Supplemental Appendix

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

By Shai Bernstein, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada

Matching Algorithm of Patent Data with Infutor

In this section, we discuss how we construct our inventor identifiers in the USPTO patent dataset. The idea is to use individuals' address histories in Infutor to verify the address at which each inventor lived during the patent application and track inventors who moved. Precisely, in the patent data, we define "identifiers" (ID) as a unique combination of the inventor's first name, last name, city name, and the state in which he lived. In the raw USPTO data, we begin with 14,991,282 patents granted worldwide and restrict to 7,350,977 patents granted in the U.S and then to 7,342,269 patents for which state and city name are not missing. Since the Infutor dataset spans 1990 to 2016, we necessarily restrict the patent data to these years only. At the start, there are 6,229,618 patents filed and ultimately granted to 1,351,024 unique IDs. We discuss our disambiguation steps in detail as follows.

We begin by matching each observation in the patent data to address histories in Infutor. Our matching criterion is such that state, city, last name, and the first three letters of first name exactly match between the two datasets: 1,034,288 unique IDs have at least one Infutor match at this stage. In this matched subset, each ID can still map to multiple individuals in Infutor. Thus, we successively apply disambiguation restrictions, where we identify unique matches in a given step and then remove them from the pool before applying the next restriction.

First, we impose that first names match exactly and that middle initials do not conflict between the two datasets (i.e., they agree or at least one or both are missing). Now, given typographical inconsistencies in first names observed between the two datasets, we allow first names to match weakly. Second, we impose that first names can contain each other (e.g., "Timothy" and "Tim") and that middle initial does not conflict. Third, we allow for alternate first names (e.g., "Richard" and "Rick") and for minor misspellings (e.g., "Stephen" and "Steven"'), while maintaining that middle initials do not conflict. In these steps, we identify 876,431 IDs, each mapping exactly to one individual in Infutor: 680,252, 144,434, and 51,745 IDs from the first, second, and third steps, respectively.

Next, we disambiguate the remaining 157,797 IDs for which state, city, last name, and the first three letters of first name match precisely, but for which we cannot find unique Infutor matches in the first three steps from above. In each of these steps, we always condition on observing at least one patent for which the application year falls between the beginning and ending address years (allowing plus or minus two years). In addition to this condition, in each step, we successively apply a stricter matching criterion. IDs are identified as unique matches at the end of each step if we only find exactly one Infutor match. For brevity, we define the following terms. Middle initials "match strictly" if both are non-missing and agree across the two datasets and " match weakly" if at least one or both are missing. First names "match strictly" if they agree across the two datasets and "match weakly" if one contains the other, is an alternate name for the other, or contains minor misspellings by at most two characters.

In the fourth step, we require that middle initials match weakly and that first names match weakly. The fifth step imposes that first names match strictly, while maintaining middle initials match weakly. In the sixth step, we require that middle initials match strictly and that first names match weakly. The seventh step imposes that both first names and middle initials match strictly. The eighth

step considers the cases where both middle names (rather than middle initials) and first names match exactly across the two datasets. We identify 51,666 IDs with unique Infutor matches in these steps: 30,714, 14,457, 2,442, 652, and 3,401, respectively. These first eight disambiguation steps together identify 928,097 IDs.

Finally, we turn to the 316,736 IDs in the patent data for which we cannot find matches in Infutor using precise matches on state, city, last name, and the first three letters of the first name. To do so, we match these observations to Infutor using exact matches on state, city, and last name, completely ignoring first name. In essence, these observations are those with inconsistent first three letters of first name. Then, we condition these potential matches on a strict middle initial match and a weak first name match (as defined above). This final step yields 3,529 IDs.

In summary, out of the 1,351,024 IDs in the patent data, we find 931,629 IDs with unique Infutor matches, indicating a match rate of 69.0% and corresponding to 879,989 unique individuals in Infutor. We summarize all of these data construction steps in Table A1 below.

A1. Validation Tests

We begin by comparing the proportion of county-level immigrants based on the entire Infutor dataset and our new classification methodology to the proportion of foreign born individuals at the county level in the 2000 Census.[1] To do so, we first geocode individuals in the Infutor dataset to US counties based on their exact 2000 street address. From this mapping and our immigrant classification procedure, we then calculate the immigrant proportion of the 2000 county population. We perform this calculation several times as we apply different SSN assignment cutoffs between ages of 20 to 25. We finally run regressions of the proportion of foreign born individuals as measured by the Census on our constructed measures. In each regression, we use the 2000 population size as reported by the 2000 Census as weights.

[1] The 2010 CENSUS does not have the proportion of immigrants at the county level.

Figure A4 in the Appendix reports the R^2 of these regressions. The x-axis denotes the minimum gap between the SSN assignment year and birth year that is required to classify an individual as an immigrant. Comfortingly, all of our specifications produce R^2 of approximately 90%. This test illustrates that our immigrant classification procedure captures well the cross-sectional variation in immigrant shares across US counties. Figure A5 provides binscatters of these regressions. While we match the cross-sectional variation extremely well, these results also illustrate that, on average, the proportion of foreign born in a county according to the 2000 census is slightly above 1.5 times the proportion of immigrants predicted by our method. This is expected, however, because the Infutor data only contains adults and legal immigrants, while the Census counts all age groups as well as undocumented immigrants.[2]

To explore whether our immigrant classification method can do even a better job in explaining variations of immigrants shares when we focus on adults only we use the ACS. The ACS allows us to not only incorporate individuals age but also, importantly, identify the age in which immigrants arrived to the US. In principle, this allows us to identify in the ACS exactly those immigrants we propose to identify in Infutor. Due to confidentiality restriction, we cannot work with the data at the county level. To have a representative sample at each age, we use the ACS at the state level rather than at the county level and calculate the proportion of the population that is both foreign born and immigrated after they had reached 20 years of age. Similar to what we did previously, we then regress the proportion of the state population of a certain age that is both foreign born and immigrated after the age of 20, as reported by the ACS, against the same statistic constructed through Infutor.

[2]In Figure A6 in the Appendix we plot the combined R^2 and regression coefficients for age thresholds between 10 years old to 30. As expected, the lower the age threshold, the lower the regression coefficient, implying that the share of foreigners, based on this classification is increasing, as we classify younger and younger individuals as immigrants. However, it is important to note the changes in the R^2. As we approach the age threshold of 20, our ability to explain variations in immigrants across counties increases, and stabilizes around the age of 20, consistent with the notion that around that age threshold we are indeed able to separate immigrant and US-born individuals based on the age in which they received their social security number.

Figure A7 illustrates the fit of these regressions through binscatters using the 2005 ACS for several adult age groups. For example, panel (a) provides the binscatter for adults in ages of 40-44. The R^2 in that case if 94%, and consistent with the notion that we have a more comparable group now, explains better the cross-section variation of immigrant proportion. Moreover, it is also useful to note that the under-representation of immigrants declines, again, consistent with the fact that we no longer pool immigrants that arrived as kids to the US. We find similar results when we focus on age groups 45-49, 50-54, and 55-59, when the R^2 ranges between 94%-97%.

The ACS shows approximately 30% more immigrants than our data, this is expected because our immigrant classification does not account for illegal immigrants. Indeed, the Department of Homeland Security estimates that 34% of immigrants were illegal in 2014. This matches very closely with the 30% under count of immigrants in Infutor, further validating our methods.

Additional Derivations

Derivation of equation (7):

$$\frac{Y_{jt}-\overline{Y}}{\overline{Y}} = \beta_{imm}\left(\frac{N_{jt}^{imm}-\overline{N}^{imm}}{1+\overline{N}^{imm}}\right) + \beta_{nat}\left(\frac{N_{jt}^{nat}-\overline{N}_{j}^{nat}}{1+\overline{N}_{j}^{nat}}\right)$$

Since we focus on teams that did not experience a team member death, it must be that $\mathbb{1}_{jt}^{death} = 0$ for all j and t. Next, we substitute for members' human capital in the team innovation production function, yielding:

$$
\begin{aligned}
Y_{jt} &= \left(\prod_{i \in J_j} \left(1 + N_{it}^{imm}\right)^{\beta^{imm}} \left(1 + N_{it}^{nat}\right)^{\beta^{nat}} \right)^{\frac{1}{|J_j|}} g\left(|J_j|\right) \varepsilon_{jt} \\
&= \Bigg\{ \underbrace{\Big(\prod_{i \in J_j} \left(1 + N_{it}^{imm}\right) \Big)^{\frac{1}{|J_j|}}}_{\equiv \left(1 + N_{jt}^{imm}\right)} \Bigg\}^{\beta_{imm}} \Bigg\{ \underbrace{\Big(\prod_{i \in J_j} \left(1 + N_{it}^{nat}\right) \Big)^{\frac{1}{|J_j|}}}_{\equiv \left(1 + N_{jt}^{nat}\right)} \Bigg\}^{\beta_{nat}} g\left(|J_j|\right) \varepsilon_{jt} \\
&= \left(1 + N_{jt}^{imm}\right)^{\beta_{imm}} \left(1 + N_{jt}^{nat}\right)^{\beta_{nat}} g\left(|J_j|\right) \varepsilon_{jt}
\end{aligned}
$$

We take a first-order approximation with respect to N_j^{imm} and N_j^{nat} around a base value $(\overline{N}^{imm}, \overline{N}^{nat}, \overline{Y})$:

$$
f(N_{jt}^{imm}, N_{jt}^{nat}) = f(\overline{N}^{imm}, \overline{N}^{nat}) + \left. \frac{\partial Y_{it}}{\partial N_{jt}^{imm}} \right|_{(\overline{N}, \overline{Y})} (N_{jt}^{imm} - \overline{N}^{imm}) + \left. \frac{\partial Y_{it}}{\partial N_{jt}^{nat}} \right|_{(\overline{N}, \overline{Y})} (N_{jt}^{nat} - \overline{N}^{nat})
$$

$$
Y_{jt} = \overline{Y} + \left(\frac{\beta_{imm}}{1 + \overline{N}^{imm}} \overline{Y} \right) (N_{jt}^{imm} - \overline{N}^{imm}) + \left(\frac{\beta_{nat}}{1 + \overline{N}^{nat}} \overline{Y} \right) (N_{jt}^{nat} - \overline{N}_j^{nat})
$$

Rearranging this yields equation (7) as desired. Note that in practice we calculate the average value across teams in the placebo-deceased group and across post-death years as our base value.

Derivation of equation (9):

$$
\frac{Y_{jt} - \overline{Y}}{\overline{Y}} = \frac{h_{jt} - \overline{h}}{\overline{h}} - \beta^{death}
$$

Since we focus on teams that experienced a team member death, it must be the case that (i) for the real-deceased teams we have $\mathbb{1}_{jt}^{death} = 1$ for all j and some $t > t_j^{death}$; and (ii) for the placebo-deceased teams we have $\mathbb{1}_{jt}^{death} = 0$ for

all j and t. Given known values of β^{imm} and β^{nat}, h_{jt} is identified. Similarly, we take a first-order approximation of the innovation production function with respect to h_{jt} and $\mathbb{1}_{jt}^{death}$ around a base value $(\overline{h}, \overline{\mathbb{1}}^{death}, \overline{Y})$, which is the average value across teams in the placebo-deceased group and across post-death years (suggesting $\overline{\mathbb{1}}^{death} = 0$).

$$f(h_{jt}, \mathbb{1}_{jt}^{death}) = f(\overline{h}, \overline{\mathbb{1}}^{death}) + \left.\frac{\partial Y_{it}}{\partial h_{jt}}\right|_{(\overline{h}, \overline{\mathbb{1}}^{death}, \overline{Y})} (h_{jt} - \overline{h}) + \left.\frac{\partial Y_{it}}{\partial \mathbb{1}_{jt}^{death}}\right|_{(\overline{h}, \overline{\mathbb{1}}^{death}, \overline{Y})} (\mathbb{1}_{jt}^{death} - \overline{\mathbb{1}}^{death})$$

$$Y_{jt} = \overline{Y} + \left(\frac{\overline{Y}}{\overline{h}}\right)(h_{jt} - \overline{h}) + \left(-\beta^{death}\overline{Y}\right)\mathbb{1}_{jt}^{death}$$

Note that in the post-death period the death effect estimate of $\mathbb{1}_{jt}^{death}$ must equal 1 since it takes a value 1 for the real-deceased teams and a value 0 for the placebo-deceased teams. Recognizing this fact and rearranging this last equation yield equation (9) as desired.

Figure A1. Validation with CENSUS 2000 - Population Sizes (millions)

Note: Scatterplot at the county level comparing Infutor county population to CENSUS 2000 adult population.

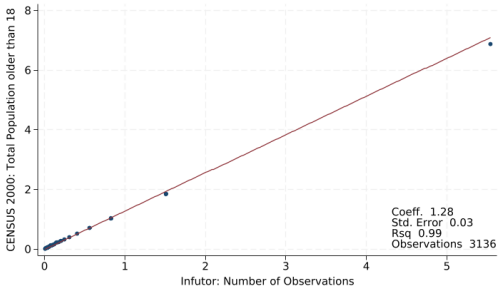


Figure A2. Validation of Pre-1950 Assignment Year Imputation

Note: Binscatter of encoded group numbers for each assignment year, controlling for area code fixed effects.

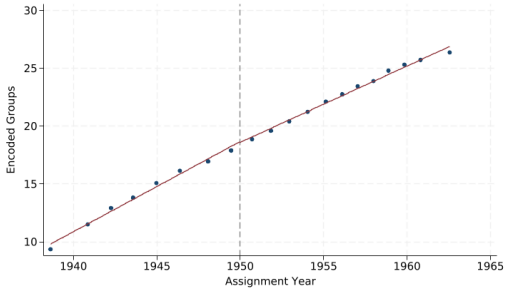


Figure A3. SSN Issuance Age Distribution

Note: Quantiles of the age of SSN issuance distribution by assignment year.

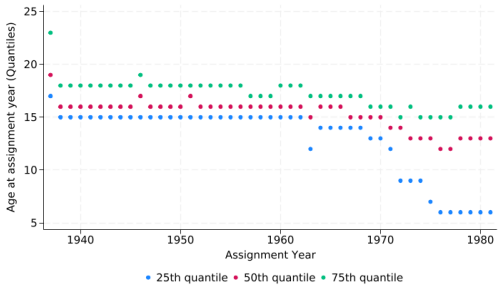


Figure A4. Validation with CENSUS 2000

Note: R^2 of regressing the proportion of foreign born in CENSUS 2000 against Infutor immigrant share by county. All regressions weighted by county population.

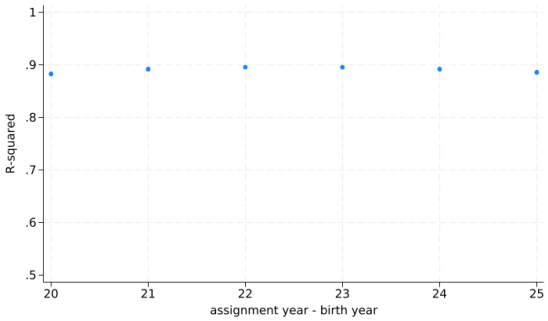
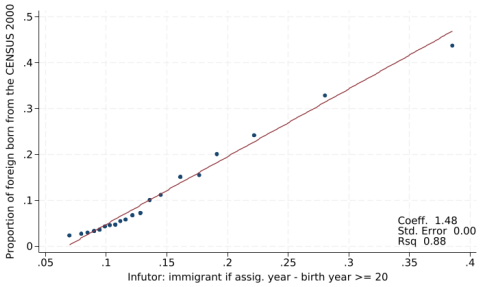


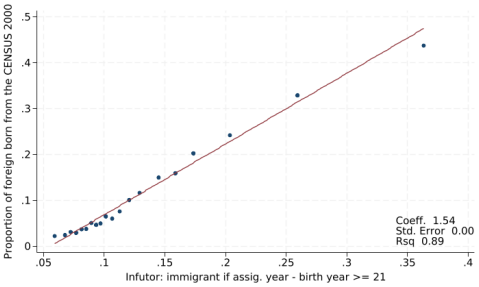
Figure A5. Validation with CENSUS 2000 – Binscatters

Note: Binscatters of the proportion of foreign born in CENSUS 2000 against Infutor immigrant share by county, for selected classification thresholds.

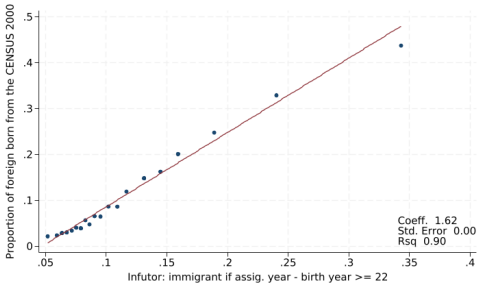
(A). Assig. year − birth year ≥ 20



(B). Assig. year − birth year ≥ 21



(C). Assig. year − birth year ≥ 22



(D). Assig. year − birth year ≥ 23

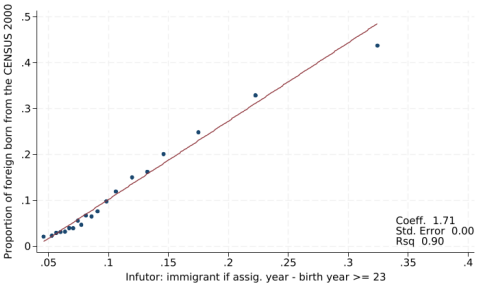


Figure A6. Validation with CENSUS 2000

Note: R^2 and slope coefficient of regressing foreign born share in CENSUS 2000 against Infutor immigrant share by county.

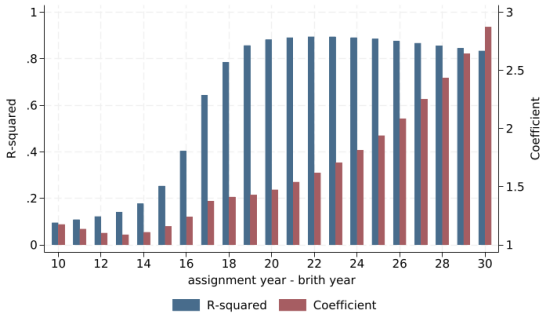
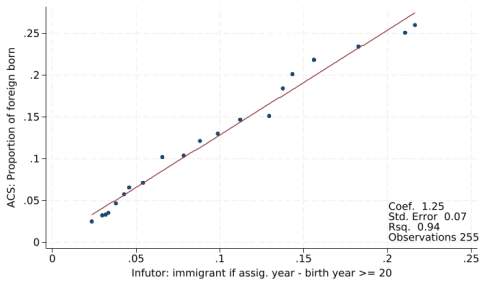


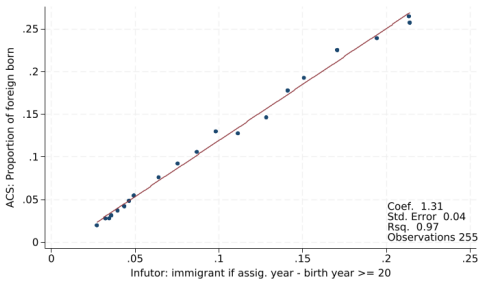
Figure A7. Validation against ACS by Selected Age Bins in 2005

Note: Binscatters of the proportion of immigrants in the state by age level in the ACS against the same proportion in Infutor. Each age bin had a separate regression, weighted by the number of individuals in each state and age level.

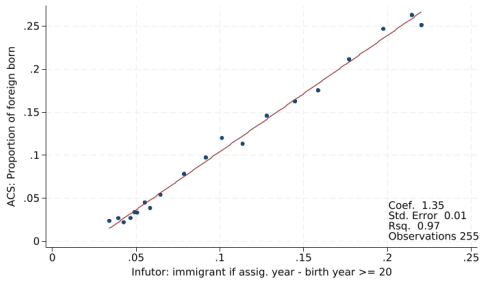
(A). 40-44 years old



(B). 45-49 years old



(C). 50-54 years old



(D). 55-59 years old

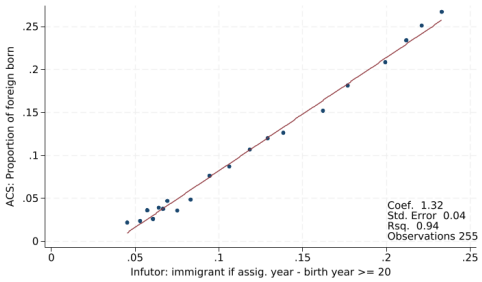


Figure A8. Share of Immigrant Contribution, Equal Credits

Note: Same categories as Figure 1, but apportioning full patent value to each co-inventor. Blue bars include all patents; red bars include solo-author patents only.

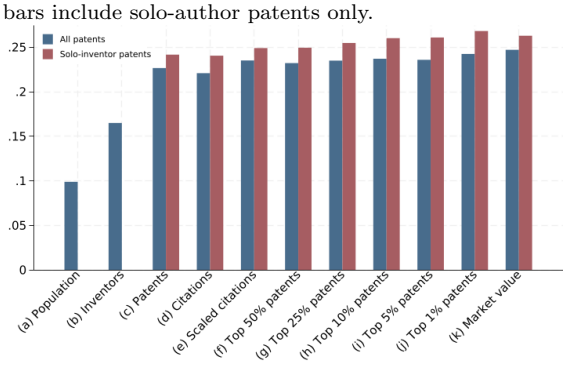


Figure A9. Team Size Distribution by Immigration Status

Note: Scaled citations by share of immigrant inventors within team, by team size.

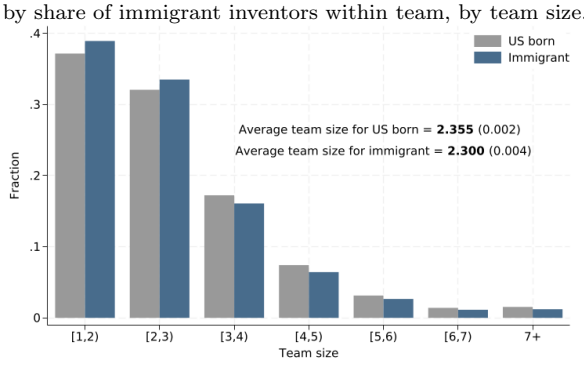


Figure A10. Team-level Share of Immigrant Contribution

Note: Scaled citations by share of immigrant inventors within team, by team size.

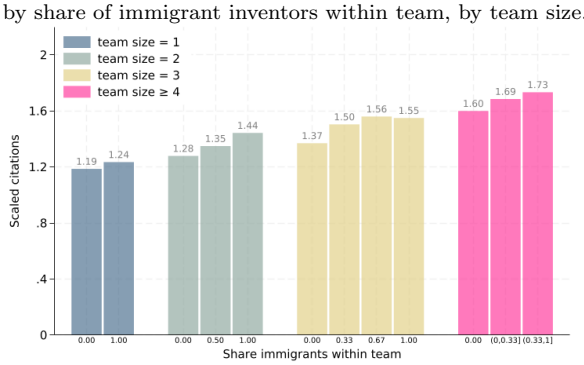
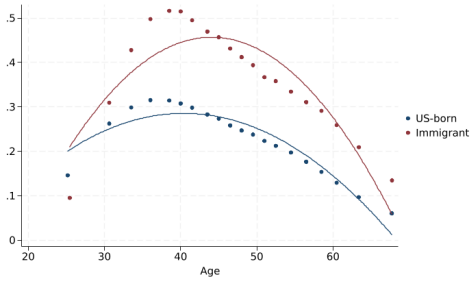


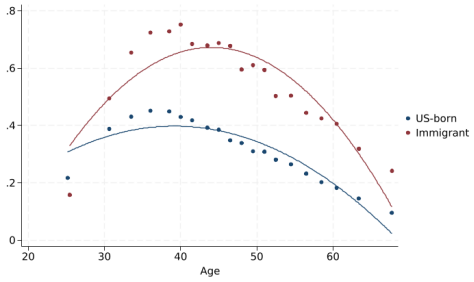
Figure A11. Productivity over the Life Cycle - First Patent in 1990s

Note: Same variables as Figure 3. Only inventors with first patent between 1990 and 1999.

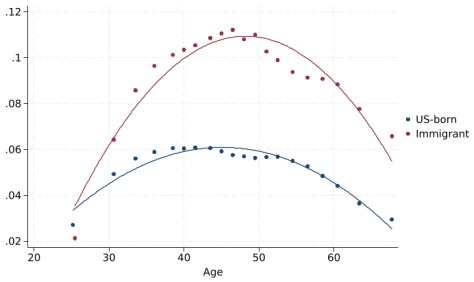
(A). Number of Patents



(B). Total Adjusted Citations



(C). Top Patents



(D). Market Value

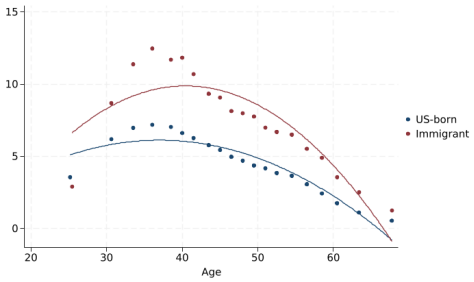
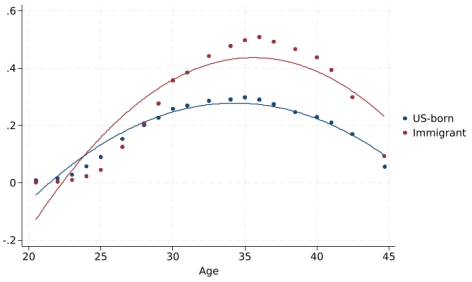


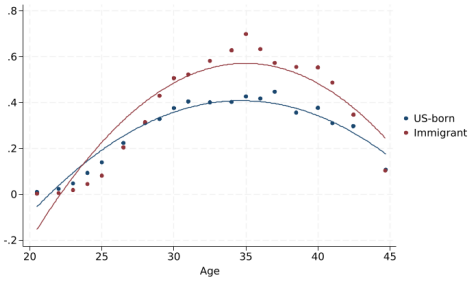
Figure A12. Productivity over the Life Cycle - 1970s Year of Birth

Note: Same variables as Figure 3. Only inventors born between 1970 and 1979.

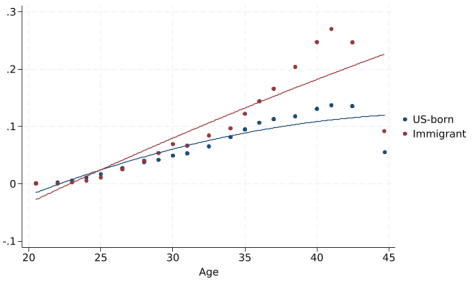
(A). Number of Patents



(B). Total Adjusted Citations



(C). Top Patents



(D). Market Value

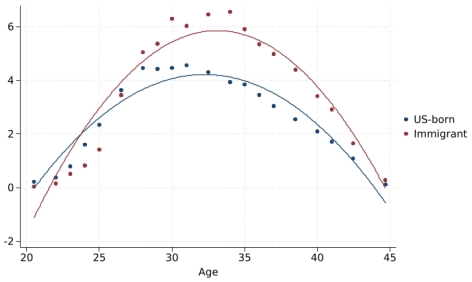


Figure A13. Productivity over the Life Cycle - Regressions

Note: Regression includes individual FE, Year FE, and age interacted with immigrant FE.

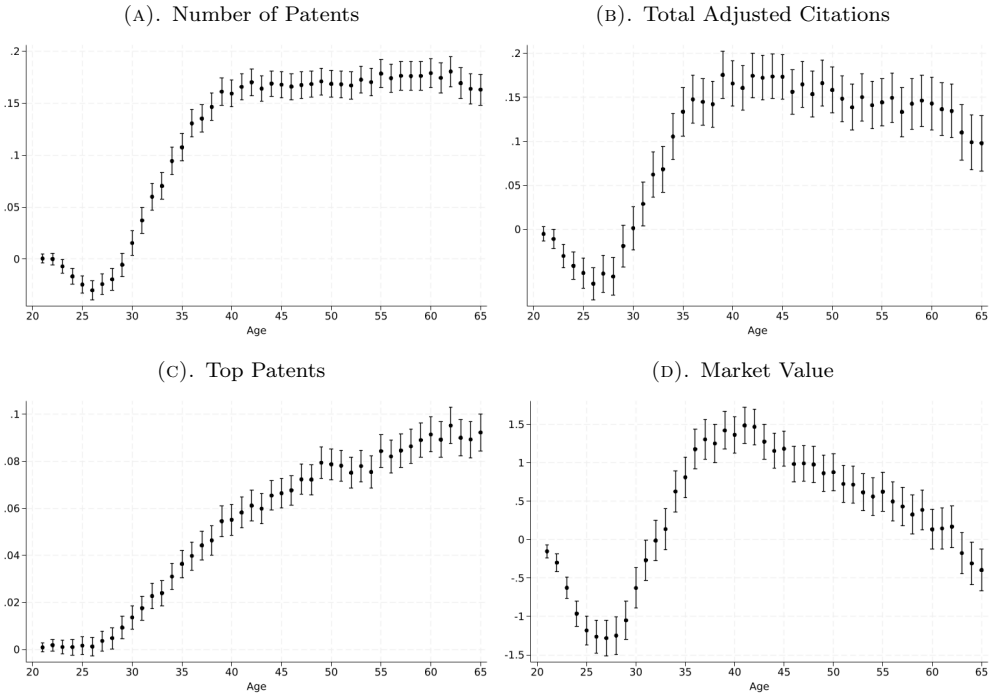


Figure A14. Collaboration with US-born and Foreign Inventors

Note: Number of unique US-born vs. foreign collaborators per immigrant inventor. Controls for year-of-birth FE. $\rho = 0.999$ with Balsmeier et al. (2015) identifiers.

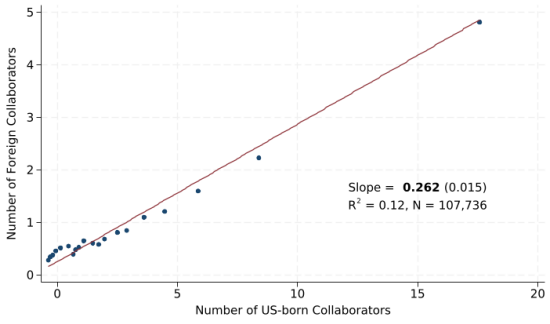


Table A1. Sample Construction

Note: Identifier (ID) is a combination of state, city, last name, and first name in the patent data. "Middle1" stands for the first letter of middle name. "Weak middle1 match" is when middle1 is missing in at least one dataset (USPTO or Infutor). "Strict middle1 match" is when middle1 is non-missing and the same across the two datasets. "Middle 1 consistent" is either weak or strict middle1 match. "First3" stands for the first three letters of first name. "Weak first name match" is when first name is either contained within or an alternate name for each other or contains misspelling. "Strict first name match" is when first name matches exactly across the two datasets.

Data Processing Steps	# Patents	# Unique IDs
Panel A. Cleaning Raw Patent Data		
– Start with raw USPTO data	14,991,282	
– Keep only U.S. patents	7,350,977	
– Keep if state and city are not missing	7,342,269	
– Keep if application year is between 1990 and 2016	6,229,618	1,351,024
	(100%)	(100%)
Panel B. Disambiguating Inventors using Infutor Data		
– State, city, last name, and first3 match *and*:		
⋆ Step 1: Middle1 consistent, exact first name match	3,287,046	680,252
⋆ Step 2: Middle1 consistent, contained first name match	673,939	144,434
⋆ Step 3: Middle1 consistent, alternate or misspelled first name	237,992	51,745
– At least one patent with consistent application-address year *and*:		
⋆ Step 4: Weak middle1 match, weak first name match	79,869	30,714
⋆ Step 5: Weak middle1 match, strict first name match	84,531	14,457
⋆ Step 6: Strict middle1 match, weak first name match	7,868	2,442
⋆ Step 7: Strict middle1 match, strict first name match	3,388	652
⋆ Step 8: Both middle name and first name match exactly	17,589	3,401
– State, city, last name match *and*:		
⋆ Step 9: Middle1 consistent, exact/contained/alternate/misspelled first name	14,993	3,532
– Number of IDs with unique Infutor matches	4,407,215	931,629
	(70.7%)	(69.0%)
– Number of unique inventors/individuals in Infutor	931,629	879,989

Table A2. Prediction of KPSS Economic Value

Note: This table reports the relationship between KPSS economic value and patent application and assignee-level characteristics, following a similar imputation in Kline et al. (2019). To reduce computational burden, we randomly select 10% of the data for model fitting. Coefficient estimates are based on a Poisson model with technology class random effects. The sample is the subsample of granted patents for which the Kogan et al. (2017) measure of economic value is available in our analysis sample. The dependent variable is the KPSS measure of economics value in millions of dollars. Standard errors are reported in parentheses. Number of claims measures the number of claims in the published U.S. patent application. $\log(\sigma_v)$ reports the log of the estimated standard deviation of the technology class random effects. χ^2 reports a likelihood ratio test statistic against a restricted Poisson model without random effects.

	KPSS Value
1(number of claims = 1)	0.2929
	(0.0080)
log(number of claims)	0.1933
	(0.0010)
Application year	0.0025
	(0.0005)
(Application year)^2	0.0041
	(0.0000)
Decision/grant year	0.0285
	(0.0007)
(Decision/grant year)^2	-0.0105
	(0.0000)
Constant	2.2617
	(0.0366)
$\log(\sigma_v)$	0.5689
	(0.0422)
Technology class	442
χ^2	280177.08
N	142,914

Table A3. Share of Immigrant Contribution – Different Team Sizes

Note: This table shows the share of immigrant contribution across different metrics listed in panels (c)-(k) of Figure 1. The statistics are displayed in row 1 for all patents granted between 1990-2016 and in rows 2-6 broken down by team size or the number of inventors on the patent.

Outcome:	number inventors (1)	number patents (2)	raw citations (3)	scaled citations (4)	top 50% patents (5)	top 25% patents (6)	top 10% patents (7)	top 5% patents (8)	top 1% patents (9)	economic value (10)
All patents (100%)	0.165	0.233	0.229	0.242	0.240	0.244	0.247	0.247	0.254	0.252
Team size 1 (54.6%)	0.181	0.242	0.241	0.249	0.250	0.255	0.260	0.261	0.268	0.263
Team size 2 (26.4%)	0.184	0.229	0.227	0.242	0.235	0.239	0.241	0.239	0.245	0.246
Team size 3 (11.3%)	0.185	0.220	0.216	0.234	0.225	0.228	0.230	0.228	0.235	0.245
Team size 4 (4.5%)	0.188	0.212	0.210	0.229	0.216	0.219	0.219	0.218	0.219	0.241
Team size $\geq$5 (3.1%)	0.179	0.202	0.192	0.201	0.206	0.204	0.203	0.200	0.198	0.233

Table A4. Robustness to Staggered Diff-in-Diff Methods

Note: We replicate our baseline diff-in-diff estimates for the effect on inventor death on surviving collaborators and then repeat the analysis using the Callaway and Sant'Anna (2021) method.

Specification:	Baseline ATE Results			Callaway & Sant'Anna		
Inventor death:	All	Native	Immigrant	All	Nat	Imm
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A. Number of Patents			
$\beta_{postdeath}$	-0.090	-0.077	-0.183	-0.137	-0.123	-0.294
	(0.011)	(0.011)	(0.046)	(0.017)	(0.019)	(0.043)
			Panel B. Scaled Citations			
$\beta_{postdeath}$	-0.142	-0.119	-0.320	-0.195	-0.171	-0.410
	(0.015)	(0.015)	(0.062)	(0.027)	(0.027)	(0.095)
			Panel C. Top Patents			
$\beta_{postdeath}$	-0.028	-0.021	-0.075	-0.031	-0.024	-0.099
	(0.003)	(0.003)	(0.013)	(0.005)	(0.005)	(0.015)
			Panel D. Economic Market			
$\beta_{postdeath}$	-1.678	-1.478	-3.195	-2.826	-2.588	-4.440
	(0.208)	(0.209)	(0.814)	(0.373)	(0.395)	(1.139)

Table A5. Inventor Death, Log Specification

Note: This table shows the difference-in-difference OLS estimates of the inventor death full sample. The sample is the same as defined in table 3 and all variables are as defined in table 1. Standard errors appear in parentheses and are clustered at the deceased inventor level.

Dependent Variable:	Log(1 + Number of Patents)			Log(1 + Adjusted Citations)		
	All	Immigrant	US-born	All	Immigrant	US-born
	(1)	(2)	(3)	(4)	(5)	(6)
Post Death	-0.035	-0.045	-0.034	-0.032	-0.049	-0.030
	(0.003)	(0.011)	(0.003)	(0.003)	(0.011)	(0.003)
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.541	0.548	0.539	0.475	0.480	0.473
Number of Deceased Inventors	159,662	8,015	151,647	159,662	8,015	151,647
Observations	6,811,858	503,706	6,308,152	6,811,858	503,706	6,308,152

Table A6. Inventor Death Treatment Effect Heterogeneity

Note: Each heterogeneity dimension has mean 0 and standard deviation 1 across the real-deceased inventors in our death analysis sample. Full details are given in the footnote of Table 5.

Dependent variables:	Number of Patents (1)	Scaled Citations (2)	Top Patents (3)	Economic Value (4)
Post-death × treat	–0.050	–0.098	–0.014	–1.143
	(0.011)	(0.014)	(0.003)	(0.177)
Post-death × treat × immigrant	–0.158	–0.203	–0.062	–1.728
	(0.049)	(0.063)	(0.013)	(0.829)
Post-death × treat × z-score:				
1. Deceased inventor's age	0.041	0.070	0.017	0.754
	(0.013)	(0.019)	(0.003)	(0.277)
2. Deceased inventor's year	–0.096	–0.044	–0.028	0.051
	(0.014)	(0.017)	(0.004)	(0.246)
3. Deceased inventor's cumulative patents and citations	–0.044	–0.021	–0.032	1.390
	(0.019)	(0.028)	(0.005)	(0.362)
4. Average surviving coauthors' age	–0.132	–0.121	–0.029	–1.907
	(0.014)	(0.020)	(0.004)	(0.240)
5. Average surviving coauthors' cumulative patents and citations	0.342	–0.061	0.097	–2.877
	(0.053)	(0.058)	(0.011)	(0.552)
6. Collaboration recency: time to most recent app pre-death	0.162	0.119	0.029	1.820
	(0.012)	(0.017)	(0.004)	(0.215)
7. Collaboration network: number of unique coauthors pre-death	–0.022	–0.002	0.010	–1.744
	(0.022)	(0.031)	(0.006)	(0.372)
8. Collaboration size: average team size on co-patents pre-death	–0.040	–0.023	–0.008	0.071
	(0.014)	(0.021)	(0.004)	(0.314)
9. Knowledge gap: backward citations and overlapping technology classes	–0.057	–0.060	–0.016	–0.423
	(0.012)	(0.017)	(0.004)	(0.257)
10. Commuting zone number of patents per capita	–0.004	0.007	–0.007	0.306
	(0.015)	(0.020)	(0.004)	(0.222)

Table A7. Team-level Productivity Residuals

Note: Production residuals, averaged across years for each team, are regressed on the share of immigrants on that team, controlling for team size. Standard errors appear in parentheses and are clustered at the deceased inventor level.

Dependent variable	Team-level Residuals		
Mean dep. var.	0.0734	0.0734	0.0734
	(1)	(2)	(3)
Share immigrant	0.0046		0.0015
	(0.0005)		(0.0005)
Team size (in logs)	−0.0796	−0.0795	
	(0.0002)	(0.0002)	
R^2	0.134	0.134	0.000
N	653,255	653,255	653,255

*

REFERENCES

Balsmeier, Benjamin, Alireza Chavosh, Guan-Cheng Li, Gabe Fierro, Kevin Johnson, Aditya Kaulagi, Doug O'Reagan, Bill Yeh, and Lee Fleming. 2015. "Automated Disambiguation of US Patent Grants and Applications." *Working Paper*.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of econometrics*, 225(2): 200–230.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2019. "Who Profits From Patents? Rent-Sharing at Innovative Firms." *Quarterly Journal of Economics*, 134(3): 1343–404.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological Innovation, Resource Allocation, and Growth." *Quarterly Journal of Economics*, 132(2): 665–712.