global Knowledge Markets

## Share of Patents Collaborated with Foreign Inventors




Note: Collaboration with foreign inventors in own patent production

# Collaboration Spillovers

- Do immigrants create significant innovation spillovers on US-born innovators, increasing total US innovation?
- **Challenge:** endogenous team formation and collaborative research efforts
- Ideal experiment:
  - Explore implications of an exogenous end to an on-going collaboration
  - Do such disruptions lead to differential effects when triggered by immigrants versus natives?

# Identification Strategy

- Identification strategy:
  - Exploit pre-mature deaths - occur before or at the age of 60
  - Causal effect is identified through a difference-in-differences setting

- Data Construction
  - Deceased inventors are identified from the Social Security Death Master File (DMF)
  - Match 1-1 placebo deceased inventors based on:
    - Immigrant status, age at death, the cumulative number of patent applications, year of death, and cumulative number of coauthors, grouped into ventiles.
  - There are 9,405 real deceased inventors and placebo deceased inventors.

# Matched Sample Comparisons

**Panel A: Real vs. Placebo Deceased Demographics**

| | Real Deceased | | | Placebo Deceased | | |
|---|---|---|---|---|---|---|
| | Mean | Median | Std. Dev. | Mean | Median | Std. Dev. |
| Age | 51.13 | 53 | 7.05 | 51.13 | 53 | 7.05 |
| Year | 2,004.42 | 2,005 | 4.89 | 2,004.42 | 2,005 | 4.89 |
| Immigrant status | 0.10 | 0 | | 0.10 | 0 | |
| Cumulative patents | 3 | 2 | 2.65 | 3 | 2 | 2.65 |
| Co-authors | 3.45 | 2 | 4.35 | 3.18 | 2 | 3.51 |
| Team size | 3.37 | 3 | 2.20 | 3.32 | 3 | 2.49 |
| Adjusted citations | 3.97 | 1.19 | 9.38 | 3.72 | 1.05 | 10.74 |
| Top patents | 0.50 | 0 | 1.40 | 0.47 | 0 | 1.48 |
| Economic Value | 76.11 | 23.11 | 265.43 | 64.50 | 20.79 | 189.51 |
| Female | 0.07 | 0 | | 0.10 | 0 | |
| Sample size | 3,947 | | | 155,711 | | |

# Matched Sample Comparisons - Co-authors

**Panel B: Real vs. Placebo Co-Inventor Characteristics**

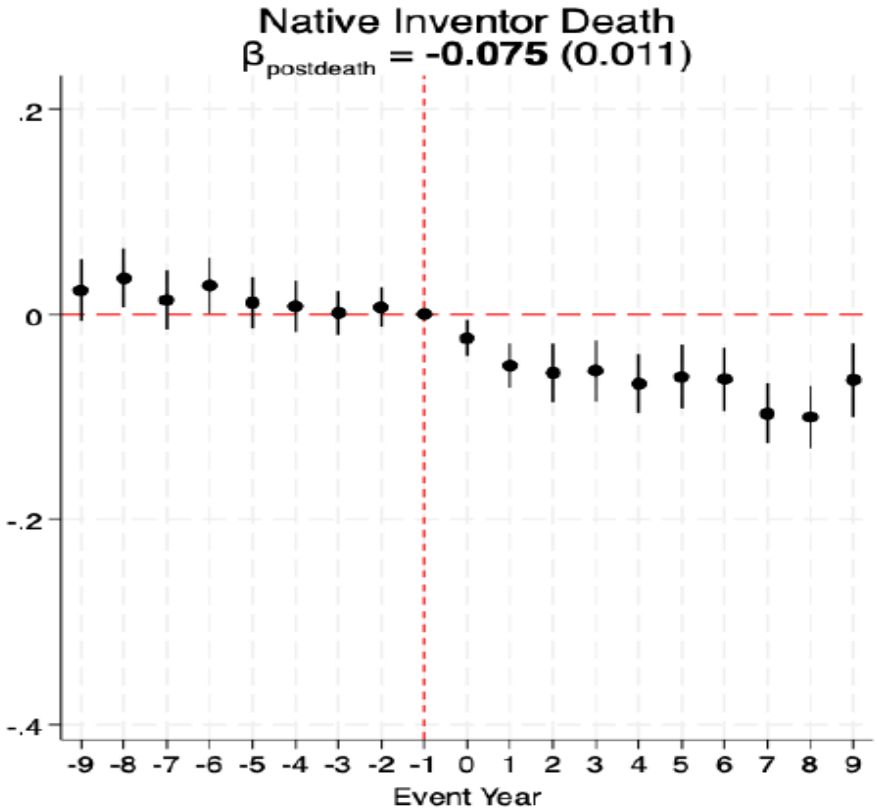
| | Real Deceased | | | Placebo Deceased | | |
|---|---|---|---|---|---|---|
| | Mean | Median | Std. Dev. | Mean | Median | Std. Dev. |
| Age | 48.37 | 49 | 16.61 | 46.56 | 48 | 20.58 |
| Immigrant status | 0.15 | 0 | | 0.20 | 0 | |
| Co-patents pre-treat | 1.91 | 1 | 2.26 | 1.87 | 1 | 2.26 |
| Cumulative patents | 8.63 | 3 | 20.46 | 6.95 | 3 | 16.06 |
| Adjusted citations | 12.58 | 3.24 | 35.70 | 10.07 | 2.39 | 29.09 |
| Top patents | 1.65 | 0 | 4.79 | 1.30 | 0 | 3.80 |
| Economic Value | 209.40 | 49.97 | 585.34 | 165.03 | 39.44 | 539.83 |
| Female | 0.10 | 0 | | 0.11 | 0 | |
| Sample size | 15,471 | | | 369,509 | | |

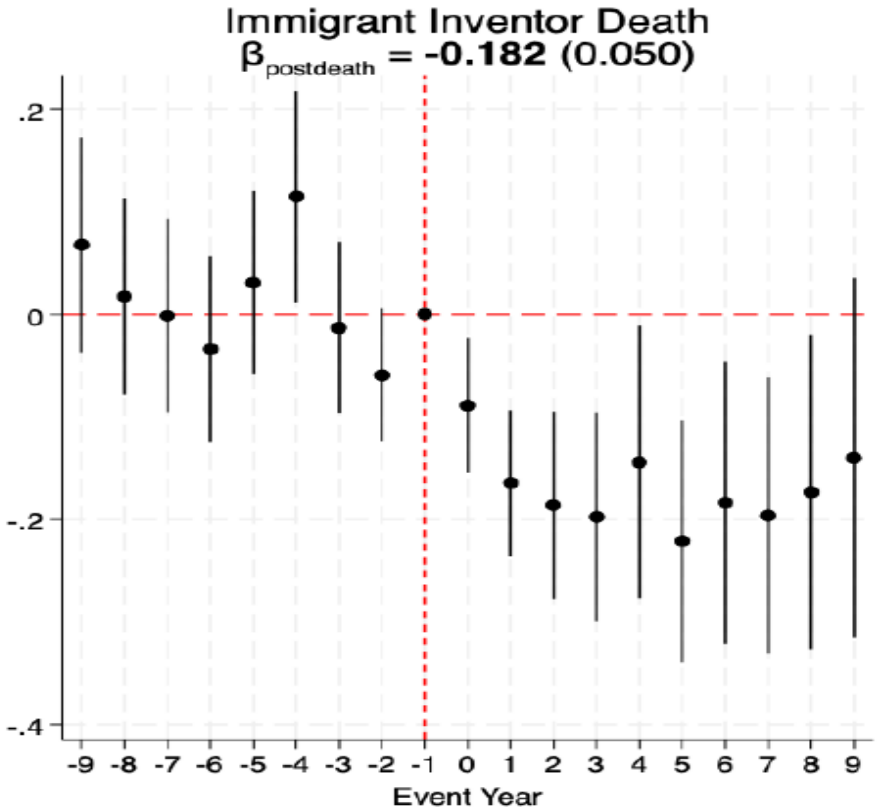
# Treatment Effects

| Dependent Variable: | Number of Patents | | | Adjusted Citations | | |
|---|---|---|---|---|---|---|
| | All | Immigrant | US-born | All | Immigrant | US-born |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Post Death | -0.087*** | -0.182*** | -0.075*** | -0.154*** | -0.337*** | -0.130*** |
| | (0.011) | (0.050) | (0.011) | (0.015) | (0.063) | (0.015) |
| Control Post Mean | 0.850 | 1.111 | 0.815 | 1.049 | 1.465 | 0.993 |
| Percent Change | -10.3% | -16.3% | -9.2% | -14.7% | -23.0% | -13.1% |
| Match Group × Event Year FE | yes | yes | yes | yes | yes | yes |
| Individual FE | yes | yes | yes | yes | yes | yes |
| $R^2$ | 0.560 | 0.573 | 0.557 | 0.366 | 0.319 | 0.377 |
| Number of Deceased Inventors | 159,658 | 8,017 | 151,641 | 159,658 | 8,017 | 151,641 |
| Observations | 6,769,647 | 502,103 | 6,267,544 | 6,769,647 | 502,103 | 6,267,544 |

# Positive Externalities and Spillovers

## Annualized Number of Patents




Note: Specifications include co-inventor and year fixed effects:

$$Y_{it} = \sum_{k=-9}^{9} \beta_k^{real} 1_{L_{it}^{real}=k} + \sum_{k=-9}^{9} \beta_k^{all} \mathbb{1}_{L_{it}^{all}=k} + \sum_{m=1976}^{2012} \gamma_m \mathbb{1}_{t=m} + \alpha_i + \epsilon_{it}$$

# Mechanisms Driving Immigrant Effects

**What drives the gap between spillovers of immigrants and natives?**

- Control for additional dimensions of observable treatment effect heterogeneity. Can we explain the gap?
- Treatment effect heterogeneity based on: inventor's age, year of death, cumulative patents and citations, average coauthors' ages, average coauthors' cumulative patents and citations, collaboration recency, number of unique prior coauthors, number of co-patents, similarity in prior work, number of local patents per-capita.

**Finding:**

Gap between immigrants and natives grows slightly larger suggests there is something unique to immigrants

# Controlling for Heterogeneity

| Dependent variable: | Number of Patents (1) | Scaled Citations (2) | Top Patents (3) | Economic Value (4) |
|---|---|---|---|---|
| **Panel A. Differentials *Before* Controlling for Treatment Effect Heterogeneity** | | | | |
| | -0.1065 | -0.2072 | -0.0547 | -1.4923 |
| | (0.0496) | (0.0620) | (0.0129) | (0.8085) |
| **Panel B. Differentials *After* Controlling for Treatment Effect Heterogeneity** | | | | |
| | -0.1584 | -0.2425 | -0.0649 | -1.6927 |
| | (0.0491) | (0.0623) | (0.0126) | (0.7904) |
| **Panel C. Absolute Difference between Estimates in Panels A and B** | | | | |
| | 0.0518 | 0.0352 | 0.0101 | 0.2004 |
| | (0.0136) | (0.0159) | (0.0035) | (0.2335) |

# Mechanisms: Team-Specific vs. General Human Capital

**Findings suggest that immigrants have substantial contributions to US innovation:**

- Both directly through their own output and indirectly through positive externalities/spillovers onto their collaborators.

**Team vs General Human Capital:**

- Lost productivity from destroying team that would have produced patents **(Team Specific)**
- Lost productivity even among teams that never involved dying inventor **(General**
- Use simple structural framework to decompose these channel, quantify aggregate immigrant contribution

# Decomposing Immigrants Contribution

- Inventor i at time t has general human capital $h_{it}$:

$$h_{it} = \underbrace{A_{it}}_{\text{ability-direct}} \underbrace{(1 + N_{it,nat})^{\beta_{nat}}}_{\text{native-spillovers}} \underbrace{(1 + N_{it,imm})^{\beta_{imm}}}_{\text{immigrant-spillovers}}$$

# Decomposing Immigrants Contribution

- Inventor i at time t has general human capital $h_{it}$:

$$h_{it} = \underbrace{A_{it}}_{\text{ability-direct}} \underbrace{(1 + N_{it,nat})^{\beta_{nat}}}_{\text{native-spillovers}} \underbrace{(1 + N_{it,imm})^{\beta_{imm}}}_{\text{immigrant-spillovers}}$$

- Team $j$ made up of a set of $J$ members at time $t$. Team-specific human capital $h_{jt}$ :

$$h_{jt} = \left( \prod_{i \in J_j} h_{it} \right)^{\frac{1}{|J_j|}}$$

# Decomposing Immigrants Contribution

- Inventor i at time t has general human capital $h_{it}$:

$$h_{it} = \underbrace{A_{it}}_{\text{ability-direct}} \underbrace{(1 + N_{it,nat})^{\beta_{nat}}}_{\text{native-spillovers}} \underbrace{(1 + N_{it,imm})^{\beta_{imm}}}_{\text{immigrant-spillovers}}$$

- Team $j$ made up of a set of $J$ members at time $t$. Team-specific human capital $h_{jt}$ :

$$h_{jt} = \left( \prod_{i \in J_j} h_{it} \right)^{\frac{1}{|J_j|}}$$

- Innovation output at time $t$ by team $j$, $Y_{jt}$ :

$$Y_{jt} = \left(1 - \beta^{death} \mathbb{1}_{jt}^{death}\right) h_{jt} g(|J_j|) \varepsilon_{jt}$$

- $N_{it,k}$ :number of unique prior co-authors
- $A_{it}$ :individual idiosyncratic innovative productivity
- $\beta_m$ general human capital productivity spillovers from collaboration
- $\beta^{death} \mathbb{1}_{jt}^{death}$: team-specific human capital loss from death
- $g(|J_j|)$ : productivity benefit of team size
- $\varepsilon_{jt}$ : team-specific idiosyncratic productivity

# General Human Capital Effects

- Taking first order taylor expansion around base mean:

$$\underbrace{\left(\frac{Y_{jt} - \bar{Y}}{\bar{Y}}\right)_{j \in \mathcal{J}_1}}_{\text{change in output due to death}} = \beta_{nat} \underbrace{\left(\frac{N_{j,nat} - \bar{N}_{nat}}{1 + \bar{N}_{nat}}\right)_{j \in \mathcal{J}_1}}_{\text{change in native coauthors}} + \beta_{imm} \underbrace{\left(\frac{N_{i,imm} - \bar{N}_{imm}}{1 + \bar{N}_{imm}}\right)_{j \in \mathcal{J}_1}}_{\text{change in immigrant coauthors}}$$

- Only study teams $\mathcal{J}_1$, that do not contain dying/placebo dying inventor. Impact only through general human capital
- Changes in general human capital reflect changes in number of (living) co-authors (native/immigrant) production function parameters
- Estimate changes in number of immigrant/native co-authors due to premature death of an immigrant/native.
- Repeat estimate for teams containing dying inventor. Excess productivity loss here due to team-specific human capital

# Model Parameters

Panel A. Parameters

| Parameter: | $\beta^{imm}$ | $\beta^{nat}$ | $\beta^{death}$ |
|---|---|---|---|
| | (1) | (2) | (3) |
| | 0.718 | 0.290 | 0.166 |
| Governing: | human capital | human capital | inventor death |
| Inventor: | immigrant | US-born | any |

- Team-Specific human capital loss only 16% larger than general human capital effects

# Decomposing Immigrants Contribution

**Question:** What is the overall contribution of immigrants on US Innovation?

We use simple structural framework to decompose immigrant contribution:

- US-born production without immigrants account for **60%** of aggregate innovation
- Immigrant output without US-born accounts for **15%** of aggregate innovation
- Credit immigrants for increase in US-born output due to immigrant collaboration: Immigrants account for **32%** aggregate innovation
- Credit US-born for increase in immigrant output due to US-born collaboration: US-born account for **68%** of aggregate innovation

# Conclusion

Using novel approach to identify immigrant status we find that:

1. Immigrant contribution to innovation is disproportionate relative to their share of the inventor population
2. Immigrants generate significant spillovers on native inventors. Immigrants create larger spillovers
3. 32% of aggregate US innovation since 1990 can be attributed to immigrants, 39% more than their observed direct output
4. Expanding high-skill immigration with unique knowledge backgrounds likely increases US innovation overall