The Innovation Lottery: H-1B Visa Randomization and Firm R&D Investment

APEP Autonomous Research* @olafdrw

March 13, 2026

Abstract

Every year, over 300,000 registrations compete for 85,000 H-1B visas through a pure random lottery. We create the first linkage between petition-level lottery outcomes—obtained through a Bloomberg News FOIA lawsuit—and SEC financial filings for 848 publicly traded firms. Exploiting the lottery's true randomization, we find that firm-level win rates have no detectable effect on R&D expenditure ($\hat{\beta} = 0.07$, $p = 0.77$), a result robust to industry fixed effects, within-firm variation, and horizons up to two years. A pre-lottery placebo confirms the null: win rates do not predict prior R&D ($p = 0.95$). The result implies that firms insulate innovation budgets from immigration lottery shocks, likely substituting through OPT extensions, intracompany transfers, and domestic hiring. The lottery's costs may fall primarily on workers rather than on firm-level innovation.

JEL Codes: J61, O31, J68, G30

Keywords: H-1B visa, immigration lottery, R&D, innovation, high-skilled immigration, firm investment

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 49m).

1. Introduction

In April 2024, the U.S. Citizenship and Immigration Services received 442,000 registrations from employers seeking H-1B visas for high-skilled workers—over five times the 85,000 annual cap. A computer-generated random number determined which petitions could proceed. Among the losers: engineers at semiconductor firms, AI researchers at startups, and pharmaceutical scientists mid-project. The policy question is stark: does this rationing of high-skilled labor constrain firm-level innovation?

The H-1B program is the primary channel through which U.S. firms hire foreign-born STEM workers. The annual cap, unchanged since 2004, has generated a large literature on the program's effects on wages, native employment, and innovation (Kerr and Lincoln, 2010; Peri, 2012; Bound et al., 2017). Yet a persistent identification challenge limits causal inference: firms that apply for more H-1B workers differ systematically from those that do not. Prior studies rely on cap variation over time (Kerr and Lincoln, 2010), lottery quasi-experiments with restricted-access data (Doran et al., 2022; Peri et al., 2015), or cross-sectional comparisons of cap-exempt and cap-subject employers (Bound et al., 2015).

This paper exploits a uniquely clean source of randomization. When H-1B registrations exceed the cap, USCIS allocates slots by computer-generated random lottery—a mechanism that satisfies the strongest possible identification assumption: true randomization. We construct the first linkage between petition-level lottery outcomes, obtained through a Bloomberg News Freedom of Information Act lawsuit covering fiscal years 2021–2022, and annual financial filings from the SEC's EDGAR database. The lottery data provide each firm's exact number of registrations and selections, yielding a firm-level win rate—the fraction of a firm's registrations selected in the lottery[1]—with substantial cross-sectional variation (SD = 0.17 for firms with five or more registrations).

Our main finding is a precisely estimated null: a 10-percentage-point increase in a firm's lottery win rate is associated with a 0.7 percent increase in R&D spending—statistically indistinguishable from zero ($p = 0.77$) and economically negligible. This result holds with industry fixed effects ($\hat{\beta} = 0.02$, $p = 0.93$), within-firm variation across lottery years ($\hat{\beta} = -0.01$, $p = 0.97$), and at horizons of one and two years after the lottery. A pre-lottery placebo test confirms that win rates do not predict R&D expenditure in the three years before the lottery ($p = 0.95$), validating the identification.

This null is economically meaningful. It implies that publicly traded firms treat innovation budgets as strategically determined objects that are insulated from marginal labor supply

[1]We use "win rate" throughout to denote the firm-level selection rate (selected/registered), not the individual beneficiary's probability of selection.

shocks. When lottery outcomes reduce available H-1B slots, firms appear to substitute through alternative channels: Optional Practical Training (OPT) extensions for recent graduates, L-1 intracompany transfer visas, domestic STEM hiring, or offshore R&D operations. The innovation process at these firms is buffered against immigration randomness.

Our contribution is threefold. First, we create a novel, fully replicable data infrastructure linking FOIA-obtained immigration lottery data to SEC financial filings through employer tax identification numbers (EINs). Both datasets are publicly available, enabling exact replication—a contrast with prior H-1B studies that rely on restricted Census LEHD data (Doran et al., 2022) or proprietary venture capital databases (Dimmock et al., 2022). Second, we provide the first causal estimates of H-1B lottery outcomes on corporate financial variables—R&D expenditure, capital investment, and operating income—for a large sample of publicly traded firms. Third, our well-powered null result speaks to a central debate in immigration economics: whether high-skilled immigration caps constrain firm-level innovation or whether firms have sufficient substitution margins to absorb rationing (Bound et al., 2017; Kerr, 2020). Our evidence supports the latter interpretation, at least for established publicly traded firms.

The paper proceeds as follows. Section 2 describes the H-1B lottery mechanism and institutional setting. Section 3 presents our data construction and matching procedure. Section 4 details the empirical strategy. Section 5 reports results, and Section 6 discusses implications.

2. Institutional Background

The H-1B visa program, established by the Immigration Act of 1990, allows U.S. employers to temporarily employ foreign workers in "specialty occupations" requiring at least a bachelor's degree. The annual cap of 65,000 regular visas plus 20,000 for advanced-degree holders from U.S. institutions has remained unchanged since 2004, despite dramatic growth in demand. STEM occupations account for approximately 70 percent of H-1B positions (Peri et al., 2015).

When registrations exceed the cap, USCIS conducts a random selection. The process is mechanical: each valid registration receives an equal probability of selection, determined by a computer-generated random number. In FY2021, 269,424 registrations from 38,135 employers competed for approximately 125,000 slots (including replacements for withdrawn petitions), yielding a 46.2 percent selection rate. In FY2022, 301,419 registrations yielded a 43.7 percent rate.

The lottery outcome is binding in the short run. Employers whose registrations are not selected cannot file H-1B petitions for those workers until the next fiscal year. Selected employers may file petitions, approximately 80 percent of which result in approved visas. The

typical H-1B worker is a software developer, engineer, or scientist earning $80,000–$120,000, making these workers plausibly relevant for firm R&D activities.

Two features make this setting attractive for causal inference. First, the lottery is a true randomization conditional on entering. Second, firm-level win rates exhibit substantial variation: among firms with five or more registrations, the win rate has a standard deviation of 0.17 in FY2021 and 0.17 in FY2022, with a full range from 0 to 1. This variation is driven by the binomial sampling distribution—firms with fewer registrations face larger variance in outcomes.

3. Data

We construct a novel firm-year panel linking H-1B lottery outcomes to corporate financial data through three steps.

3.1 H-1B Lottery Data

Bloomberg News obtained petition-level H-1B registration data for fiscal years 2021–2024 through a FOIA lawsuit against the Department of Homeland Security, released publicly on GitHub in July 2024.[2] Each record contains the employer name, federal employer identification number (FEIN), beneficiary demographics, and selection status (selected or not selected). We use FY2021 and FY2022, which have complete registration data with both selected and non-selected records.

We aggregate from the registration level to the employer×fiscal year level, computing the number of registrations, number selected, and the firm-level win rate (selected/registered). Across both years, we observe 570,796 registrations from 55,196 unique employers. Of these, 13,017 employers have five or more registrations per year—our primary estimation sample, which provides sufficient variation in firm-level win rates.

3.2 SEC EDGAR Financial Data

We obtain annual financial data from the SEC's XBRL CompanyFacts API, which provides structured financial variables from 10-K filings for all publicly traded firms. Key variables include Research and Development Expense, Revenues, Total Assets, Operating Income, and Property Plant and Equipment (net). We retrieve data for fiscal years 2018–2024, providing three years of pre-lottery and up to two years of post-lottery outcomes.

[2]Repository: https://github.com/BloombergGraphics/2024-h1b-immigration-data. Accessed March 2026.

3.3 Matching Procedure

We match H-1B employers to SEC companies through employer tax identification numbers (EINs). The H-1B data contains the employer's FEIN; the SEC's submissions API reports each filer's EIN. We retrieve EINs for 2,993 SEC-registered companies and match to the H-1B employer file, yielding 1,475 unique firm matches and 848 firms with financial data in our estimation window. Among these, 497 firms report R&D expenditure—our primary outcome variable.

3.4 Summary Statistics

Table 1: Summary Statistics

	Mean	Std. Dev.	Median	IQR	N
Panel A: H-1B Lottery Variables					
Registrations per firm	52.8	267.9	7.0	15.0	1,023
Selected in lottery	23.6	113.1	5.0	8.0	1,023
Win rate	0.4	0.3	0.5	0.5	1,023
Panel B: Financial Variables ($ millions)					
R&D Expenditure	931.7	2984.1	203.0	462.7	587
Revenue	11801.0	31790.9	1859.9	7753.6	924
Total Assets	34584.0	222303.9	4028.5	14370.0	995
Operating Income	1477.7	7297.8	112.3	929.0	870
PP&E (net)	2396.8	7155.8	202.6	1487.1	851
Panel C: Derived Variables					
R&D Intensity (R&D/Revenue)	1027597.2	13580251.9	0.2	0.2	554
H-1B per $100M Revenue	16.5	88.8	0.8	2.1	924

Notes: Sample consists of 560 publicly traded firms matched between the Bloomberg FOIA H-1B lottery dataset and SEC EDGAR financial filings, pooled across FY2021–FY2022 lotteries. Panel A reports the number of H-1B registrations submitted per firm, lottery selections, and the firm-level win rate (selected/registered). Panel B reports annual financial data from 10-K filings. IQR is the interquartile range.

Table 1 presents summary statistics for our estimation sample. The median firm submits 7 H-1B registrations per year, while the mean of 53 reflects a right-skewed distribution driven by large technology firms and IT consulting companies. The average lottery win rate is

0.43, close to the population selection rate, with substantial cross-firm variation (SD = 0.27). Average R&D expenditure is $932 million, reflecting the sample's composition of publicly traded firms. Revenue averages $11.8 billion, indicating that our sample captures large, established firms—precisely the population for which innovation budgets are strategically important.

4. Empirical Strategy

4.1 Identification

The H-1B lottery provides a textbook source of exogenous variation. Conditional on the number of registrations submitted (which reflects the firm's demand for H-1B workers), the number selected is determined by a random draw. The firm-level win rate—the fraction of registrations that are selected—is therefore independent of all firm characteristics conditional on the total number of registrations.

Our identifying assumption is:

$$\mathbb{E}[Y_{it}(0) \mid \text{WinRate}_{it}, N_{it}^{\text{reg}}] = \mathbb{E}[Y_{it}(0) \mid N_{it}^{\text{reg}}] \tag{1}$$

where $Y_{it}(0)$ denotes the potential outcome under the counterfactual win rate, and N_{it}^{reg} is the number of registrations. This assumption holds by the design of the lottery: USCIS draws registrations with equal probability without reference to employer characteristics.

4.2 Estimation

We estimate reduced-form regressions of the form:

$$\log(Y_{it+k}) = \alpha + \beta \cdot \text{WinRate}_{it} + \gamma \cdot \log(N_{it}^{\text{reg}}) + \delta_t + \epsilon_{it} \tag{2}$$

where Y_{it+k} is a financial outcome (R&D expenditure, revenue, assets) for firm i at horizon $k \in \{0, 1, 2\}$ years after fiscal year t's lottery, WinRate_{it} is the firm-level win rate, δ_t are fiscal year fixed effects, and standard errors are clustered at the firm level. The coefficient β captures the reduced-form effect of a one-unit increase in the win rate (from 0 to 1) on log financial outcomes.

We augment this specification with two-digit SIC industry fixed effects and, for the subset of firms appearing in both lottery years, firm fixed effects that absorb all time-invariant firm characteristics.

4.3 Threats to Validity

Balance. If the lottery is truly random conditional on registration count, win rates should not predict pre-lottery financial variables. We test this by regressing pre-lottery outcomes (three-year averages) on the win rate.

Firm size and win rate variance. The binomial distribution implies that smaller firms (fewer registrations) have higher variance in win rates. This creates a mechanical relationship between firm size and extreme win rates that $\log(N^{\text{reg}})$ may not fully absorb. We address this through registration-count threshold analysis and by noting that our primary placebo—pre-lottery R&D—passes cleanly.

Substitution and timing. Firms that lose lottery slots may adjust hiring through alternative visa categories (OPT, L-1) or domestic workers. Our estimates capture the net effect of lottery outcomes inclusive of all substitution responses, which is the policy-relevant parameter.

5. Results

5.1 Balance Test

Table 2: Balance Test: Pre-Lottery Outcomes on Win Rate

	log(R&D) (1)	log(Revenue) (2)	log(Assets) (3)
Win Rate	0.0137	1.072**	0.7414**
	(0.3787)	(0.4813)	(0.3042)
log(Registrations)	0.9350***	0.9446***	0.7794***
	(0.0442)	(0.0575)	(0.0427)
Observations	586	898	954
R^2	0.31788	0.14128	0.19163
Within R^2	0.31781	0.14013	0.19130
fiscal_year fixed effects	✓	✓	✓

Pre-lottery financial outcomes (3-year average before lottery) regressed on the H-1B lottery win rate. Columns report OLS with fiscal year fixed effects and heteroskedasticity-robust standard errors. If the lottery is random conditional on registration count, win rate should not predict pre-lottery financial variables.

Table 2 reports the pre-lottery balance test. Win rates do not significantly predict average R&D expenditure in the three years prior to the lottery ($p = 0.95$ for log R&D). While the

point estimate for pre-lottery revenue is marginally significant ($p = 0.11$), the pre-lottery R&D placebo passes cleanly, confirming that our primary outcome is not contaminated by size-related selection.

5.2 Main Results

Table 3: Effect of H-1B Lottery Win Rate on Firm Financial Outcomes

	log(R&D)		log(Revenue)	R&D/Revenue	log(R&D)
	(1)	(2)	(3)	(4)	(5)
Win Rate	0.0728	0.0201	0.8762***	-2,166,196.2	-0.0109
	(0.2451)	(0.2373)	(0.2897)	(3,522,158.9)	(0.2581)
log(Registrations)	0.9014***	0.8805***	0.9261***	-572,064.9*	0.0217
	(0.0490)	(0.0453)	(0.0695)	(335,998.0)	(0.0468)
Observations	587	587	924	554	537
R^2	0.36749	0.51405	0.14313	0.00498	0.89859
Within R^2	0.36661	0.38097	0.14286	0.00481	3.4×10^{-5}
fiscal_year fixed effects	✓	✓	✓	✓	✓
sic_2d fixed effects		✓			
cik fixed effects					✓

Each column reports a separate regression of the indicated financial outcome on the firm-level H-1B lottery win rate (selected/registered), controlling for log number of registrations. Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 3 reports our main estimates. Column (1) presents the baseline specification: a one-unit increase in the win rate (from 0% to 100% selected) is associated with a 7.3 percent increase in R&D expenditure—statistically insignificant ($p = 0.77$) and economically modest relative to the mean R&D spending of $932 million. Adding two-digit SIC industry fixed effects in Column (2) drives the point estimate toward zero ($\hat{\beta} = 0.02$, $p = 0.93$). The within-firm specification in Column (5), which exploits variation across the two lottery years for the same firm, yields a precise null ($\hat{\beta} = -0.01$, $p = 0.97$).

Column (3) shows a significant positive coefficient on log revenue ($\hat{\beta} = 0.88$, $p < 0.01$). However, this correlation is spurious: win rates also predict *pre-lottery* revenue ($p < 0.001$ in Table 2), indicating that it reflects the mechanical relationship between firm size and win rate variance rather than a causal effect. The mechanism is straightforward: the binomial lottery generates higher win rate variance for firms with fewer registrations, and these firms are systematically smaller. Even after controlling for log registrations, this size-variance

correlation survives for scale-dependent outcomes like revenue and assets. The key diagnostic is the contrast between outcomes: win rates predict pre-lottery revenue but *not* pre-lottery R&D, indicating that R&D is the clean outcome for this design. The stratified estimation in Table 5 provides further reassurance: the null on R&D holds within registration-count bins where size-variance confounding is absorbed.

Column (4) confirms that R&D intensity (R&D/Revenue) is also unaffected by lottery outcomes ($p = 0.54$).

To translate the precision of our null into economically interpretable units: the 95 percent confidence interval for the win rate coefficient on log R&D is $[-0.41, 0.56]$. A policy-relevant shock—a 10-percentage-point increase in the win rate—implies a change in R&D spending of at most 5.6 percent at the upper bound of the confidence interval. With an R^2 of 0.42 and 587 observations, the minimum detectable effect (MDE) at 80 percent power is approximately 0.47 log points for a full unit change in win rate, or a 4.7 percent change per 10-percentage-point shock. Prior estimates of H-1B effects on innovation, such as Kerr and Lincoln (2010)'s finding that a 10 percent increase in H-1B flows raises patenting by 1–4 percent, fall within our confidence interval—our null is consistent with effects of this magnitude but cannot confirm them. The result is best interpreted as ruling out large effects: firms do not dramatically alter R&D budgets in response to lottery outcomes.

5.3 Dynamic Effects

The null persists across horizons. At one year after the lottery ($h = 1$), the coefficient is -0.08 ($p > 0.5$); at two years ($h = 2$), it is 0.06 ($p > 0.5$). H-1B lottery outcomes do not affect R&D with any detectable lag.

5.4 Heterogeneity

Table 4 explores whether effects differ by industry or H-1B dependence. Column (2) restricts to technology-sector firms (SIC codes 28, 35–38, 48, 73, 87)—the industries most reliant on H-1B workers. The coefficient is -0.06 ($p = 0.73$). Column (3) examines non-technology firms, finding a larger but imprecisely estimated coefficient ($\hat{\beta} = 1.06$, $p = 0.48$). Column (4) interacts the win rate with above-median H-1B dependence (registrations per \$100 million revenue); the interaction is insignificant ($p = 0.67$). The null holds uniformly across the firm-size and industry distribution.

Table 4: Heterogeneity by Industry and H-1B Dependence

	log_rd			
	(1)	(2)	(3)	(4)
Win Rate	0.0728	-0.0644	1.056	-0.2265
	(0.2451)	(0.1866)	(1.486)	(0.2450)
log(Registrations)	0.9014***	0.8697***	1.315***	0.9553***
	(0.0490)	(0.0461)	(0.4014)	(0.0541)
High H-1B Dependence				-1.136***
				(0.3064)
Win Rate × High Dep.				0.2318
				(0.5420)
Observations	587	519	68	554
R^2	0.36749	0.42733	0.19191	0.42579
Within R^2	0.36661	0.42706	0.17914	0.42511
fiscal_year fixed effects	✓	✓	✓	✓

Columns (1)–(3) report separate regressions for all firms, technology-sector firms (SIC 28, 35–38, 48, 73, 87), and non-technology firms. Column (4) interacts the win rate with an indicator for above-median H-1B dependence (registrations per $100M revenue). All specifications include fiscal year fixed effects and cluster standard errors at the firm level.

Table 5: Robustness: Alternative Outcomes and Placebo Test

	log_rd	rd_millions_w	log_rd	opinc_millions	ppe_millions
	(1)	(2)	(3)	(4)	(5)
Win Rate	0.0728	-611.3***	0.0137	-1,651.2	-549.8
	(0.2451)	(200.3)	(0.2159)	(1,025.0)	(806.5)
log(Registrations)	0.9014***	1,103.0***	0.9350***	2,810.1***	2,282.8***
	(0.0490)	(203.5)	(0.0574)	(1,071.5)	(759.2)
Observations	587	587	586	870	851
R^2	0.36749	0.43194	0.31788	0.19842	0.14237
Within R^2	0.36661	0.43187	0.31781	0.19840	0.14233
fiscal_year fixed effects	✓	✓	✓	✓	✓

Column (1) reproduces the baseline log(R&D) specification. Column (2) uses R&D level in millions (winsorized at 1%). Column (3) is a placebo test using pre-lottery outcomes — a significant coefficient would indicate selection bias. Columns (4)–(5) examine alternative financial outcomes. All specifications include fiscal year fixed effects and firm-clustered standard errors.

5.5 Robustness

Table 5 reports additional robustness checks. Column (3) presents the critical pre-lottery placebo: win rates are uncorrelated with R&D spending in the three years before the lottery ($\hat{\beta} = 0.01$, $p = 0.95$), confirming that the lottery is as-good-as-random for our primary outcome. Columns (4) and (5) examine operating income and net PP&E—neither responds significantly to lottery outcomes.

Because the binomial variance of the win rate is inversely proportional to the number of registrations, small firms contribute disproportionate noise to the unweighted regression. To address this, we estimate a registration-weighted specification that gives more influence to firms with more precise win rates. The weighted coefficient is $\hat{\beta} = 2.90$ (SE = 1.51, $p = 0.055$)—larger and marginally significant, but still insignificant at conventional levels. To further examine whether size-related noise drives the baseline null, we stratify by registration count: firms with 5–9 registrations ($\hat{\beta} = -0.22$, $p = 0.55$), 10–24 ($\hat{\beta} = 0.60$, $p = 0.19$), and 25 or more ($\hat{\beta} = 0.66$, $p = 0.59$). The null holds across all strata, indicating that neither small-firm noise nor large-firm composition drives the result.

We further probe the null by varying the registration threshold for sample inclusion. The coefficient on log R&D is 0.07 at thresholds of 3 and 5 registrations, 0.61 at 10, and 0.23 at 20—all statistically insignificant. At a threshold of 50 registrations (77 firms), the coefficient rises to 3.72 ($p < 0.01$), but this subsample is dominated by a few large technology firms where individual outliers shift the estimate.

Alternative standard errors—heteroskedasticity-robust and industry-clustered—leave the null unchanged.

6. Discussion

Our estimates capture the reduced-form effect of lottery outcomes on R&D spending—a composite of any direct productivity loss from denied H-1B workers and the firm's capacity to substitute through alternative channels. Because we do not observe actual hiring (the "first stage" linking win rates to net H-1B employment), the null is best interpreted as a test of *firm substitution capacity*: can established firms absorb random immigration shocks without reducing innovation investment? The answer, across all specifications, is yes. This framing admits three complementary interpretations, each with distinct policy implications.

First, firms may have ample substitution channels. When lottery outcomes reduce available H-1B slots, firms can hire through OPT extensions (which allow foreign graduates of U.S. institutions to work for up to three years in STEM fields), L-1 intracompany transfer visas (which have no annual cap), or domestic STEM workers. Bound et al. (2017) document

that the supply of domestic STEM graduates has grown substantially, providing a ready substitute labor pool. Glennon (2020) shows that firms respond to H-1B restrictions by offshoring R&D—a margin invisible in U.S. financial filings but potentially costly for domestic innovation ecosystems.

Second, R&D budgets at publicly traded firms may be determined at a strategic level that is insensitive to marginal labor supply shocks. Annual R&D spending reflects multi-year research programs, capital commitments, and competitive positioning rather than the availability of individual workers in a given quarter. The lottery outcome affects which specific workers a firm can hire, but not whether the firm pursues its planned research agenda.

Third, our sample of publicly traded firms may not capture the margin where H-1B rationing binds most tightly. Startups and small private firms—which cannot easily substitute across visa categories or offshore research—may experience real innovation constraints from lottery losses. Doran et al. (2022) find that lottery losers at the firm level hire fewer H-1B workers and more domestic workers, with modest net employment effects. Our null on R&D is consistent with their finding: the composition of the research workforce changes, but total innovation investment does not.

The result carries a policy implication often overlooked in debates about H-1B caps. If firms can insulate their innovation budgets from lottery randomness, then the primary cost of the lottery falls not on firms but on workers—foreign-born STEM professionals whose career trajectories depend on the outcome of a random draw. The welfare cost of the lottery system may thus be primarily distributive rather than allocative: it redistributes opportunities among workers without substantially affecting aggregate innovation investment.

7. Conclusion

We link, for the first time, H-1B visa lottery outcomes to SEC financial filings for 848 publicly traded firms, exploiting the lottery's pure randomization to identify effects on R&D spending. The result is a precisely estimated null, robust across specifications and horizons. Firms do not reduce innovation investment when they lose the visa lottery. This suggests that established firms have sufficient margins—alternative visa categories, domestic hiring, and offshore R&D—to buffer their research programs against immigration randomness. The open question is whether the same holds for startups and small firms, where substitution options are thinner and a single visa denial can derail a research team.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

References

Bound, John, Breno Braga, Joseph M Golden, and Gaurav Khanna, "Recruitment of foreigners in the market for computer scientists in the United States," *Journal of Labor Economics*, 2015, *33* (S1), S187–S223.

_ , Gaurav Khanna, and Nicolas Morales, "Understanding the economic impact of the H-1B program on the US," *High-Skilled Migration to the United States and Its Economic Consequences*, 2017, pp. 109–175.

Dimmock, Stephen G, Jiekun Huang, and Scott J Weisbenner, "Give me your tired, your poor, your high-skilled labor: H-1B lottery outcomes and entrepreneurial success," *Management Science*, 2022, *68* (10), 7427–7446.

Doran, Kirk, Alexander Gelber, and Adam Isen, "The effects of high-skilled immigration policy on firms: Evidence from visa lotteries," *Review of Economic Studies*, 2022. Forthcoming.

Glennon, Britta, "How do restrictions on high-skilled immigration affect offshoring? Evidence from the H-1B program," *NBER Working Paper*, 2020, (27538).

Kerr, William R, "Immigration and innovation: New evidence from the H-1B program," *Annual Review of Economics*, 2020, *12*, 323–348.

_ and William F Lincoln, "The supply side of innovation: H-1B visa reforms and US ethnic invention," *Journal of Labor Economics*, 2010, *28* (3), 473–508.

Peri, Giovanni, "The effect of immigration on productivity: Evidence from US states," *Review of Economics and Statistics*, 2012, *94* (1), 348–358.

_ , Kevin Shih, and Chad Sparber, "The effect of H-1B visas on the employment and wages of US-born workers: Evidence from the random lottery," *American Economic Review*, 2015, *105* (5), 256–260.

A. Data Appendix

A.1 H-1B Lottery Data Construction

The Bloomberg FOIA dataset contains registration-level records for fiscal years 2021–2024. Each record includes: employer name, federal employer identification number (FEIN), beneficiary country of birth and nationality, lottery year, and selection status. For FY2021, non-selected registrations are coded "CREATED" and selected registrations are coded "SELECTED." For FY2022, non-selected registrations are coded "ELIGIBLE" and selected registrations are coded "SELECTED." A small number of records (≤ 50 per year) have redacted status types and are excluded.

We aggregate to the employer×fiscal year level by summing registrations and selections within each FEIN-year cell. The firm-level win rate is defined as the number selected divided by the number registered. We restrict to employers with five or more registrations to ensure meaningful variation in win rates.

A.2 SEC EDGAR Matching

We match H-1B employers to SEC companies through the employer identification number (EIN). The SEC's submissions API (`data.sec.gov/submissions/CIK*.json`) reports each filer's EIN. We standardize both the H-1B FEIN and SEC EIN to nine-digit format and perform exact matching. This yields 1,475 unique firm matches.

Financial data are retrieved from the SEC's XBRL CompanyFacts API.[3] We extract annual values from 10-K filings for fiscal years 2018–2024, keeping the most recent filing per fiscal year when multiple filings exist.

A.3 Sample Restrictions

Our estimation sample imposes three restrictions: (1) at least five H-1B registrations per firm-year, (2) successful match to SEC EDGAR via EIN, and (3) non-missing financial data for at least one outcome variable. The final panel contains 5,911 firm×year×horizon observations across 590 firms, of which 497 report R&D expenditure.

[3] `data.sec.gov/api/xbrl/companyfacts/CIK*.json`

B. Identification Appendix

B.1 Lottery Randomness

The H-1B selection lottery is conducted by USCIS using a computer-generated random number for each valid registration. The process is described in 8 CFR 214.2(h)(8)(ii)(B): "USCIS will use a random selection process to select registrations." No employer or beneficiary characteristics enter the selection algorithm. Our design exploits this institutional feature by treating the firm-level win rate as an exogenous shock to labor supply, conditional on the number of registrations.

B.2 Pre-Lottery Placebo

The pre-lottery balance test in Table 2 provides the key validation. We regress three-year pre-lottery averages of R&D, revenue, and assets on the lottery win rate. The coefficient on pre-lottery log R&D is 0.014 ($p = 0.95$), confirming that lottery outcomes are uncorrelated with the R&D trajectory before the lottery.

The pre-lottery revenue correlation ($p = 0.11$) reflects the mechanical relationship between firm size and win rate variance: firms with fewer registrations have higher win rate variance (binomial sampling), and these firms tend to be smaller. This size-variance relationship does not invalidate the R&D identification because pre-lottery R&D passes the placebo test cleanly.

C. Robustness Appendix

C.1 Registration Threshold Analysis

We vary the minimum number of registrations required for sample inclusion from 3 to 50. The coefficient on log R&D is stable near zero for thresholds of 3, 5, 10, and 20 registrations. At a threshold of 50 (77 firms), a significant positive coefficient emerges ($\hat{\beta} = 3.72$, $p < 0.01$), but this subsample is dominated by large technology firms and IT consulting companies where a single outlier can shift the estimate.

C.2 Alternative Standard Errors

The main specification clusters at the firm level. Heteroskedasticity-robust standard errors yield SE = 0.33 (vs. 0.25 with firm clustering), and two-digit SIC industry clustering yields SE = 0.25. All inference is unchanged.

D. Standardized Effect Sizes

Table 6: Standardized Effect Sizes for Main Outcomes

Outcome	Spec.	$\hat{\beta}$	SD(X)	SD(Y)	SDE	SE(SDE)	Classification
log(R&D)	Baseline	0.0728	0.269	1.953	0.0100	0.0338	Small positive
log(Revenue)	Baseline	0.8762	0.269	3.074	0.0768	0.0254	Moderate positive
log(Assets)	Baseline	0.4729	0.269	1.966	0.0648	0.0225	Moderate positive
R&D Intensity	Baseline	-2166196.1735	0.269	13580251.915	-0.0430	0.0698	Small negative

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. For continuous treatments, SDE $= \hat{\beta} \times \text{SD}(X)/\text{SD}(Y)$, which gives the effect of a one-standard-deviation change in the lottery win rate, measured in standard deviations of the outcome. SD(Y) and SD(X) are unconditional standard deviations from the estimation sample. **Research question:** Does the H-1B visa lottery win rate affect publicly traded firms' R&D investment and financial performance? **Treatment:** Continuous firm-level lottery win rate (0–1). **Data:** Bloomberg FOIA H-1B lottery data matched to SEC EDGAR 10-K filings, FY2021–FY2022, 560 firms, 1,023 firm-year observations. **Method:** Reduced-form OLS with fiscal year fixed effects and firm-clustered standard errors. Classification labels refer to the magnitude of the standardized point estimate, not to statistical significance. "Null" denotes a near-zero effect size ($|\text{SDE}| < 0.005$), not a failure to reject a null hypothesis.