The Impact of Immigration on Firms and Workers: Insights from the H-1B Lottery*

Agostina Brinatti† Mingyu Chen‡
Parag Mahajan§ Nicolas Morales¶ Kevin Shih‖

March 2024

Abstract

We study how random variation in the availability of highly educated, foreign-born workers impacts firm performance and recruitment behavior. We combine two rich data sources: 1) administrative employer-employee matched data from the US Census Bureau; and 2) firm-level information on the first large-scale H-1B visa lottery in 2007. Using an event-study approach, we find that lottery wins lead to increases in firm hiring of college-educated, immigrant labor along with increases in scale and survival. These effects are stronger for small, skill-intensive, and high-productivity firms that participate in the lottery. We do not find evidence for displacement of native-born, college-educated workers at the firm level, on net. However, this result masks dynamics among more specific subgroups of incumbents that we further elucidate.

JEL Codes: F22, J61
Keywords: Immigration, firm dynamics, productivity, H-1B visa, high-skilled migration

*We thank John Bound, Ethan Lewis, Chad Sparber, and seminar participants at UVA, FRB Dallas, Peking University, Penn State, Bank of Canada, Wharton Migration and Organizations Conference, SEA Annual Meetings (Ft. Lauderdale), NBER Immigrants and the US Economy Conference, Barcelona Summer Forum, SOLE (Philadelphia), Richmond Fed, Princeton, University of Delaware and George Washington University. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1582. (CBDRB-FY24-P1582-R11150). This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. This research benefited financial support from the Upjohn Institute. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Board of Governors. This work does not relate to Mingyu Chen's position at Amazon.

†Department of Economics, University of Michigan. Email: brinatti@umich.edu
‡IZA Institute of Labor Economics. Email: mingyuc@alumni.princeton.edu
§Department of Economics, University of Delaware. Email: paragma@udel.edu
¶Federal Reserve Bank of Richmond. Email: nicolas.morales@rich.frb.org
‖Queens College, City University of New York. Email: Kevin.Shih@qc.cuny.edu

1 Introduction

The impact of immigration on a receiving economy is heavily influenced by individual firms' recruitment decisions and subsequent performance (Kerr et al., 2015). Yet, two major barriers have limited prior literature from studying the role of firms in shaping the consequences of immigration. First, immigrants tend to cluster in labor markets that feature high-productivity firms, which confounds the relationship between immigration and firm-level outcomes. Second, with some notable exceptions, firm-level data on foreign-born employment is scarce. Hence, a majority of the literature has measured a firm's exposure to immigration at the labor market level (Dustmann and Glitz, 2015; Mitaritonna et al., 2017), ignoring adjustments that occur within the firm and heterogeneity across firm responses.

In this paper, we overcome both of these barriers in a particularly high-stakes setting: the hiring of foreign-born college graduates through the H-1B visa program, the main pathway for high-skill immigrants to work in the US. We study the impact of a plausibly exogenous shock to a firm's ability to hire new immigrants generated by the first large-scale H-1B visa lottery in 2007. We do so by merging employer-employee linked data from the US Census Bureau to a measure that closely approximates a firm's lottery success. Using this rich data set, we find that lottery wins help firms increase their employment, revenues, and survival probabilities. While there are limited effects on college-educated natives, we find heterogeneous responses across firms and specific subgroups of incumbent workers.

We begin by constructing a proxy of a firm's success in the 2007 H-1B lottery. In 2007, for the first time, *all* applications for H-1B workers went through a lottery to determine which visa applicants were allowed to work in the United States. Through a Freedom of Information Act (FOIA) request from the United States Citizenship and Immigration Services (USCIS), we obtained a data set on granted H-1B visas—including lottery wins—along with the Employer Identification Numbers (EINs) of associated employers. We combine this data set with information on intended applications for H-1B visas submitted by firms to the Department of Labor—Labor Conditions Applications (LCAs). We take several steps to ensure that our measure of intended applications closely tracks true lottery applications. Then, using these two sources, we construct our firm-level, lottery-induced shock to immigrant hiring: the number of visas the firm won in the lottery divided by the number of visas the firm intended to apply for—the lottery "win rate."

Next, we merge our lottery measure to restricted access US Census Bureau economic data from the Longitudinal Employer-Household Dynamics (LEHD) program. Our LEHD data set contains the universe of employer-employee matches of twenty-five US states between 2002 and 2011 and includes wages, employment histories, and workers' country of birth, among other variables. We augment the LEHD data with information on firm-level revenue, employment, and payroll from the Longitudinal Business Database (LBD), a comprehensive panel of non-

farm, private sector US firms. The LEHD and LBD provide a unique opportunity to study how workers and firms in the Untied States respond to immigration restrictions. Our main sample consists of firms that participated in the H-1B lottery in 2007 and were successfully matched to the Census. These 20,000 firms alone employed over 16 million workers in 2006, accounting for roughly 13% of total nonfarm, private US employment.

We proceed by setting up an event study analysis to understand the impact of the lottery win rate on firm performance and hiring. This empirical approach allows us to concisely evaluate the plausibility that our lottery measure is exogenous and examine its impact on firm-level outcomes. Importantly, we show that firms with more 2007 lottery success were not trending differently than firms with less 2007 lottery success prior to 2007 across several outcomes, including usage of the H-1B program itself. Quarterly data from the LEHD further allow us to show that even though the lottery took place in April of 2007, responses to lottery success largely started in the fourth quarter of 2007, precisely when workers tied to lottery winning applications could be hired.

Turning to our main firm-level results, we first assess how total firm employment increases when a firm has success in the 2007 H-1B lottery. According to our estimates, one additional lottery win increases firm employment by 0.83 workers relative to a firm that loses the lottery, and this effect is persistent for at least five years after the lottery occurs. The employment effects are stronger for lottery-winning firms with less than 10 employees, who see a full one-for-one increase in their employment relative to lottery-losing firms. Larger firms see much smaller effects on overall employment, suggesting that small firms might be more constrained in their ability to find alternate workers in the event of losing the lottery.

Next, we look at how employment composition at the firm changes after winning the lottery. We corroborate that there is a "first-stage" by showing that lottery winners employ more "likely-H-1B immigrants"—young, low-tenure immigrant workers with college degrees—after 2007. This is a key contribution of our work, as previous studies of H-1B lotteries lacked the demographic information on workers necessary data to establish that winning the lottery is indeed associated with more *immigrant* workers at the firm.

We find limited evidence that lottery wins impact a firm's employment of native workers. We estimate that each lottery win increases employment of native college workers by 0.1 employees, on net (with a 95% CI lower bound of -0.09). Across firm size classes, we similarly rule out crowd out effects on native college workers of more than 0.1 per lottery win with a high degree of confidence. While overall college graduates do not seem to respond significantly to additional H-1B workers, we do find a slight crowd-out of natives that are more similar to the H-1B immigrants such as young, college-educated, low-tenure natives.We also find that the hiring of non-college workers tends to rise in response to lottery wins, consistent with increased demand for complementary workers.

We next evaluate how other firm performance indicators respond to winning the lottery. First, we find that the lottery has important implications for firm survival: firms that win all of their lottery applications are 2.5 percentage points more likely to remain active than firms that lose all of their applications. We also find expansions in firm revenues and payroll that coincide with the increase in scale found in our employment results. Finally, we find suggestive but statistically imprecise evidence that revenues per worker and average wages paid to the employees of winning firms increase in response to lottery wins.

All told, our main firm-level results indicate that lottery wins enable firms to scale up without generating large amounts of substitution away from native workers. Plugging our estimates into a theoretical model with heterogeneous firms and heterogeneous labor inputs yields an elasticity of substitution of 4.6 between immigrant and native college-educated workers.

While our main firm-level results provide new insights on how firms adjust scale and employment composition in response to the availability of immigrants, they may mask important heterogeneous effects across firms and incumbent workers. To dig deeper into these dynamics, we present two sets of additional analyses. First, we explore which firms respond the most to lottery luck. Firms that are skill-intensive, immigrant-intensive, pay high wages, and have higher labor productivity are more likely to expand in terms of employment, and even crowd in native workers, after winning the lottery. We interpret these results as a sign that high-wage and high-productivity firms may be more dependent on workers with specialized skills who may only be available from abroad.

Finally, the employer-employee matched data in the LEHD offers a unique opportunity to differentiate between how the lottery impacts firm-level aggregates and how the lottery impacts *individuals*. We tackle this latter question by estimating individual-level difference-in-difference models that compare the career trajectories of incumbent workers at lottery-winning firms to those of similar workers at lottery-losing firms. We find distinct effects across groups of workers. On one hand, young, college-educated workers with low tenure at the firm experience a 4-5% wage increase when working at a lottery-winning firm, regardless of nativity. On the other hand, high-tenure, young, college-educated natives at lottery-winning firms experience a 5% wage reduction and a 3.6% higher-likelihood of leaving the firm. Our results therefore indicate that it is this *higher*-tenure, young, native college group that may be most substitutable with immigrants among incumbent workers.

Our work makes four main contributions. First, we are among the first papers to use the rich, employer-employee panel data from the LEHD to study the impact of immigrant hiring on firm-level decision-making. Second, we bring credible identification to this question by exploiting exogenous variation induced by the H-1B lottery. Third, we provide new evidence on the degree of substitutability between immigrants and natives within the firm, a key parameter for understanding the welfare impacts of immigration. Fourth, we provide new evidence on how

high-skilled immigration restrictions impact firm performance and incumbent worker career trajectories, key questions for policy-making.

A nascent, growing literature studies the impact of immigrant workers on firm performance and the role of the firm in determining the economic impact of immigration using employer-level data (see, e.g., Amuedo-Dorantes et al., 2023a; Arellano-Bover and San, 2023; Brinatti and Morales, 2023; Brinatti and Guo, 2023; Clemens and Lewis, 2022; Doran et al., 2022; Kerr et al., 2015; Mahajan, forth.; Mayda et al., 2020; Mitaritonna et al., 2017). To our knowledge, the only other paper using the LEHD to study the impact of skilled immigration on firms is Kerr et al. (2015), who study how changes to the H-1B cap affected employment composition within a sample of 319 large firms. In contrast, we exploit lottery variation and use the near-universe of firms that intended to hire H-1B workers in 2007.

A limited number of papers have used the H-1B lottery as a source of firm-level variation in immigrant hiring (Clemens, 2013; Dimmock et al., 2021; Doran et al., 2022; Glennon, 2020; Mandelman et al., 2024). Of particular note is Doran, Gelber and Isen (2022), who use the H-1B lotteries in 2005 and 2006 to evaluate the impact of immigrants on firm employment and patenting. These lotteries were held among the subset of applications that were received on the day that the H-1B cap was met in each year, ultimately covering 2,750 firms that applied for an H-1B visa. In contrast, we focus on the lottery of 2007, where H-1B visa applications exceeded the cap on the first day of filing and were therefore *all* placed into a random lottery. Our analysis covers 20,000 LBD-enumerated firms and 13,500 LEHD-enumerated firms applying for cap-subject H-1B visas in 2007. As such, we complement Doran, Gelber and Isen (2022) by analyzing a more representative sample of H-1B employers.[1] A key additional contribution of our work is that our data contains significant detail on the composition of employment within firms that has not been studied in prior H-1B literature. We are therefore able to directly examine how hiring an H-1B immigrant worker impacts the employment of other worker groups within the firm, including similarly-skilled native workers.

Other papers have focused on the impact of the H-1B program on alternative outcomes such as multinational activity (Glennon, 2020; Morales, 2023), entrepreneurship (Dimmock et al., 2021; Mandelman et al., 2024), occupation choice (Bound et al., 2018; Khanna and Morales, 2021), innovation (Choudhury and Kim, 2019; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010), local employment opportunities in H-1B-related occupations (Peri et al., 2015a), and local productivity (Peri et al., 2015b).[2] We contribute to this literature by looking at the hiring behavior and composition of employment of individual US firms, as well as the adjustment of

[1] Doran, Gelber and Isen (2022) also analyze the effect of 2007 lottery wins on employment and patents as a robustness check on their 2005-06, using a methodology that follows the broad contours of ours but with some important differences. We discuss how we compare results in Section E.

[2] The H-1B program also has the potential to impact higher education (Kato and Sparber, 2013). Literature finds international students, many of whom eventually apply for an H-1B, cross-subsidize domestic students at US universities and increase college selectivity (Bound et al., 2021; Chen, 2021; Chen et al., 2020; Shih, 2017).

individual worker career paths in response to H-1B worker inflows.

Finally, other papers have also combined rich micro-data and causal identification to study the impact of immigration on firms and workers in other contexts. Brinatti and Guo (2023) look at how denials in the H-1B visa in 2017 pushed immigrants to migrate to Canada, and measure the response of Canadian firms. Clemens and Lewis (2022) and Amuedo-Dorantes et al. (2023a) study the H-2B program, aimed predominantly at hiring temporary, noncollege workers in the service sector. Similarly, Egger et al. (2022) exploits the exogenous allocation of refugees across Swiss regions to look at their impact on firms and workers. Dustmann et al. (2016) and Beerli et al. (2021) study the effect of cross-border workers on firms and incumbent workers in Germany and Switzerland, respectively. Most of these results broadly align well with ours: firms that hire immigrants expand more in terms of scale and with either negligible or small positive impacts on native employment while results for individual incumbent workers are mixed. Our setup focuses on the hiring of high-skill immigrants and has the advantage of combining three main ingredients: 1) having a large set of treated firms (the universe of H-1B applicants), 2) using rich employer-employee matched data from the United States, and 3) exploiting a firm-level source of exogenous variation, the H-1B lottery.

2 Overview of the H-1B Lotteries

The H-1B visa was created in 1990 to provide authorization for college-educated foreign nationals to work in specialty occupations in the United States. Since fiscal year (FY) 2004, H-1B annual quotas for new employment at for-profit employers have remained fixed at 65,000 visas under the Regular Cap and 20,000 visas under the Advanced Degree Exemption (i.e., the ADE Cap, for applicants with a master's degree or higher from a US educational institution). These quotas have been binding in each year since FY 2004.[3]

H-1B visas are allocated on a first-come, first-served basis. In the mid-2000s demand for H-1B visas started growing. Figure 1 displays the number of days to reach the H-1B cap from the start of the filing period for each application season, denoted by calendar years 2001-2019. From 2003 application period onward, the cap was reached in successively fewer days from the start of filing.

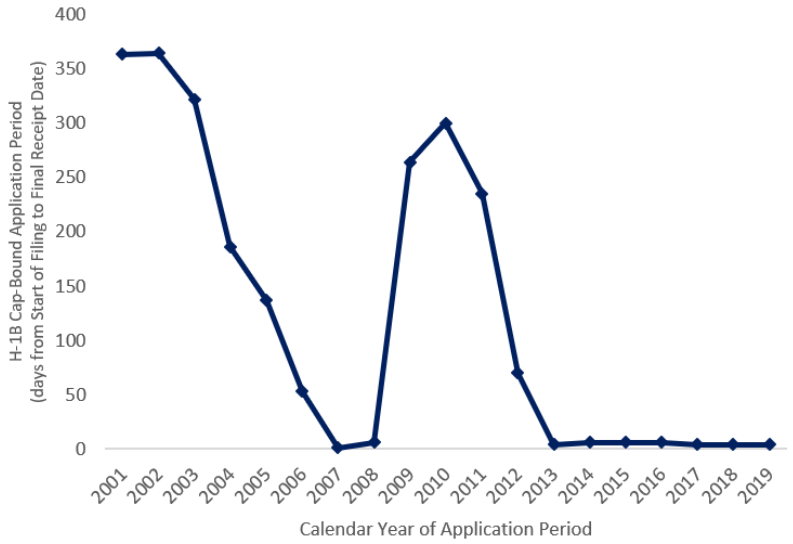
On the first day of the filing period in 2007 (i.e. April 2nd, 2007), USCIS received an unprecedented number of applications that already exceeded the Regular Cap. USCIS held a lottery to distribute all 65,000 visas among the 123,480 applications received on April 2 and 3 of 2007. A lottery was not held to distribute the ADE Cap of 20,000 visas, as the volume of ADE applications remained below the cap until the end of April.[4] This marked the first time that

[3]Universities and other nonprofit/government entities are exempt from these H-1B quotas

[4]In 2005, procedures were updated to allow for randomized selection of petitions under two scenarios. First, petitions received on the "final receipt date"—defined as the date in which the number of petitions exceeds the

the entire Regular Cap was distributed by random lottery – USCIS held small pilot lotteries in 2005 and 2006 to assess feasibility on a sample of just under 3,000 firms (Doran, Gelber and Isen, 2022).[5]

Figure 1: Days-in-Filing for the H-1B Visa Cap, 2001-2019



Note. Figure illustrates the number of days that it took for applications for new, cap-subject H-1B visa workers to meet the cap. Final receipt dates are provided by the USCIS, and we calculate days in filing by counting the number of days from the start of the filing period (generally April 1 or April 2) to the final receipt date.

Hence, the 2007 lottery created an unexpected shock to all firms that applied for H-1B workers. Compared to later lotteries held in 2008 and in each year since 2013, the 2007 lottery was both much-less anticipated and also less complicated. In later years, lotteries were separately held for the Regular Cap and the ADE Cap, and losing applicants in the ADE Cap lottery participated in the Regular Cap lottery.

Finally, it is important to note that some alternatives to the H-1B visa were available to firms that lost the lottery.[6] Interviews conducted by the Government Accountability Office showed that employers that were unsuccessful in the H-1B lottery were able to secure their preferred worker through other, "sometimes more costly" means.[7] This has two implications: 1) whether

cap—would be subject to a randomized lottery. Second, if the final receipt date is the first day of the application period, the entire cap would be randomly allocated across applications received on the first two days of the filing period. See *70 FR 23775, May 2005*: https://www.govinfo.gov/content/pkg/FR-2005-05-05/pdf/FR-2005-05-05.pdf.

[5]The final receipt dates for the FY2006 Regular visa cap, the FY 2006 ADE visa cap, the FY 2007 Regular visa cap, and the FY 2007 ADE visa cap, were August 10, 2005, January 17, 2006, May 26, 2006, and July 26, 2006, respectively.

[6]Alternate pathways to hire foreign-born workers with college degrees include the Optional Practical Training program (OPT), which provides only one to two years of work duration for international students in the years we study. The L-1 visa is available to multinational firms transferring workers to US offices. Lastly, alternative visas were available for specific nations. These include the TN visa for Mexicans and Canadians, the E-3 visa for Australians, and the H-1B1 visa for Chileans and Singaporeans.

[7]See report GAO-11-26 here: https://www.gao.gov/products/gao-11-26.

lottery wins leads to greater skilled immigrant hiring is ultimately an empirical question – we are the first to estimate this relationship in Section 5.4; and 2) even in cases where the lottery does not affect *who* is hired, the additional costs required to overcome lottery losses may still impact measures of firm performance like survival and revenue generation, which we study in Section 5.5.

2.1 The H-1B Application Process

To understand our empirical measurement of firm success in the lottery, we clarify the timeline of the H-1B application process. Employers submit H-1B visa applications on behalf of workers they wish to hire. First, firms must file a Labor Conditions Application (LCA) with the Department of Labor, in which they attest that H-1B workers will not harm incumbent workers, and also provide information about the job, the work start and end dates, the work location, and the associated salary.

After LCA approval, employers may then submit an H-1B visa application, the primary document of which is the I-129 form that provides information about the employer, demographic and other personal information about the worker, the occupation, the work start and end dates, and the wage/salary. The I-129 form is tied to a specific worker and firms need to pay filing fees that range from $2,000 to $10,000, not including attorney fees, which can be substantial.[8]

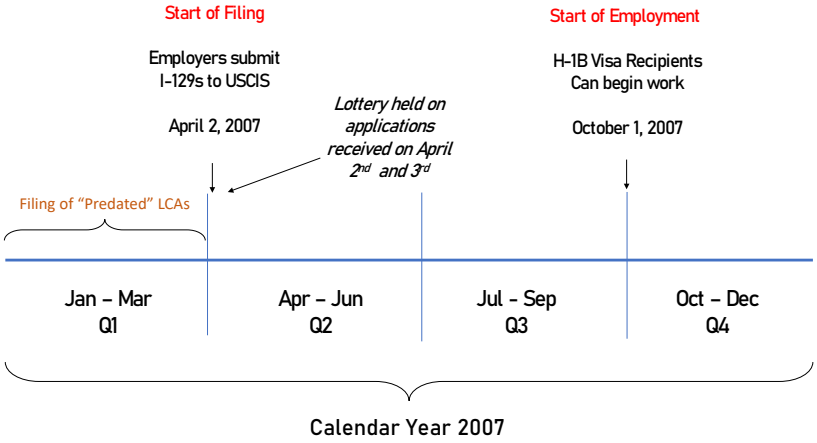
Figure 2 visually depicts the H-1B application timeline for 2007. LCAs were filed in Q1 2007, which we explain in greater detail in Section 4. USCIS began accepting H-1B applications (i.e., I-129 petitions and associated documents) on April 2, 2007. The lottery was then held on those applications received on April 2 and 3.

An important feature crucial to our analysis is that only data on lottery winners was retained. USCIS returned the applications of lottery losers without further processing, and hence there is no data available on the exact number of applications (i.e., I-129 petitions) each firm submitted. We describe how we overcome this using LCA data in Section 4.

Finally, for lottery winners, October 1, 2007 marked the earliest date the H-1B worker could begin working. Hence, Q4 2007 (which is equivalently Q1 of FY 2008) is the earliest workers selected in the lottery could actually start working for the firm.

[8] In addition, a complete H-1B application must also include a copy of the approved LCA, the formal job offer letter, and other supporting documents (e.g., educational transcripts, etc.).

Figure 2: H-1B I-129 Application Timeline



Note. Figure illustrates the application timeline for the 2007 application season. Quarters (Q1-Q4) correspond to the 2007 calendar year.

3 Data Description

3.1 Administrative Employer-Employee Matched Data

We obtain access to a rich collection of employer-employee level data from the US Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) for the period 2002-2011. The LEHD contains the universe of individual worker histories and their associated firms, for which we were granted access to records for twenty-five US states.[9] These states accounted for 53% of the H-1B I-129 petitions for new employment, 48% of total college-educated workers, and 52% of total college-educated immigrants in the United States during our sample period (2002-2011).[10]

The LEHD data contains information on quarterly earnings for each individual-employer pair, taken from employer-reported information to state unemployment agencies. The data also includes individual characteristics such as date of birth, gender, race, education level, and place of birth. Many of these, such as place of birth, are taken from administrative sources such as the Social Security Administration. In some cases—particularly for education—the data is imputed by the Census Bureau using observed education for individuals who participate in the American Community Survey. More details on the LEHD can be found in Vilhuber (2018).

We complement LEHD data by linking it to the Longitudinal Business Database (LBD) at the firm level. The LBD consists of the universe of private sector establishments in the United

[9]We were permitted access to LEHD data from the following states: Arizona, Arkansas, California, Colorado, District of Columbia, Delaware, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Maine, Maryland, Missouri, Montana, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, Pennsylvania, Tennessee, Texas, Washington, and Wyoming.

[10]Calculations done with I-129 records for FY 2003-2012 and Public Use American Community Survey for 2002-2011.

States, with information on employment, payroll, revenues, firm age, industry, and exact location. Furthermore, a majority of LBD firms can be matched to firm-level revenues annually. The LBD includes ownership linkages between establishments, which allow us to aggregate employment and payroll to the firm level. While much previous work on the H-1B program has been limited to the sample of publicly-held firms using databases like Compustat (Mayda et al., 2020), the LBD allow us to look at the near-universe of firms that participate in the program, and the LEHD allows us to examine the near-universe of individual workers in the twenty-five states for which we obtained access.

3.2 H-1B Data

Our H-1B visa information comes from two different sources. First, we obtained individual-level H-1B visa application records from form I-129 through a Freedom of Information Act request from the US Citizenship and Immigration Services (USCIS) for the period 2001-2011. These data include detailed demographic information about the worker, firm-level information, and information about the job position. Worker-level information includes the country of origin, age, and degree type. Firm-level information include firm name, address, and federal tax identification numbers (EIN). Information about the job position include the occupation, salary, and work start and end dates. As noted before, USCIS only retained information of those applications that won the lottery. Hence, the I-129 records are useful in measuring H-1B lottery winners, as explained in Section 4.

A second source for H-1B visa applications are records of Labor Condition Applications (LCA) from the Department of Labor's Office of Foreign Labor Certification for 2001 through 2017. These records provide information from each application, detailing the date of filing, application status, employer name and address, job position, job code, number of positions requested, work start and end dates, work locations, wage/salary associated with the position, and the prevailing wage associated with the area-occupation. We use the LCAs to construct a proxy for lottery-subject visa applications, as described in Section 4.

We merge the Census administrative data with the I-129 and LCA records in two stages. In the first stage, I-129 records are matched to LCA records by first searching for exact matches on firm name and address. All LCA records are retained to keep those with zero lottery wins. Remaining unmatched I-129 records are then linked to LCAs using a fuzzy matching approach developed by Flaeen and Wasi (2015) as follows: fuzzy matches initially are searched over name–street–city–state–zip, then name–city–state–zip, then name–city–state, then name–state, and finally name only. This procedure results in 98% of I129 records being matched to a corresponding LCA record, and 81% of the linked I129-LCA records have EINs (to be used in the second stage of matching).

In the second stage, we linked I129-LCA data to US Census data via the federal tax identi-

fication number (EIN) that is common in both data sets. Records remaining unmatched—including records with zero lottery wins—are then linked via fuzzy matching approach based on the same successively less stringent manner as above. Once completed, we manually inspected the list of companies and removed poor matches to ensure that the quality of the match is high.

4 H-1B Lottery Success: Measurement & Identification

The 2007 H-1B Regular Cap lottery (henceforth, lottery) generated random, exogenous variation in the ability of firms to hire skilled immigrants. Our ideal measure of firm-level lottery success would divide the number of successful lottery applications by a given firm by the total number of lottery applications.

Calculating the number of lottery wins—our ideal numerator—for each firm can be achieved with high precision using granular applicant information contained in our USCIS I-129 records. We classify winning lottery petitions as those received on April 2 or 3 of 2007, from private sector firms (excluding universities and other non-profit institutions). We exclude ADE applications by removing individuals on an F-1 visa that possessed a master's degree or higher at the time of application. Finally, we retain only those applications that requested an employment start date on or after October 1, 2007 – the first day that winning H-1B workers could legally commence employment. These restrictions result in 64,723 winning I-129 petitions, remarkably close to the actual cap of 65,000.

However, while the USCIS kept these detailed records on successful lottery applications, it did not keep records on unsuccessful lottery applications – losing applications were returned to sender without any record. Hence, measuring each firm's exact number of lottery applications—our ideal denominator—is not possible as USCIS I-129 administrative records only contain data on lottery winners. Thus, we next turn to several steps we take to proxy for the number of lottery applications at the firm level.

4.1 Proxying Applications

One potential proxy for the number of lottery applications—with precedent in the H-1B literature—is the number of LCAs filed by a firm prior to April 1, 2007. However, using raw counts of LCAs as a proxy for total lottery applications will induce measurement error in win rates. LCAs are filed prior to the I-129 petition, carry no explicit filing fee, and do not require firms to eventually submit an I-129. Hence, the actual number of H-1B applications is less than or equal to the number of LCAs. Win rates calculated using LCAs are weakly smaller than the ideal win rate using actual H-1B applications.

Such measurement error might induce an endogenous correlation between win rates and firm

outcomes. First, this mismeasurement could reflect a fixed firm-trait – for example, if large firms with in-house legal teams always file many excess LCAs, win rates constructed using LCAs as a proxy for applications would be correlated with firm size. Second, such error may arise from a pre-trend or pre-lottery shock to firm outcomes. For example, after having submitted their LCAs, a firm might experience a negative shock that precludes submission of the I-129 application. Under this scenario, win rates using LCAs would be endogenously related to pre-lottery changes in firm outcomes.

We develop several refinements and an empirical strategy to more accurately proxy applications and reduce the scope for endogeneity, thereby improving upon prior research (Dimmock et al., 2021; Glennon, 2020; Peri et al., 2015a) which has only relied on simple counts of LCAs. We then demonstrate through balance tests, evaluation of pre-trends, and other checks, that our corrected measure (which refer to as *Win Rate*$_j$ and define explicitly later) captures plausibly exogenous variation in win rates within the confines of our difference-in-differences estimation strategy.

First, analysis of the pattern and timing of LCA filings suggests that LCAs filed between March 1 and April 3 of 2007 are more likely lottery-bound than LCAs filed on other dates.[11] Second, we remove LCAs filed during this period that were for non-lottery H-1B applications by calculating the total *non-lottery* I-129 filings of each firm (e.g., renewals, change of status, etc.) between March 1 and April 3 and subtracting this number from the LCA count.[12]

Third, we address remnant mismeasurement by identifying and removing observations whose proxied applications (i.e. the count of LCAs after completing the two refinements above) remain improbably small or large given the number of lottery wins that we observe. This "filtering" procedure exploits information on the true win probability for any application, which USCIS publicly reported to be $\frac{65{,}000}{123{,}480} = 0.5264$. The number of winning applications for a given firm follows a binomial distribution with success probability 0.5264 and number of trials given by the number of applications.

Given this, we remove observations for which the probability of observing the pair of our proxied application count and number of lottery wins is less than 1%.[13] To clarify by way of example, the probability of observing a firm with 0 lottery wins and a proxied number of applications equal to 7 (or more) would be less than 0.01 when the binomial success probability is 0.5264. Our filtering procedure therefore removes observations with 0 observed wins and 7 or more

[11] We essentially find strong patterns of "predating" (Peri et al., 2015a) that begins to occur on March 1st, 2007. We also require that LCAs in this date range maintain a six-month duration between LCA submission date and employment start date (i.e., employment start dates between August 1 and October 3). We detail these issues related to "predating" and the six-month restriction in Appendix Section B.

[12] Rules stipulate that each I-129 application requires a previously successful, corresponding LCA application. This applies to renewals, change of status, and other types of applications to amend H-1B petitions. We remove any observations for whom Likely-Lottery Applications$_j \leq 0$ after this step.

[13] Further details of our filtering strategy and its utility can be found in Appendix Section C.

proxied applications.

These refinements culminate in our preferred win rate proxy:

$$\text{Win Rate}_j \equiv \frac{\text{Lottery Wins}_j}{\text{Likely-Lottery Applications}_j}, \tag{1}$$

where the denominator *Likely-Lottery Applications*$_j$ reflects the three refinements previously described. We now summarize the implications of these refinements on our win rate measure.

4.2 Win Rate Summary Statistics

Table 1 summarizes our measurement of Win Rate$_j$. Columns (1) and (2) of Table 1 are constructed at the Employer Identification Number (EIN) level using only the linked I129-LCA data.[14] Column (3) contains disclosed statistics from US Census data, reported at the "firmid" level – the Census Bureau's internal firm identifier. "Firmid" is more aggregated than EIN, representing the highest possible organizing unit for business entities.[15]

Several key points emerge from Column (1), which is tabulated from our *unfiltered* sample of linked I-129-LCA data. Despite our initial restrictions – only counting LCAs from March 1-April 3 and removing non-lottery I-129 filings from LCA counts – our proxy for applications *Likely-Lottery Applications*$_j$ still overestimates the true number of lottery applications reported by USCIS (182,951 vs 123,480), resulting in an estimated overall fraction of winning applications of 30% (i.e. 55,565/182,951) instead of 53%. The average win rate calculated across EINs (46%) is much larger than this overall fraction of winners (30%). These differences suggest that without further refinement, nonrandom measurement error in win rates may be sizable.

The utility of our filtering method can be seen by comparing Column (2) to Column (1). Removing probabilistic outliers greatly reduces the average application count (e.g., from 9 in Column (1) to 4 in Column (2)), while only slightly reducing average wins. Thus, filtering brings the average win rate across firms (47%) into parity with the overall fraction of winners within this sample (35,903/78,080≈46%)—a key indicator that our win rates from the filtered sample are more likely to capture lottery variation. Column (3) similarly displays disclosed statistics from our primary analysis sample after filtering and also after aggregating EINs to the "firmid" level (hereafter, firms). Compared to Column (1), average applications are also smaller, and the average win rate in this filtered sample (41%) is once again nearly identical to

[14] In order to avoid excessive disclosure avoidance review we report these exclusively from I-129/LCA data at the EIN level.

[15] The variable "firmid" is the Census Bureau's internal identifier for "firms," which represent the highest possible organizing unit for a business entity. Firms may comprise of one or multiple different establishments. The EIN, in many cases, does not represent the firm, as firms may often be associated with various EINs, each representing subdivisions of the business.

Table 1: Estimated FY 2008 Lottery Characteristics

	Sample		
	(1)	(2)	(3)
	EIN sample	EIN sample (filtered)	LBD firmid sample (filtered)
Panel A: *Lottery Wins*$_j$			
Mean	2.77	1.89	1.62
Std. Dev.	(32.15)	(11.68)	(5.52)
Total	55,565	35,903	32,170
Panel B: *Likely-Lottery Applications*$_j$			
Mean	9.11	4.11	3.68
Std. Dev.	(93.21)	(21.89)	(10.81)
Median	1	1	–
Total	182,951	78,080	73,180
Panel C: *Win Rate*$_j$			
Mean	0.46	0.47	0.41
Std. Dev.	(0.42)	(0.42)	(0.42)
Panel D: Other Lottery Characteristics			
EIN Count	20,072	18,963	–
Prop. EINs with *Likely-Lottery Applications*$_j = 1$	0.52	0.55	–
Prop. EINs with *Wins*$_j = 0$	0.35	0.36	–
Prop. EINs with *Lottery Wins*$_j = 1$	0.44	0.46	–
Prop. EINs with *Lottery Wins*$_j > 1$	0.21	0.18	–

Note. Column (1) presents summary statistics for the full sample of likely-lottery participants measured with our I-129 data. Column (2) presents the results for the sample of firms with valid win rates after the filtering process explained in Section C. Columns (1) and (2) consider a "firm" as a unique EIN which stands for Employer Identification Number and are calculated using only our I-129 data. Column (3) calculates sample statistics for LBD firms, which is the firm identifier at the Census. EINs match with but do not correspond exactly to the LBD firm IDs used in our analysis. There are often multiple EINs per LBD firm ID. With the exception of this table, an LBD firm ID is generally what we mean when we reference firm j.

the overall share of wins (32,170/73,180≈40%).[16]

In Panel D we observe skewness in both wins and applications, which helps contextualize the results in Section 5. While the average number of wins is almost two and the average number of applications is four, standard deviations are large, indicating the presence of a small number of firms that file many applications. Panel D shows that roughly 50% of EINs only file one application, and hence either win zero or exactly one H-1B petition. The share winning more than one is less than 20%. Given that our regressions are conducted at the firm-year level and do not weight by firm size, much of our identifying variation comes from firms that apply for only one H-1B worker.

[16]We are unable to report EIN level statistics in Panel D for Column (3) since the unit of analysis is at the firmid level.

5 Empirical Analysis

5.1 Difference-in-Difference Event Study Approach

We first detail how we operationalize our empirical approach and then discuss identification, before turning to results. We estimate continuous difference-in-difference event study models specified by Equation (2):

$$y_{jt} = \sum_{\tau \neq b} \beta_\tau \left[\text{Win Rate}_j \times \mathbb{1}(\tau = t)\right] + \Gamma X_{jt} + \alpha_j + \alpha_{kt} + \varepsilon_{jt}, \tag{2}$$

where j denotes the firm. For LEHD outcomes available at a quarterly frequency, t indexes a calendar quarter, in which case the omitted period is $b = 2007Q1$ (the quarter before the lottery took place). For outcomes available at an annual frequency, t indexes a calendar or fiscal year, in which case $b = 2006$. Outcomes y_{jt} include firm-level variables such as revenues, total employment, and employment of specific subgroups. In most cases, we apply the inverse hyperbolic sine transformation to outcomes to approximate log changes while retaining zeros.[17]

Firm fixed-effects (α_j) account for time-invariant differences across firms that might correlate with win rates and outcomes, including a firm's tendency to apply for excess LCAs. Four-digit industry-by-time dummies (α_{kt}) help absorb aggregate shocks (e.g. the Great Recession) that affect all firms similarly. Finally, our control variables (X_{jt}) include pre-lottery firm employment interacted with time fixed effects.[18] These control variables play two roles: 1) absorbing the potentially large amount of residual variation specific to industry and firm size that was unrelated to the lottery but nonetheless created by the Great Recession; 2) removing the influence of systematic, time varying differences in the propensity to apply for excess LCAs that are either correlated with initial size or learned within industry.

Equation (2) directly helps us assess concerns regarding the win rate proxy $Win\ Rate_j$. Generally, we can use estimated $\hat{\beta}_\tau$ for pre-lottery τ to evaluate whether firms with different win rates were following similar trends before the lottery. When estimating Equation (2) at a quarterly frequency, we can be even more precise: $\hat{\beta}_\tau, \tau \leq 2006Q4$ represent pure balance tests, while $\hat{\beta}_\tau, \tau \in \{2007Q2, 2007Q3\}$ allow us to examine whether firms began reacting to the results of the lottery in the two quarters between the lottery itself and the H-1B worker start date of October 1, 2007 (the first day of 2007Q4). While we expect immediate post-lottery effects[19], the direct impact of hiring of an immigrant worker through the H-1B program should be most

[17] In Section D.1, we show that our results are robust to alternate ways transforming our key outcome variables following the suggestions of Roth and Chen (2022).

[18] We interact the inverse hyperbolic sine of a firm's employment in March 2007 (pre-lottery) interacted with time-period indicators.

[19] Lottery results were known by May for most firms. Winning firms may have immediately undertaken other investments.

reflected in $\hat{\beta}_\tau, \tau \geq 2007Q4$.

The other utility of Equation (2) is that it allows us to examine post-lottery dynamics in detail. We are particularly interested in whether the lottery generates a permanent advantage for winners or whether such an advantage dissipates. When estimates from Equation (2) are roughly in line with a one-time, permanent jump in a secondary outcome y_{jt}, we sometimes present estimates from a standard continuous difference-in-difference specification:

$$y_{jt} = \beta \left[\text{Win Rate}_j \times \text{Post}_t\right] + \Gamma X_{jt} + \alpha_j + \alpha_{kt} + \varepsilon_{jt}, \tag{3}$$

where the coefficient of interest, β, measures the impact of the win rate on firm-level outcomes for the post-lottery period relative to the pre-lottery period. Here, LEHD outcomes are measured in calendar Q4 of a given year, and t always indexes a year. In most specifications, $Post_t$ is therefore an indicator for $t \geq 2007$.[20] Estimates from Equation (3) help us paint a more complete picture of the lottery's effects in a concise way while avoiding excessive disclosure review burden on the US Census Bureau.

We limit the sample to firms that apply for at least one lottery-subject LCA in 2007 according to our proxy and who survive the outlier filtering strategy laid out in Section 4.1. In the end, we can analyze roughly 13,500 firms using the LEHD and roughly 20,000 firms using the LBD in fully balanced panels from 2002-2011. The former sample is restricted to lottery applicants operating in at least one of the twenty-five states in our LEHD data listed above but allows us to analyze employment of specific subgroups. The latter sample contains all filtered lottery applicants in 2007 but only allows us to analyze total firm employment, total firm payroll, total revenues, and other outcomes that can be generated from these three basic measures. These panels are fully balanced, as values are changed from missing to 0 if the firm is inactive (has zero total LBD payroll for the year).

5.2 Identifying Assumptions

Causal identification in our setting requires that our win rate measure, conditional on covariates specified above, is unrelated to other unobserved determinants of firm employment and performance. Crucial to this is whether our corrections to the win rate sufficiently mitigate potentially endogenous measurement error.

As stated earlier, excess LCA filing that is reflective of a fixed-firm trait is accounted for by firm fixed-effects. The difference-in-differences design effectively eliminates any confounding influence from fixed factors, like firm size. Additionally, we find no evidence of pre-lottery shocks or pre-trends that cause firms to submit fewer H-1B applications than their LCAs filed. Throughout the presentation of results in Section 5, our event study analysis finds no significant

[20] The lone exception is LBD-measured employment, which is measured in March of a given year. For this outcome, $Post_t$ is therefore an indicator for $t \geq 2008$.

pre-trends across the full range of our outcome variables in the years or quarters leading up to the 2007 lottery. Key outcomes do not respond until Q4 2007, when winning H-1B workers are legally allowed to begin working at firms.

We also perform a more targeted balance check to assess the scope for pre-lottery shocks by unconditionally regressing pre-lottery changes in firm outcomes on their win rate. We measure changes in outcomes in the years prior to the lottery, but also in the first and second quarters of 2007, between when LCAs are filed (i.e., January-March 2007) and when H-1B applications must be submitted (April 2007). The results of these balance checks are presented below, in Table 2.

Table 2: Pre-lottery Balance Test for Key Outcomes

	Δ Employment LBD March 06 - March 07	Δ Pay LBD March 06 - March 07	Δ Average Wage LBD March 06 - March 07	Δ Employment LEHD Q1 06 - Q1 07	Δ Employment LEHD Q1 07 - Q2 07
Win Rate$_j$	0.001	-0.006	-0.059	0.023	-0.004
	(0.032)	(0.149)	(0.106)	(0.086)	(0.012)
N Obs	20,000	20,000	20,000	13,500	13,500

Note. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. These are cross-sectional regressions of our measured win rate on growth rates of key outcomes, with no other controls included. Number of firms is rounded as per the US Census Bureau's disclosure rules. Columns 1-3 present the results for the growth rate in employment, payroll, and wages between March 2006 and March 2007, as measured by the LBD. Column 4 regresses the win rate on the employment growth rate between Q1 of 2006 and Q1 of 2007 for the LEHD sample of firms. Finally, Column 5 regresses the win rate on the growth rate in LEHD employment between Q1 2007 and Q2 2007.

Columns (1)-(3) examine the growth rate in employment, payroll, and average wages (payroll per worker) from March 2006 to March 2007 (just prior to when the lottery occurred). Column (4) examines the change in employment from Q1 2006 to Q1 2007 from the LEHD sample. Finally, Column (5) uses an outcome uniquely available to us through the LEHD: employment changes between Q1 2007, when LCAs are filed, and Q2 2007, when H-1B I-129 applications are filed.

Finally in appendix C we demonstrate that our corrected win rate performs similarly to win rate obtained from a simulated lottery on our data. Our filtering correction reduces the negative correlation between win rates and applications that arises because applications are on average too large, and reduces left-skewness apparent in the unfiltered distribution of win rates due to extra mass in small win rates (i.e., since applications again are too large, win rates are too small, on average).

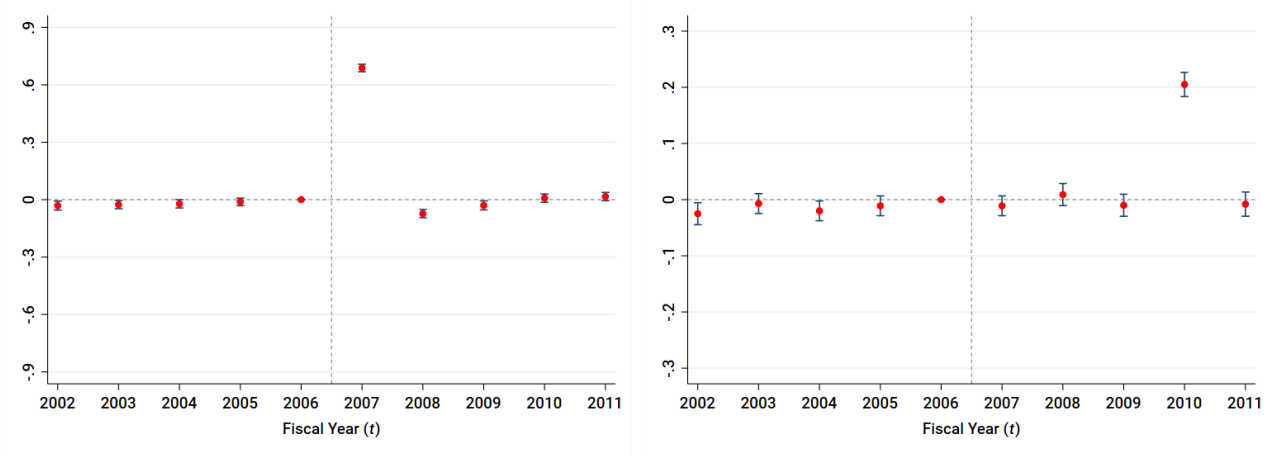
Taken together, these checks support our identification strategy, and that our corrections for measurement error help reduce the scope for endogeneity bias. We now turn to our results on H-1B hiring and asses whether greater success in the lottery translates into increased H-1B hiring.

5.3 H-1B Hiring

We begin the analysis by testing whether our measure of lottery success is consistent with the expected firm-level responses in I-129 approvals for new employment and employment renewal. A firm that is successful in the lottery should have a pronounced, mechanical increase observed I-129 approvals for new employment during 2007. Further, if successful firms exercise their option to renew H-1B workers, we expect to see higher I-129 records for renewal three years later in 2010. Perhaps most importantly, if we have isolated lottery variation, we should not expect to see significant estimates of $\hat{\beta}_\tau$ for new employment or renewals prior to 2007.

Figure 3: Effect of Lottery Win Rate on Granted H-1B Petitions

(a) New Employment (b) Renewals



Note. See Equation (2) for specification, where t indexes a fiscal year and $b = 2006$. We plot the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. Vertical dashed line separates pre-lottery years (balance tests) from post-lottery years (treatment effect estimates). Omitted period is 2006. Number of firms is roughly 20,000 (rounded as per the US Census Bureau's disclosure rules). Standard errors clustered at the firm level. The dependent variables is the inverse hyperbolic sine of the count of I-129 petitions for new employment (Panel a) or renewals (Panel b) filed in year t for workers starting in fiscal year $t + 1$. These data are from the USCIS. Regressions include firm fixed effects, industry-time fixed effects, and the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects.

Figure 3 presents estimates from Equation (2) for I-129 approvals and provides several pieces of validation for our approach. First, it shows negligible divergence in pre-trends between winning and losing firms in I-129 approvals for new H-1B workers. Interestingly, the expected, mechanical spike in 2007 is followed by significant drops in I-129 approvals for new H-1B workers in 2008 and 2009. We believe the most plausible explanation is that 2007 lottery losers were especially motivated to apply for H-1B workers in subsequent years. Meanwhile, Figure 3b shows that winning firms are substantially more likely to have I-129 records for H-1B employment renewals in 2010. We do not find an effect on renewals in any other year. Given that renewals are not subject to the lottery, we view these precise zeros as further corroboration of our identifying assumptions.

5.4 Workforce Composition

We start our main results by asking how the hiring of H-1B workers alters the worker composition of a firm, utilizing the unique data contained in the LEHD. We focus on two critical questions: 1) does the provision of H-1B visas found in Figure 3 translate into the utilization of workers with characteristics of H-1B visa holders in the LEHD data? And, 2) the canonical question of whether immigrant workers crowd out native counterparts with similar characteristics, studied here in the context of the H-1B program. Our answers to these questions are largely contained in Figure 5, and we provide further context and details in the text below.

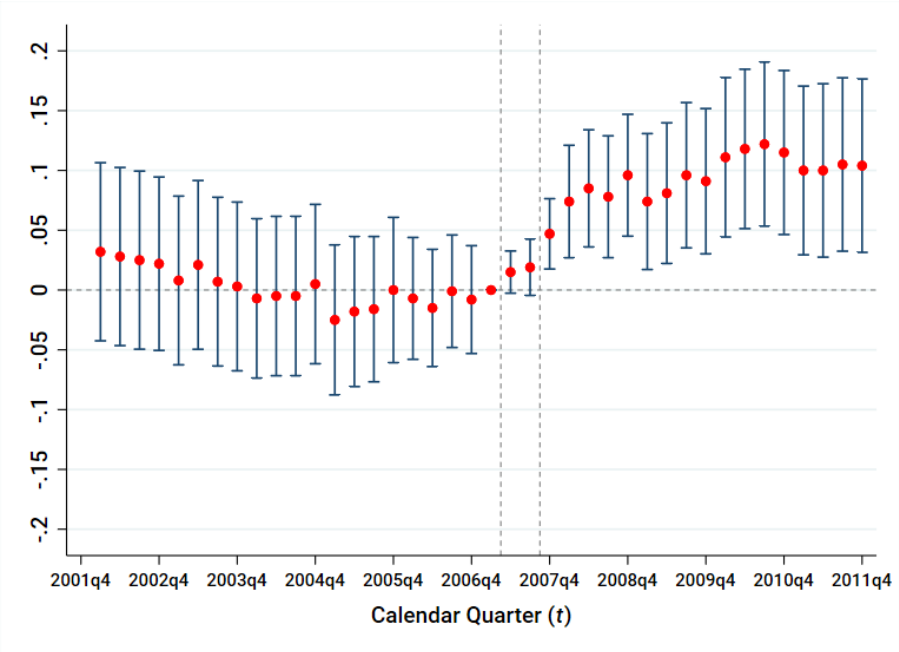
First, however, we ask the basic question of whether lottery wins increased total firm employment. Figure 4 plots estimated coefficients from Equation (2) where the outcome is the inverse hyperbolic sine of total firm employment. Echoing results from Section 5.3, lottery winners and losers were not trending differently in terms of total employment prior to the realized outcomes of the lottery. We then see a discrete jump after the start of the H-1B hiring period (2007Q4) that lasts throughout our study period—a permanent increase in employment that lasts beyond the initial H-1B employment spell of three years. As we discuss further in Section 5.4.2, our results imply that each lottery win leads to 0.83 additional workers in total at the firm starting in 2007Q4 (implying an estimate of 0.17 workers crowded out per H-1B lottery win, on net).

To delve into this initial result, we next test whether lottery wins translate to increased employment of "H-1B-like" immigrant workers. This check is a key contribution of our project relative to prior literature that has studied the H-1B lotteries (Dimmock et al., 2021; Doran et al., 2022; Glennon, 2020). Due to the lack of firm-level data on immigrant and native employment composition, prior empirical literature was not able to confirm whether H-1B lottery wins increased immigrant employment at the firm level. Meanwhile, a standard theoretical framework with perfectly competitive markets for highly educated, young foreign-born workers would imply null effects. Hence, it is important to first ask whether a larger immigrant college workforce is an important mechanism through which lottery success affects firm-level outcomes.

As we cannot directly identify workers under H-1B visas in the LEHD, we define H-1B-like immigrants as those workers who were born outside of the United States and Puerto Rico, who have a college degree, who are between the ages of 25 and 40, and who have less than three years of tenure at the firm. This definition is based on the characteristics of most H-1B workers but neither captures all H-1B workers nor excludes all non-H-1B workers.[21] We also test the effect of lottery wins on a firm's total immigrant college workforce—the number of foreign-born

[21]Some reasons why this measure is likely broader than the true H-1B workers at the firm is that it may capture those working on an L-1 status, OPT, or a green card. Further, such workers may even be naturalized citizens. Even under an H-1B status, they might have been working at other US firms the years before and thus would not be subject to the 2007 lottery. However, this measure may also miss some H-1B workers, such as those who are under 25, above 40, or who are incorrectly coded as not having a college degree in the LEHD.

Figure 4: Effect of Lottery Win Rate on Employment on Total Firm Employment



Note. See Equation (2) for specification, where t indexes a calendar quarter and omitted period $b = 2007Q1$. We plot the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. First vertical dashed line separates pre-lottery quarters (placebo tests) from quarters in which lotteries may have affected behavior but before 2007 lottery H-1B workers could be hired (anticipation tests). Second vertical dashed line separates pre-lottery-hire quarters from post-lottery quarters (treatment effect estimates). Number of firms is roughly 13,500 (rounded as per the US Census Bureau's disclosure rules). Standard errors clustered at the firm level. Dependent variables in each subfigure are the inverse hyperbolic sine of the worker count from the labeled group. These outcomes are measured using the LEHD. Regressions include firm fixed effects, industry-time fixed effects, and the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects.

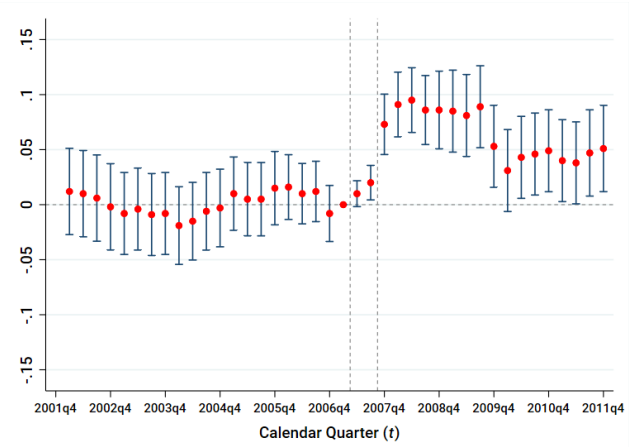
workers classified as having a college degree in the LEHD—and the effect of lottery wins on the total immigrant workforce, regardless of educational attainment. These measures are less likely to miss H-1B immigrant workers but are also likely to contain more non-H-1B immigrant workers.

Figures 5a, 5c, and 5e show that lottery-winning firms increase their use of workers in each category. First, Figure 5a shows a clear increase in winners' employment of H-1B-like immigrant workers immediately after the 2007 lottery. Given that this group has less than three years of tenure at the firm, the hump-shaped response displayed in Figure 5a is specifically consistent with increased hiring of H-1B workers in 2007, as seen in Figure 3a. Figures 5c and 5e further show a permanent increase in the employment of foreign-born workers with college degrees and foreign-born workers more generally during our study period, potentially due to the response in renewals seen in Figure 3b.

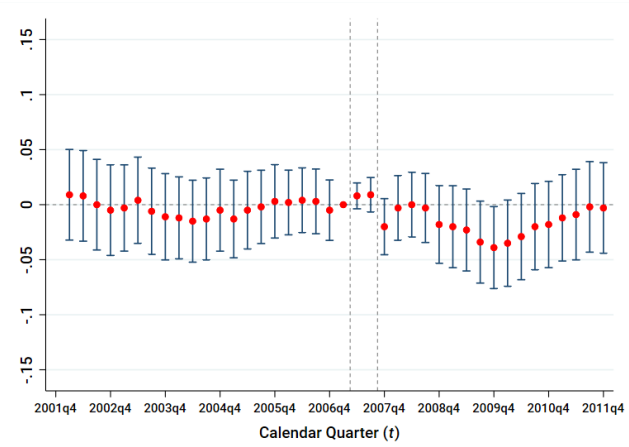
Our estimates imply that firms increase their hiring of (college-educated) immigrant workers by around 10% when they win all of their lottery applications. As discussed further in Section 5.4.2, this translates into each H-1B lottery win leading to 0.29 additional immigrant workers with a college degree. Thus, while there is a permanent increase in immigrant hiring among

Figure 5: Effect of Lottery Win Rate on Employment of Selected Subgroups

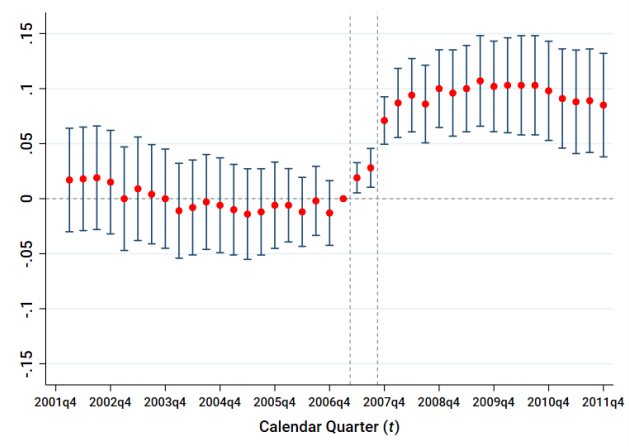
(a) H-1B-Like Immigrants



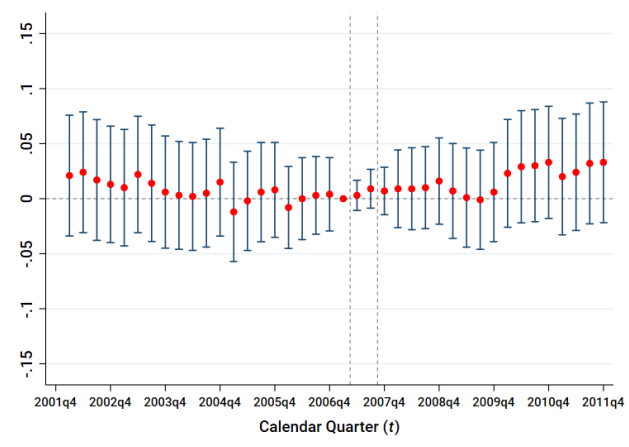
(b) H-1B-Like Natives



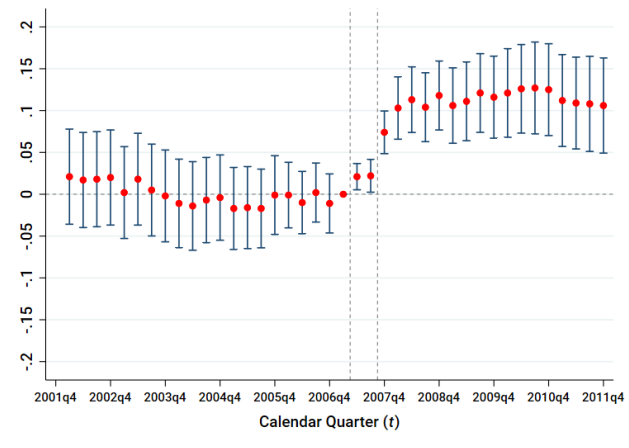
(c) College Immigrants



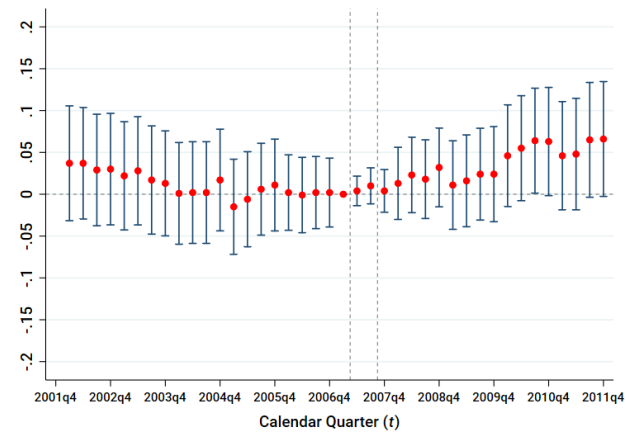
(d) College Natives



(e) All Immigrants



(f) All Natives



Note. See Equation (2) for specification, where t indexes a calendar quarter and omitted period $b = 2007Q1$. We plot the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. First vertical dashed line separates pre-lottery quarters (placebo tests) from quarters in which lotteries may have affected behavior but before 2007 lottery H-1B workers could be hired (anticipation tests). Second vertical dashed line separates pre-lottery-hire quarters from post-lottery quarters (treatment effect estimates). Number of firms is roughly 13,500 (rounded as per the US Census Bureau's disclosure rules). Standard errors clustered at the firm level. Dependent variables in each subfigure are the inverse hyperbolic sine of the worker count from the labeled group. These outcomes are measured using the LEHD. Regressions include firm fixed effects, industry-time fixed effects, and the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects.

H-1B lottery winners, there is substantial scope for lottery losers to find other ways to hire immigrant workers.[22] Still, in total, the left panels of Figure 5 demonstrate that the H-1B lottery wins permanently increase firm-level immigrant intensity.

To further explore whether we are capturing H-1B hiring in the LEHD, we split H-1B-like immigrants by nationality. Results in Table A3 of Appendix Section D.2 show that lottery wins lead to large increases in employment of Indian and other non-North-American H-1B-like immigrants but declines in employment of Mexican and Canadian H-1B-like immigrants, who can also be hired on TN visas. Thus, we find specific evidence for one alternate channel through which lottery losing firms may fill their demand for highly educated immigrant workers.

We next turn to our key estimates on immigrant-native substitutability in the right panels of Figure 5, which can be directly compared to their corresponding left panels. Figure 5b presents evidence of limited substitution away from native workers who we consider most substitutable with H-1B workers—those native workers who are 25-40 years old, with less than three years of tenure at the firm, and with a college degree.[23] Meanwhile, Figure 5d does not show any evidence of a decline in the native college workforce more broadly, and Figure 5f similarly does not show a decline in the overall native workforce at a lottery-winning firm. As described in Section 5.4.2 below, our results imply that each H-1B lottery win increases employment of native college workers by 0.1 (-0.09,0.29). We further explore the possibility of complementarities between H-1B and non-H-1B immigrant workers, along with effects on incumbent native workers in Section 7.

We note that across all panels in Figure 5, we find at-most limited evidence of anticipation effects that occur after the lottery has been resolved but before H-1B workers are hired ($\hat{\beta}_{2007Q2}$ and $\hat{\beta}_{2007Q3}$). In particular, firms do not appear to shed native workers—on net—in anticipation of hiring new immigrant workers through lottery wins. If anything, there may be small, positive anticipatory effects on the hiring (or retention) of immigrant workers with college degrees. Part of this small anticipation effect might be due to H-1B workers being at the firm before the lottery takes place under an OPT. After the lottery happens, workers with OPTs at losing firms are more likely to leave than those at winning firms.

5.4.1 Differences by Size

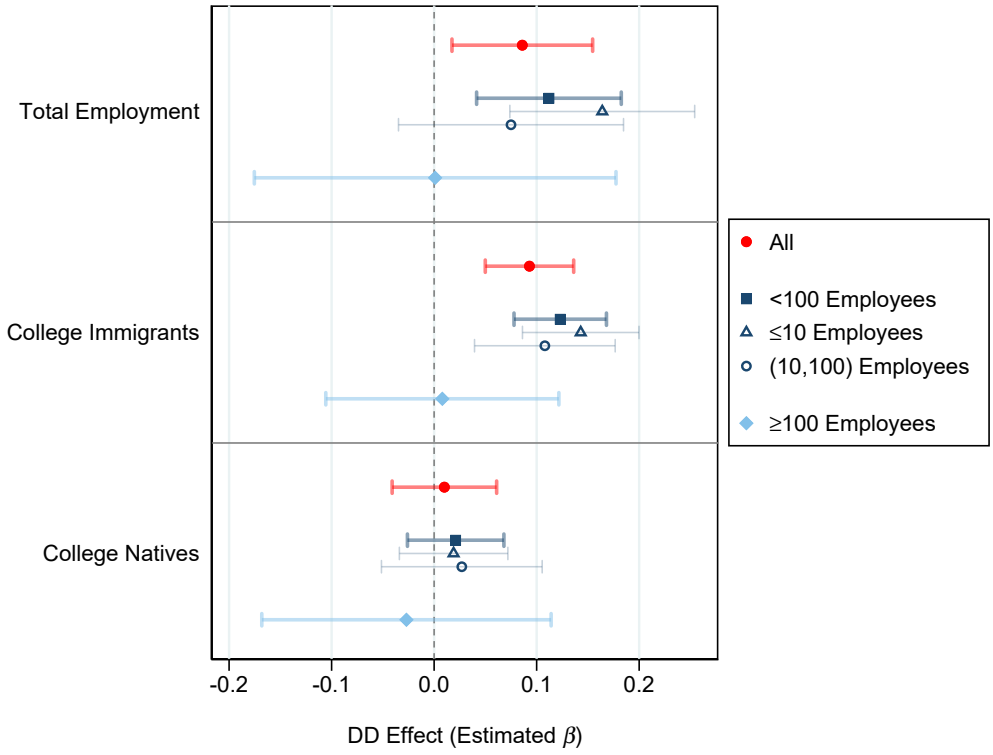
As discussed in Section 4, much of our identifying variation comes from firms that apply for exactly one H-1B worker. For these, likely smaller, firms, the USCIS essentially conducts coin flip to determine whether the firm will receive an H-1B worker. For very large employers who tend to apply for many H-1B workers, the lottery introduces substantially less randomness in workforce composition because the lottery is held at the application level.

[22] As discussed above, this echoes qualitative findings in a Government Accountability Office report to Congress regarding the H-1B program and suggestive evidence provided in the appendix of Doran, Gelber and Isen (2022).

[23] This exactly mimics our definition of "H-1B-like immigrant" workers except for nativity.

To confirm this intuition and provide further context around our primary results on employer composition, we present heterogeneity by initial firm size in Figure 6. In it, we plot estimated coefficients from Equation (3), where the outcome is the inverse hyperbolic sine of total employment, employment of college immigrants, or employment of college natives. We run these difference-in-difference regressions, splitting employers based on whether they were at least 100 employees in 2007Q1 (light blue diamonds) or less than 100 employees in 2007Q1 (navy blue squares). We then further split the less than 100 employee results into regressions on firms with between 10 and 100 employees in 2007Q1 (hollow navy blue circles) and those with less than 10 employees in 2007Q1 (hollow navy blue triangles). For comparison, we also plot overall effects, across all size classes, in red circles.

Figure 6: Heterogeneity by Initial Employer Size



Note. Each plotted coefficient comes from a separate regression estimated using Equation (3), with t indexing a calendar year and omitted period $b = 2006$. Outcomes (labeled on the left) are transformed using the inverse hyperbolic sine. Each regression is estimated using a sub-set of firms based on their 2007Q1 employment, as labeled in the legend. Employment counts that generate sample splits are measured using the LBD.

Two key features emerge from Figure 6. First, identifying variation stems primarily from smaller firms. Intuitively, firms with large numbers of applications have low variation in win rates due to the law of large numbers. This ultimately manifests in the tighter standard errors for small firms in Figure 6 and the fact that the overall results ("All") closely match the results for firms under 100 employees. Second, even if we ignore standard errors and take point estimates at face value, the lottery appears to have no impact on larger firms—either in terms of total

employment counts *nor* native crowd out. Thus, point estimates indicate that larger firms are able to fill their demand for college immigrant workers through non-lottery channels even if they are left unexpectedly short by the lottery. Indeed, we find evidence consistent with the notion that firms of all sizes find other ways to hire college immigrants if they lose the lottery in Section 5.4.2, and that this is especially true for larger firms. We therefore conclude that our results are most relevant for the roughly 75% of our lottery participants who originally had less than 100 employees.

5.4.2 Magnitudes

Our event study coefficients can be approximately interpreted as the percent change in a given outcome for a firm had it won all of its lottery applications relative to if it had lost all of its lottery applications. To obtain an alternate interpretation that more directly speaks to magnitudes in our employment effects, we present an an instrumental variable strategy that allows us to quantify how much one additional H-1B win increases the employment of different types of workers. The second stage regression is a modified version of our difference-in-differences specification from Equation (3):

$$\text{Scaled Employment}_{jt}^{g} = \beta_m[\text{Scaled Lottery Wins}_{j,2007} \times \text{Post}_t] + \Gamma X_{jt} + \alpha_j + \alpha_{kt} + \varepsilon_{jt}. \quad (4)$$

In Equation (4), $\textit{Scaled Employment}_{jt}^{g}$ is the employment of a given group of workers g at firm j in time t divided by total firm employment in 2006Q4, as reported in the LEHD. $\textit{Scaled Lottery Wins}_{j,2007}^{g}$ is similarly the number of 2007 lottery wins divided by total firm employment in 2006Q4 (the same denominator). The interpretation of β_m, then, is the number of additional employees in group g per lottery win in the post-treatment period. Control variables remain the same as in Equation (3). Estimating Equation (4) use OLS would yield biased estimates of β_1, as the raw number of lottery wins (not divided by applications) reflects several endogenous firm-level factors, including firm size and firm demand for H-1B workers. Thus, we use $\textit{Win Rate}_{j,2007} \times \textit{Post}_t$ as an instrument for $\textit{Scaled Lottery Wins}_{j,2007} \times \textit{Post}_t$ and estimate Equation (4) using two-stage least squares (2SLS).

Table 3 presents the results for the full sample, as well as for small firms (less or equal than 10 employees) and medium size firms (between 10 and 100 employees). We focus on the responses of total employment, and three sub-groups of workers: college-educated immigrants, college-educated natives and non-college workers. We find that one additional H-1B lottery win increases total firm employment by 0.83 workers. Of those 0.83 workers, 0.29 are college-educated immigrants, 0.1 college-educated natives, and 0.44 non-college-educated workers.

These numbers highlight two main results. First, total employment at the firm level increases by less than 1 and college-educated immigrants increase by substantially less than 1. In our

view, the most plausible explanation is that lottery-losing firms are able to hire immigrants through other channels despite losing the lottery. Second, the H-1B immigrants seem to crowd-in other workers, such as non-college workers, as lottery-winning firms *expand* their usage of workers outside of the immigrant college group. Such results are consistent with, e.g., non-college workers being complements to highly-educated immigrants. We incorporate such complementarities into our theoretical framework in Section 5.6.

Given that our lottery variation is driven by smaller firms, we proceed by looking into how these magnitudes change for different employment size groups. For firms with at most ten employees as of 2007, one additional lottery win increases total employment by one, and half of that is accounted by college-educated immigrants. Interestingly, for small firms, there is a crowd-in of 0.18 college educated-natives, which indicates that the lottery eases a constraint that allows them to expand in all types of workers relative to losing firms. For the group of firms with employment between ten and one hundred, the effects are more muted. Total employment for winning firms increases by 0.2 employees for each additional lottery win and it is almost fully accounted by an increase on college-educated immigrants. The smaller magnitudes for college-educated immigrants highlight that bigger firms might find it easier to hire immigrant workers in the event of losing the lottery.

Our employment per approval magnitudes differ from those found in Doran, Gelber and Isen (2022), who study the 2005 and 2006 last-day-of-filling H-1B lotteries. In Appendix E, we discuss differences across these two studies—and why we do not believe they are contradictory—in more detail.

5.4.3 Robustness Checks

In Appendix Section D.1, we probe the robustness of our key results on total employment, employment of immigrant college graduates, and employment of native college graduates on several fronts. We show that these results are not sensitive to alternate control sets or alternate ways of approximating percent changes other than the inverse hyperbolic sine, as advocated in Roth and Chen (2022). We also show that our results do not depend on the additional assumptions required when using a continuous exposure variable in a difference-in-differences analysis (Callaway et al., 2024; de Chaisemartin et al., 2024) by presenting results using a binary version of our lottery success measure. In Figure A4 we show that employment magnitudes are comparable to using LBD employment where we have access to the full population of firms as opposed to only 25 states as in the LEHD. Finally, in Table A4 we present the difference in difference results for all of our outcome variables.

Table 3: Changes in Employment per Lottery Win

Full Sample	Outcome: Scaled Employment of Group			
	Total Employment	College Immigrants	College Natives	Non-College
Scaled Lottery Wins$_{j,2007}$ × Post$_t$	0.827*	0.287***	0.099	0.442
	(0.445)	(0.079)	(0.098)	(0.325)
Number of Firms	13,500	13,500	13,500	13,500
Number of Observations	137,000	137,000	137,000	137000
1st Stage F-Stat	318.5	318.5	318.5	318.5

≤ 10 Employees	Outcome: Scaled Employment of Group			
	Total Employment	College Immigrants	College Natives	Non-College
Scaled Lottery Wins$_{j,2007}$ × Post$_t$	1.006***	0.459***	0.176*	0.370**
	(0.333)	(0.125)	(0.090)	(0.150)
Number of Firms	5000	5000	5000	5000
Number of Observations	48000	48000	48000	48000
1st Stage F-Stat	249.7	249.7	249.7	249.7

(10,100) Employees	Outcome: Scaled Employment of Group			
	Total Employment	College Immigrants	College Natives	Non-College
Scaled Lottery Wins$_{j,2007}$ × Post$_t$	0.196	0.167***	0.040	-0.012
	(0.291)	(0.053)	(0.092)	(0.180)
Number of Firms	5000	5000	5000	5000
Number of Observations	52000	52000	52000	52000
1st Stage F-Stat	61.2	61.2	61.2	61.2

Note.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Firm and observation counts rounded as per the US Census Bureau's disclosure rules. Standard errors clustered at the firm level. The top panel presents the results for the full sample, the middle panel for firms that have less or equal than 10 employees in 2007 and the bottom panel for firms that have between 10 and 100 employees win 2007. "Scaled" refers to dividing by a given variable by 2006Q4 LEHD employment at the firm. Both employment counts (outcomes) and lottery wins (endogenous, independent variable) are scaled this way. All regressions include industry-time fixed effects, firm fixed effects and the IHS of employment in 2007 interacted with year dummies. All regressions are estimated using 2SLS with $Win\ Rate_{j,2007} \times Post_t$ as the instrumental variable.

5.5 Firm Performance

Having shown that lottery wins increase firm employment without generating net employment losses for substitutable natives, we next turn to the effect of lottery wins on other key indicators of firm performance. These analyses use outcomes measured in the LBD and are available for the universe of H-1B lottery applicants, subject to our sampling restrictions.

Table 4 presents our main firm performance results from Equation (3). First, Column (1) indicates that there are large extensive margin responses to lottery wins. Firms that win all of their lottery applications are 2.5 percentage points more likely to have positive payroll—our proxy for actively operating—in a given post-lottery year than firms that lose all of their lottery applications. A nontrivial set of lottery participants appear to be reliant on H-1B workers for survival.

These extensive margin responses also help explain the large scale responses found in Columns (2) and (3). Specifically, we find that revenues and payroll both rise by more than 20% at firms that win all of their lottery applications relative to firms that lose all of their lottery applications. The large magnitudes of these effects coincide with the notion that our identifying variation is driven by smaller firms who may be especially reliant on procuring an H-1B worker for survival and for generating revenue streams.

Table 4: Effect of Lottery Win Rate on Firm Productivity Measures

				Conditional on Active	
	(1)	(2)	(3)	(4)	(5)
	𝟙[Active]	Revenues	Payroll	Revenues per Worker	Average Wage
Win Rate$_j$ × Post$_t$	0.025***	0.266***	0.215***	0.019	0.012*
	(0.006)	(0.068)	(0.058)	(0.013)	(0.007)
Firms	20,000	20,000	20,000	20,000	20,000
Observations	199,000	199,000	199,000	199,000	199,000

Note. See Equation (3) for specification. Standard errors clustered at the firm level. Each outcome are measured using the LBD. Outcomes in Columns (2) through (5) are transformed using the inverse hyperbolic sine. Regressions include firm fixed effects, industry-time fixed effects, and the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects. Firm and observation counts rounded as per the US Census Bureau's disclosure rules. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Finally, in Columns (4) and (5), Table 4 provides suggestive but imprecise evidence that labor productivity increases among continuing firms after lottery wins. Both revenues per workers and average annual wages (payroll per worker) see marginally statistically significant increases in the post-lottery period among winners who continue to operate relative to losers who continue to operate. The positive result for wages might be an implication of the new H-1B immigrants having higher wages than the average worker at the firm. In Appendix Table A4, we show that the wage increases are indeed driven by college-educated immigrants. However, in Section 7, we also show that winning the H-1B lottery has both positive and negative wage impacts for certain incumbent workers.

In total, the results in Table 4 run counter to the notion that H-1B workers hired through the 2007 lottery simply provided "cheap labor" for winning firms. Increases in survival and scale coupled with *increases* in average wages suggest that firms relied on H-1B workers to increase productive capacity. We provide several additional estimates from Equation (3) that align with these conclusions in Appendix Section D.3, Table A4.

5.6 Theoretical framework

How can skilled immigrants increase firm employment and revenues, with minimal substitution of competing native workers? We adopt a simple framework whereby firms produce a particular

variety of good under monopolistic competition, using a nested-CES aggregate of different labor inputs. In this section we present a simple theoretical framework to illustrate how lottery-induced variation in skilled immigrant employment can generate scale effects that mitigate substitution between immigrants and natives. Under this design, lottery-induced variation in skilled immigrant employment generates scale effects that can mitigate substitution between immigrants and natives.[24] We assume workers consume a composite good Y, composed of different varieties indexed by z as in equation (5).

$$Y = \left(\int_{J_z} (y(z))^{\frac{\varepsilon-1}{\varepsilon}} dz \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{5}$$

where J_z represents the set of varieties available in the country, and $\varepsilon > 1$ is the elasticity of demand. Higher values of ε imply that consumers find it easier to switch consumption across varieties when prices change. Each variety is produced by a firm indexed by j that has a production function as in equation 6.

$$y_j = \psi_j L_j^{\delta} H_j^{1-\delta} \longrightarrow H_j = \left(\alpha I_j^{\frac{\sigma-1}{\sigma}} + (1-\alpha) N_j^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{6}$$

where ψ_j is the firm's specific productivity. Firms use only labor inputs of non-college educated workers L_j and college-educated workers H_j which are aggregated in a Cobb Douglas manner with δ representing the share of non-college labor in the total wage bill. We assume college workers are a CES composite of skilled immigrants and natives. The elasticity of substitution $\sigma > 1$ captures how substitutable college-educated immigrants and natives are within the firm, where higher values of σ indicate greater substitutability between immigrants and natives.

We analytically compute the change in demand for native college-educated labor following an exogenous change in the number of college-educated immigrants in the firm. In our setting, this change arises via the H-1B visa lottery, which generated an exogenous increase in the number of immigrant college workers hired at winning firms relative to losing firms. We obtain equation (7) by totally log-differentiating the first order condition of the firm optimization problem for native college labor, assuming competitive native labor markets[25] (we relegate intermediate steps to Appendix D.4):

[24] Similar frameworks have been applied to study firm-level immigration. For example, see Waugh (2017), Clemens and Lewis (2022), Mahajan (2022) and Brinatti and Morales (2023). Additionally, the contrast between scale and substitution actually dates back to the Hicks-Marshall laws of derived demand, for the two-factor case, by rearranging equation (4) in the appendix of in Hicks (1932).

[25] Recent literature has shown that immigration in monopsonistic labor markets can lead to sizable negative effects on native employment (Amior and Manning, 2020; Amior and Stuhler, 2023). This, however, is inconsistent with our finding of minimal substitution. Additionally our modelling choice is consistent with our finding of no firm-level wage impacts.

$$\frac{d\log N}{d\log I} = s_I \left[\frac{\sigma(1-\delta)(\varepsilon-1) - \varepsilon(\sigma-1)(1-s_L)}{(1-\delta)(\varepsilon-1)(1-s_L) - s_N\left(\sigma(1-\delta)(\varepsilon-1) - \varepsilon(\sigma-1)(1-s_L)\right)} \right], \tag{7}$$

where s_I, s_N, s_L are the wage bill of college-immigrants, college-natives and non-college labor as a share of total revenues, respectively. Equation (7) shows that the elasticity of native labor with respect to immigrants is a function of these shares, the non-college share parameter δ, the demand elasticity ε and the elasticity of substitution between immigrants and natives σ.

To gain intuition on the role of the elasticities (ε and σ), we simplify the model by assuming that there are no non-college workers, such that $\delta = s_L = 0$. The relative log change of natives to immigrants reduces to:

$$\left.\frac{d\log N}{d\log I}\right|_{\delta=s_L=0} = \frac{s_I(\varepsilon-\sigma)}{(1-s_N)\varepsilon + s_N\sigma - 1}. \tag{8}$$

Equation (8) implies that whether there is crowd-in or crowd-out of natives depends on the sign of $\varepsilon-\sigma$, as the denominator $(1-s_N)\varepsilon+s_N\sigma-1>0$. If $\varepsilon>\sigma$, the scale effect dominates: when immigrants arrive at the firm, the firm can expand output without having to reduce the price as much. Such expansion makes the firm also hire more natives instead of substituting them with immigrants. On the contrary, if $\sigma>\varepsilon$, the substitution effect dominates, and the influx of immigrants does not generate a big enough scale effect to prevent firms from reducing native employment in response to immigration. A similar intuition holds in the expanded model when $\delta>0$ and $s_L>0$.

Following the CES properties, firm revenues have to increase when there is an influx of immigrants. As shown in Appendix Section D.4, in the version of the model with non-college workers, the change in non-college workers is proportional to the change in revenues. Hence, an influx of college-educated immigrants unambiguously increases the employment of non-college graduates at the firm.

Equation 7 allows us to use our reduced form results to back out the implied substitution between immigrants and natives σ. First, we set ε to 2.92, which is the median value estimated by Broda and Weinstein (2006) for differentiated products. Second, we compute directly from the data the values for $s_L = 0.22$, $s_I = 0.15$, $s_N = 0.23$ and $\delta = 0.33$. Note that our sample of firms of H-1B applicants is quite immigrant-intensive and college-intensive.

Third, we use our estimates from Table A4 to estimate the value of the right-hand side of Equation (7). We focus on the relationship between H-1B like natives and H-1B like immigrants, as that is the immigrant group that is shocked by the lottery. The estimate for $\frac{d\log N}{d\log I}$ can be approximated by the relative diff-in-diff coefficients of H-1B like natives and H-1B like immigrants. Given our values for the other parameters, the estimated value of σ is 4.6. Among the literature finding imperfect substitutability, close to the value estimated by other papers

with similar frameworks (Brinatti and Morales, 2023; Burstein et al., 2020; Cortes, 2008).

6 Heterogeneity Analysis

As a next step, we investigate whether certain firm characteristics drive our results. We focus our analysis on three main outcomes: total firm employment, native college-educated employment, and immigrant college-educated employment. To estimate differential response to winning the H-1B lottery across firms, we set up a triple difference approach:

$$y_{jt} = \beta \left[\text{Win Rate}_j \times \text{Post}_t\right] + \gamma \left[\text{Post}_t \times Z_j\right] + \\ \delta \left[\text{Win Rate}_j \times \text{Post}_t \times Z_j\right] + \Gamma X_{jt} + \alpha_j + \alpha_t + \varepsilon_{jt} \ , \tag{9}$$

where Z_j stands for the standardized version of a continuous firm characteristic. The coefficient β will capture the response to winning the lottery for firms with average value of characteristic Z_j. Coefficient γ captures the differential time-trend for firms with one standard deviation above average of characteristic Z_j relative to firms with average characteristic Z_j regardless of their lottery outcome. Finally, the key coefficient of interest is δ, which captures the differential impact of winning the lottery for firms with one standard deviation above the mean in characteristic Z_j relative to firms with average Z_j. By comparing coefficient β with coefficient $\beta + \delta$, we can quantify the differential impact of winning the lottery across firms with different values of Z_j.

We explore four main characteristics in our analysis which we measure in 2006, before the lottery took place. First, we compute the number of college graduates as a share of firm employment to proxy for the skill intensity of the firm. Second, we compute the number of immigrant college graduates as a share of total college graduates, as a measure of immigrant intensity for college labor. Third, we compute average wage as a measure of overall worker quality and skill. Fourth, we compute the labor productivity at the firm by calculating the ratio of revenues to total employment.

As shown in Table 5, our four characteristics broadly present a similar picture. In terms of total employment, firms with a higher immigrant share, higher wages, or higher productivity expand by almost double when winning the lottery relative to firms with average value of these characteristics. To understand the interpretation, we can focus on the result for wages. A firm that pays an average level of wages to their employees and wins all of their lottery applications, increases their total employment by 6.4% relative to a firm that loses all of their applications. However, a firm that pays wages that are one standard deviation above the mean expand employment by an additional 10.8% when winning all of their lottery applications, relative to a

firm that loses all of their lottery applications. These results highlight the idea that high-wage and high-productivity firms benefit more from hiring H-1B immigrants, as they were likely constrained in hiring high-skill workers.

When looking at native employment, in the middle panel of Table 5, we see that firms with high immigrant shares, high wages, and high productivity expand their native-college employment when winning the lottery relative to firms with average values of these characteristics. This reinforces the idea that immigration eases a production constraint for high-wage, high-productivity firms that allows them to also hire more of other labor inputs. Finally, as shown in the bottom panel of Table 5, firms that have high values of the characteristics expand their immigrant-college employment by more than those with average values of the characteristic. This is suggestive that firms with higher productivity might recruit immigrants that are harder to find substitutes for in the domestic labor market.

7 Individual-Level Results

As a final step, we explore how the H-1B lottery affects the career trajectories of incumbent workers at lottery-participating firms. Since we are interested in understanding how incumbent workers respond when firms get a positive immigration shock, we focus on firms that participated in the lottery and have less than 100 employees.[26] We then extract the full job histories between 2001 and 2011 for all individuals who worked at those firms in 2006, on the eve of the H-1B lottery. Using this sample, we run difference-in-difference regressions at the worker-year level as shown in equation 10:

$$y_{it}^{g} = \beta^{g}[\text{Win Rate}_{j_0(i)}^{g} \times \text{Post}_{t}^{g}] + \Gamma X_{it}^{g} + \delta_{i}^{g} + \delta_{t}^{g} + \varepsilon_{it}^{g}$$

where the superscript g indicates that we run this regression separately for different groups of workers. The key coefficient of interest, β^{g}, captures the interaction between the lottery outcome at the firm the individual's employing firm in 2006 ($Win\ Rate_{j_0(i)}$) and an indicator for the post period (2007 or later). The subscript $j_0(i)$ stands for the lottery-participating firm that individual i was working for in time $t_0 = 2006$. We include the same controls as in Equation (3) based on firm $j_0(i)$ (regardless of where individual i works at time t). We also include individual fixed effects (δ_i) and gender-wage-bin-year fixed effects in X_{it}.[27] Effectively, we compare how the career paths of individuals working at firms that were successful at the lottery change relative to the career path of similar individuals working at firms that were unsuccessful at the lottery.

[26] As shown in Section 5, firms with more than 100 employees that win the lottery do not hire a statistically different number of immigrants than firms that lose the lottery following the results in Section 5.4.1, so we exclude them from this analysis.

[27] To determine the wage bins, we classify individuals into 5 wage groups based on their income in 2006.

Table 5: Triple-difference Estimates by Firm-level Characteristics

	Outcome: Q4 Employment			
Characteristic (Z_j):	College Share	Immigrant Share	Average Wage	Revenues per Worker
Win Rate$_j$ × Post$_t$	0.082**	0.078**	0.064**	0.054
	(0.036)	(0.037)	(0.036)	(0.036)
Z_j × Post$_t$	0.012	-0.001	0.053**	0.041**
	(0.021)	(0.023)	(0.024)	(0.020)
Win Rate$_j$ × Z_j × Post$_t$	0.056	0.076**	0.108**	0.069**
	(0.034)	(0.034)	(0.042)	(0.031)
	Outcome: Native College Employment			
Characteristic (Z_j):	College Share	Immigrant Share	Average Wage	Revenues per Worker
Win Rate$_j$ × Post$_t$	0.003	0.007	-0.008	-0.012
	(0.026)	(0.026)	(0.027)	(0.026)
Z_j × Post$_t$	-0.006	0.021	0.037**	0.010
	(0.014)	(0.016)	(0.017)	(0.013)
Win Rate$_j$ × Z_j × Post$_t$	0.026	0.041*	0.061**	0.051**
	(0.023)	(0.023)	(0.031)	(0.022)
	Outcome: Immigrant College Employment			
Characteristic (Z_j):	College Share	Immigrant Share	Average Wage	Revenues per Worker
Win Rate$_j$ × Post$_t$	0.094***	0.095***	0.082***	0.075***
	(0.024)	(0.025)	(0.023)	(0.024)
Z_j × Post$_t$	-0.008	-0.021	0.035**	0.032**
	(0.014)	(0.015)	(0.015)	(0.013)
Win Rate$_j$ × Z_j × Post$_t$	0.048**	0.047**	0.057*	0.036*
	(0.023)	(0.023)	(0.030)	(0.021)
Firms	13,500	13,500	13,500	13,500
Observations	137,000	137,000	137,000	137,000

Note.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Firm and observation counts rounded as per the US Census Bureau's disclosure rules. Standard errors clustered at the firm level. The top panel presents the estimates for the outcome of the IHS of total employment. The middle panel uses as an outcome the IHS of total employment of college-educated natives. The bottom panel uses as an outcome the IHS of total employment of college-educated immigrants. All outcome variables are computed using the LEHD. The characteristics Z_j and standardized and computed in year 2006.

We focus on two main outcomes: the evolution of log individual wages and the probability of worker being at their 2006 firm at time t. As we expect heterogeneous effects of the lottery across incumbent workers, we separate individuals into six groups g based on age, tenure, nativity and education. H-1B immigrants tend to be below the age of 40, are college educated, and by definition have low-tenure at the firm. Therefore, specifically, we separate workers by whether: 1) they were above or below 40 years of age in 2006, 2) were above or below three years of tenure at the lottery firm in 2006, 3) have a college degree, and 4) are immigrants

or natives. We combine all college workers above 40 years old into a single group, as well as pooling all non-college workers together.

Table 6 presents the results for these groups of workers. Interestingly, young college-educated workers with low tenure who worked at a firm that won the lottery see a wage increase between 4-5% relative to workers of that same group who worked at firms that lost the lottery. One interpretation of these results is that incumbent workers who are young and low tenure might be those who work most closely with and thereby learn the most from the incoming H-1B worker. Jarosch et al. (2021) measure learning from coworkers through the increase in wages of incumbents after being exposed to a productive new coworker. Non-college workers at lottery-winning firms also seem to benefit from the H-1B immigrant coworkers, as they increase their wages by 3.1% after the lottery relative to those in lottery-losing firms.

Table 6: Individual-Level Analysis

	Outcome		
Group	Log Wage	$\mathbb{1}$(At Lottery Firm)	Number of observations
Immigrants, college, low-tenure, age ≤ 40	0.050**	0.020	175,000
	(0.025)	(0.016)	
Natives, college, low-tenure, age ≤ 40	0.041*	-0.004	122,000
	(0.024)	(0.016)	
Immigrants, college, high-tenure, age ≤ 40	-0.015	-0.003	79,000
	(0.023)	(0.016)	
Natives, college, high-tenure, age ≤ 40	-0.054**	-0.036**	103,000
	(0.024)	(0.016)	
College workers, age > 40	0.013	-0.011	583,000
	(0.012)	(0.015)	
Non-college workers	0.031**	-0.001	1,374,000
	(0.013)	(0.011)	

Note.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Observation counts rounded as per the US Census Bureau's disclosure rules. Standard errors clustered at the firm level. The first column presents the results for the outcome of $\log(\text{wage}_{i,t})$ at the individual level. The second column uses as an outcome a binary variable that takes the value of 1 if the individual still works at their 2006 firm and 0 otherwise. The coefficients presented are the interaction between $\mathbb{1}$(Individual belongs to group g) $\times$ Win Rate$_{j_0(i)}$ $\times$ $\mathbb{1}$(year $\geq$ 2007). Groups are defined based on an individual's age and tenure in 2006. The Win Rate$_{j_0(i)}$ is the win rate of the firm where the employee worked at in 2006. We limit the sample to individuals working in 2006 at lottery-participating firms that had less or equal than 100 employees.

However, not all worker groups benefit from the exposure to an H-1B coworker. Young, college-educated natives who have high tenure experience 5.4% wage reductions when tied to a lottery-

winning firms relative to those who were tied to lottery-losing firms at the time of the lottery. Some of this wage reduction may be due to jon separation: these workers are 3.6% less likely to remain at the lottery firm after 2007. In conjunction, these results are consistent with the notion that firms demand higher-tenure, young, native college workers in response to increased availability of young, immigrant college workers. In effect, we allow the data to tell us that this group of incumbent workers is most substitutable to the new H-1B immigrants. We also note that the increased separation probability higher turnover of high tenure natives might also explain the positive effects on low tenure workers, who may be promoted once more senior workers move to other jobs.

All in all, these results reveal the importance of separately analyzing the effect of exposure to immigrants on incumbent firms and incumbent individuals. In Section 5, we found that individual firms tend to expand, survive, and thrive in response to H-1B lottery wins. Here, we find that young, low-tenure, college workers of all nativities are most likely to benefit from these firm-level impacts. Meanwhile, higher-tenure, young, native college gradates appear to pay a cost. These are important, individual-level heterogeneities in the response to lottery wins that a firm-level-only analysis would miss. In this sense, our individual-level analysis mirrors that in Dustmann et al. (2016), particularly in its ability to elucidate important heterogeneities in response to immigrant exposure that are not detectable without administrative panel data on workers.

8 Conclusion

All told, we fill an important gap in the literature and by providing several new, well-identified estimates to the empirical literature on the impact of high-skill immigration on US firms and workers. A key finding is that one additional H-1B worker increases total firm employment by 0.83 additional employees. For firms with less than 10 employees, the effect is even larger, as winning an additional i-129 application increases firm size by one employee. While lottery-losing firms manage to hire some immigrants regardless, lottery-winning firms expand in other types of complementary workers such as non-college graduates. Lottery-winning firms also enjoy increases in revenue generation and survival. High-wage, high-productivity firms expand more and even crowd in natives with college degrees when winning the lottery. Finally, the impact of winning the lottery on incumbent workers is small on average, but lottery wins generate positive wage spillovers for young, low tenure college graduates and negative wage spillovers for young, high tenure, native college graduates. Our findings have implications for immigration policy design and the broader welfare effects of immigration to advanced economies.

References

Amior, Michael and Alan Manning, "Monopsony and the wage effects of migration," Technical Report, Centre for Economic Performance, LSE 2020.

__ **and Jan Stuhler**, "Immigration, monopsony and the distribution of firm pay," 2023.

Amuedo-Dorantes, Catalina, Esther Arenas-Arroyo, Parag Mahajan, and Bernhard Schmidpeter, "Low-Wage Jobs, Foreign-Born Workers, and Firm Performance," IZA Discussion Papers 16438, Institute of Labor Economics (IZA) September 2023.

__ **, Kevin Shih, and Huanan Xu**, "International Student Enrollments and Selectivity: Evidence from the Optional Practical Training Program," *Economic Inquiry*, 2023, *61* (2), 253–281.

Arellano-Bover, Jaime and Shmuel San, "The Role of Firms and Job Mobility in the Assimilation of Immigrants: Former Soviet Union Jews in Israel 1990–2019," IZA Discussion Papers 16389, Institute of Labor Economics (IZA) August 2023.

Beerli, Andreas, Jan Ruffner, Michael Siegenthaler, and Giovanni Peri, "The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland," *American Economic Review*, 2021.

Bound, John, Breno Braga, Gaurav Khanna, and Sarah Turner, "The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System.," *Journal of Economic Perspectives*, 2021, *35* (1), 1–23.

__ **, 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*, 2018. (eds. G Hanson, W Kerr, and S Turner), University of Chicago Press.

Brinatti, Agostina and Nicolas Morales, "Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments," 2023. Working Paper.

__ **and Xing Guo**, "Third-Country Effects of U.S. Immigration Policy," *Working paper*, 2023.

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States," *Econometrica*, 2020, *88* (3), 1071–1112.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro Sant'Anna, "Difference-in-Differences with a Continuous Treatment," *Working Paper*, 2024.

Chen, Mingyu, "The Value of US College Education in Global Labor Markets: Experimental Evidence from China.," *Industrial Relations Section Working Papers*, 2019, (No. 627).

__ , "The Impact of International Students on US Colleges: Higher Education as a Service Export.," 2021.

__ **, Jessica Howell, and Jonathan Smith**, "Best and Brightest? International Students' Decision to Attend College in the US.," *Industrial Relations Section Working Papers*, 2020, (No. 640).

Choudhury, Prithwiraj and Do Yoon Kim, "The ethnic migrant inventor effect: Codification and recombination of knowledge across borders.," *Strategic Management Journal*, 2019, *40* (2), 203–229.

Clemens, Michael A., "Why Do Programmers Earn More in Houston than Hyderabad? Evidence from Randomized Processing of U.S. Visas," *American Economic Review, Papers & Proceedings*, May 2013, *103* (3), 198–202.

_ **and Ethan Lewis**, "The Effect of Low-Skill Immigration Restrictions on US Firms and Workers: Evidence from a Randomized Lottery," December 2022, (30589).

Cortes, Patricia, "The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data," *Journal of Political Economy*, 2008, *116* (3), 381–422.

Davis, Steven J., John Haltiwanger, and Scott Schuh, "Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts," *Small Business Economics*, 1996, *8* (4), 297–315.

de Chaisemartin, Clement, Xavier D'Haultfoeuille, Felix Pasquier, and Gonzalo Vazquez-Bare, "Difference-in-Differences for Continuous Treatments and Instruments with Stayers," *Working Paper*, 2024.

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*, 2021.

Doran, Kirk, Alexander Gelber, and Adam Isen, "The Effects of High-Skilled Immigration Policy on Firms: Evidence from H-1B Visa Lotteries," *Journal of Political Economy*, 2022.

Dustmann, Christian and Albrecht Glitz, "How Do Industries and Firms Respond to Changes in Local Labor Supply?," *Journal of Labor Economics*, 2015, *33* (3), 711–750.

_ **, Uta Schönberg, and Jan Stuhler**, "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment," *The Quarterly Journal of Economics*, 10 2016, *132* (1), 435–483.

Egger, Dennis, Daniel Auer, and Johannes Kunz, "Effects of Migrant Networks on Labor Market Integration, Local Firms and Employees," *Working Paper*, 2022.

Flaeen, Aaron and Nada Wasi, "Record Linkage Using Stata: Pre-processing, Linking, and Reviewing Utilities," *The Stata Journal*, 2015, *15* (3), 672–697.

Glennon, Britta, "How Do Restrictions on High-Skilled Immigration Affect Offshoring? Evidence from the H-1B Program," Working Paper 27538, National Bureau of Economic Research July 2020.

Hunt, Jennifer and Marjolaine Gauthier-Loiselle, "How Much Does Immigration Boost Innovation?," *American Economic Journal: Macroeconomics*, 2010, *2* (2), 31–56.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg, "Learning From Coworkers," *Econometrica*, 2021, *89* (2), , 647–676.

Kato, Takao and Chad Sparber, ""Quotas and Quality: The Effect of H-1B Visa Restrictions on the Pool of Prospective Undergraduate Students from Abroad.," *Review of Economics and Statistics*, 2013, *95* (1), 109–126.

Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln, "Skilled Immigration and the Employment Structures of US Firms," *Journal of Labor Economics*, 2015, *33* (S1), S147–S186.

Kerr, William R. and William F. Lincoln, "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention," *Journal of Labor Economics*, 2010, *28* (3), 473–508.

Khanna, Gaurav and Nicolas Morales, "The IT Boom and Other Unintended Consequences of Chasing the American Dream.," 2021. Working Paper.

Mahajan, Parag, "Immigration and Business Dynamics: Evidence from US Firms," 2022.

_ , "Immigration and Business Dynamics: Evidence from U.S. Firms," *Journal of the European Economic Association*, forth.

Mandelman, Federico, Mishita Mehra, and Hewei Shen, "Skilled Immigration Frictions as a Barrier for Young Firms," *Working Paper*, 2024.

Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Y. Shih, and Chad Sparber, "Coping with H-1B Shortages: Firm Performance and Mitigation Strategies," NBER Working Papers 27730, National Bureau of Economic Research, Inc August 2020.

Mitaritonna, Cristina, Gianluca Orefice, and Giovanni Peri, "Immigrants and firms' outcomes: Evidence from France," *European Economic Review*, 2017, *96*, 62 – 82.

Morales, Nicolas, "High-Skill Migration, Multinational Companies, and the Location of Economic Activity.," *Review of Economics and Statistics*, 2023.

Peri, Giovanni, Kevin Shih, and Chad Sparber, "Foreign and Native Skilled Workers: What Can We Learn from H-1B Lotteries?," Working Paper 21175, National Bureau of Economic Research May 2015.

_ , _ , and _ , "STEM Workers, H-1B Visas, and Productivity in US Cities," *Journal of Labor Economics*, 2015, *33* (S1), S225–S255.

Roth, Jonathan and Jiafeng Chen, "Log-like? Identified ATEs defined with zero-valued outcomes are (arbitrarily) scale-dependent," Technical Report 2022.

Shih, Kevin, "Do International Students Crowd-Out or Cross-Subsidize Americans in Higher Education?," *Journal of Public Economics*, 2017, *156*, 170–184.

Vilhuber, Lars, "LEHD Infrastructure S2014 files in the FSRDC," Technical Report, U.S. Census Bureau, Working Papers 18-27, 2018 2018.

Waugh, Michael E, "Firm dynamics and immigration: The case of high-skilled immigration," in "High-Skilled Migration to the United States and Its Economic Consequences," University of Chicago Press, 2017, pp. 205–238.

A Additional Details on the H-1B Program

The H-1B visa was created in 1990 to provide authorization for college-educated foreign nationals to work in specialty occupations in the United States. The program has undergone a variety of reforms primarily related to its quota, which was initially set at 65,000 per year. In 1998, Congress increased the quota to 115,000 for FYs 1999 and 2000, and to 195,000 for FYs 2001-2003. For FY 2004, the quota returned to its initial level of 65,000, but Congress authorized an additional quota of 20,000 visas for those with a master's degree or higher from a US educational institution (referred to as the "Advanced Degree Exemption" or ADE Cap). In 2000, US universities and other nonprofit entities were exempt from H-1B quotas, and in 2004, governmental entities also became exempt. Since FY 2004, the number of visas issued has exceeded the cap in each year mainly due to these exemptions.

The entire Regular Cap of 65,000 was distributed by random lottery for the first time in 2007, our year of study. A similar procedure was used to distribute visas under the FY 2009 cap, with 163,000 applications received for 85,000 cap visas between April 1 and April 8 of 2008, implying a similar win rate. However, ADE visa applications were first included in the lottery for the Regular visa cap (of 65,000). Nonselected ADE applications were then pooled into a second lottery for the ADE Cap (20,000). This substantially complicates analysis of the 2008 lottery.

As mentioned in Section 2, the H-1B program is the primary pathway to hire skilled workers from abroad. The alternative programs to hire high skill workers are significantly smaller. The L-1 and TN programs are less than 15% of the size of the H-1B (Morales, 2023), while the H-1B1 and E-3 visa usage is even smaller. Graduates from a US institution are allowed to work in the US for 1-2 years after graduation through the OPT. Over 40% of new H-1B visa winners in 2014 were F-1 visa holders, and at least 30% of students on OPT transition to an H-1B status (Chen, 2019). In 2008, the OPT program extended the period of work authorization from 12 months to 29 months for individuals graduating in selected STEM fields (Amuedo-Dorantes et al., 2023b).

B LCA Predating and Anticipation

To understand LCA predating, we first show how demand for H-1B visas changed over time. Figure 1 shows for each fiscal year, how long the H-1B cap took to distribute.[28] The vertical axis shows days-in-filing, calculated as the number of days between the start of the application filing period (usually April 1 or 2), and the final receipt date (i.e., the day that USCIS received enough petitions to hit the cap). Filing LCAs for the H-1B cap will likely occur around April, with start dates on or after October 1, 2007.

In earlier years, such as FY 2002, 2003 and 2004, the cap was either not exhausted or took nearly an entire year to distribute. Hence, in terms of filing, LCAs were filed more or less uniformly in calendar time. Demand noticeably increases for H-1B visas as the number of days-in-filing falls. Compared with FY 2004, H-1B distribution in 2005 took only 6 months. Each year after the days-in-filing falls and by FY 2007, all cap-subject H-1B visas were distributed within 2 months. In FYs 2008 and 2009, USCIS received an overwhelming number of applications on the first day of the filing period and held lotteries. The effects of the Great Recession are seen in FYs 2010-12 as demand slows down, and the days-in-filing increase once again to almost a full year. As the economy recovers from the Great Recession, the days-in-filing once again falls quickly. Each year since FY 2014, cap-subject H-1B visas have been distributed by lottery.

We then track the pattern of LCA filing in for each fiscal year and see how it changes with demand and the days-in-filing. Figure A1 shows the total number of approved LCA applications by week, for each H-1B filing

[28] Throughout this section, we use Fiscal Years (FY) instead of calendar years. For example, FY 2008 goes from April 2007 to March 2008.

season from FY 2003-2017. Note the weeks correspond to the calendar year, so for example, week 1 for the FY 2003 figure refers to January 1, 2002. A vertical line is displayed for the week in which the H-1B application season begins (i.e., the week of April 1, which usually corresponds to week 13 or 14 of the calendar year). The weeks go up until week 40, the week of October 1 and the start of the fiscal year.

In times of low demand, when the H-1B cap takes a long time to distribute, we see roughly uniform LCA filing throughout the calendar year. As days-in-filing starts to decrease, we start to see a mass of LCA applications grow starting around April 1. FY 2008 and 2009 are clear that a huge mass of LCA applications are filed and approved not only on April 1, but also in the weeks prior. As the Great Recession occurs, this LCA predating slows down. As the recovery happens and demand once again surges, predating behavior increases as the number of LCA applications filed and approved grows around April 1. Since FY 2014, it is clear that predating is now the norm, and most LCA applications appear to be filed before April 1.

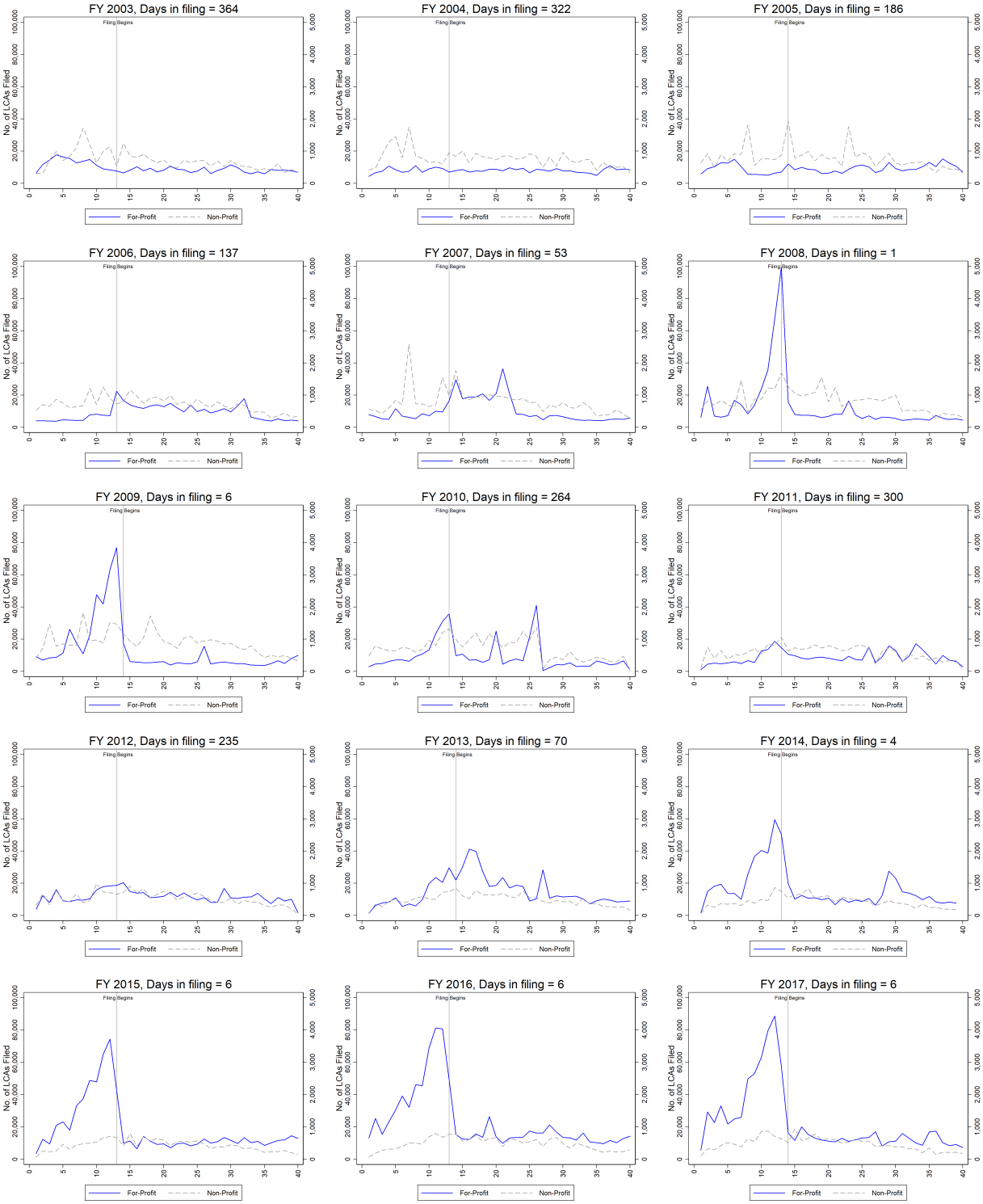
Such predating behavior presents at least two issues. First, this implies particular selection of firms into the lottery. In its simplest form, this means firms that could anticipate high demand and a lottery, ended up predating LCA applications to ensure they could submit a complete H-1B application by April 1. Those firms that either were unaware of overall demand, were unaware of the lottery protocol, or for some other reason could not match with a worker and get a completed application ready in time, did not partake. Ultimately, the firms that select into the lottery might be different from the set of firms that apply in nonlottery years. This needs to be considered when thinking about the external validity of these findings.

To get a sense of how characteristics of H-1B applicants may differ, we plot the share of applications going to computer-related occupations and to the top two H-1B receiving countries, India and China, for each fiscal year in Figure A2. Periods that had high demand, a very small number of days-in-filing, and large predating behavior, are shaded in blue. A clear pattern emerges that the share going to computer scientists and to India/China tends to rise during these periods. Hence, it is possible that firms with strong networks to computer scientists in India were those that either were able to anticipate the lotteries and/or able to quickly match with workers so that they could apply by April 1.

A second issue pertains to the internal validity of this natural experiment. Firms anticipating the lottery may be able to partially mitigate its impact. For example, a firm that anticipates a lottery and believes the odds to be equivalent to a coin flip can mitigate the lottery rationing by applying for double the amount of workers it actually needs. If the overall win-rate of the lottery is 50% and a firm needs to hire 1,000 workers, then it can get close to this target in expectation by filing 2,000 applications. However, it should be noted that this is not a trivial task. Companies need to file for LCA approval, and actually match with real workers in order to file the H-1B application. They also need to pay filing fees associated with all 2,000 applications and attorney fees where applicable. In practice, it is more likely that this mitigation strategy could reasonably be pursued by very large firms looking to hire a large number of workers, rather than smaller startup companies. In later checks, we remove the largest applicants from the analysis as a robustness check on this.

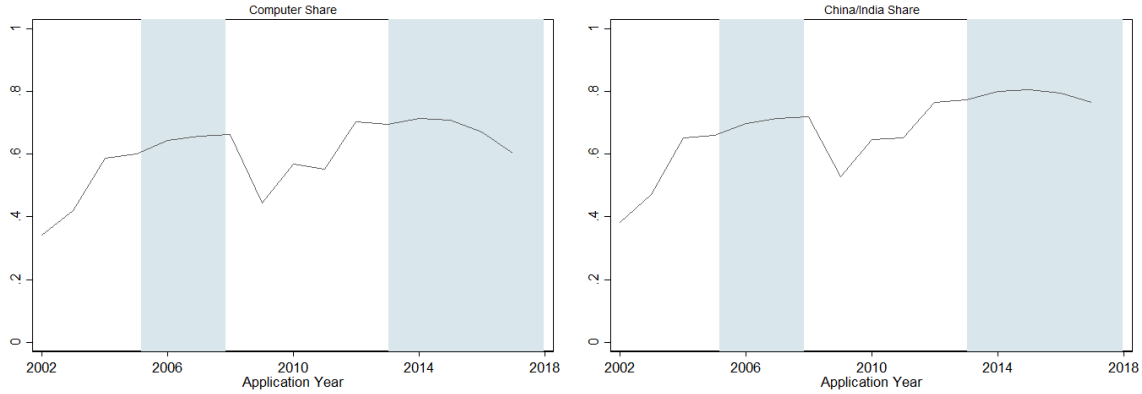
While LCAs exhibit predating, with many employment start dates listed earlier than October 1, 2007, rules stipulate that employment start dates listed on I-129 applications not be earlier than October 1. As a result, "predating" manifests in I-129 forms through an earlier than usual employment end date. Instead of a normal three-year duration (e.g., October 1, 2007-October 1, 2010), many applications list end dates that are one or two months early (e.g. October 1, 2007-September 1, 2010).

Figure A1: LCA Application Filing by Week, FY 2003-2017



Note. Figure shows LCA filing by week for each calendar year. Note that 1 refers to week 1, hence the week of January 1. Week 13/14 is marked by the vertical black line and refers to the week of April 1, the start of the H-1B filing period. Week 40 refers to the week of October 1, or the start of the subsequent fiscal year for which firms are applying for H-1B visas. We separate applications into for-profit (cap-subject) and nonprofit (cap-exempt).

Figure A2: Compositional Changes in Periods with High Predating



Note. Figure shows the share of approved I-129s awarded to computer-related occupations, and also to individuals from India and China for each fiscal year between 2002-2018. Regions shaded in blue show periods with a very small number of days-in-filing, and also an associated high amount of LCA predating activity.

C Filtering Probabilistic Outliers

We identify probabilistic outliers based on the likelihood of observing the stated number of applications given the number of wins we observe and the overall likelihood of success for any given application. To do this, we use the negative binomial distribution, which describes the number of failures (applications - wins) expected given a number of trials (applications), successes (wins), and true win probability. From publicly available numbers referred to above, we infer that the true win probability for any application should be $\frac{65{,}000}{123{,}480} = 0.5264$. We then filter out any observations whose application-win combinations are low likelihood given this probability of success—those that occur with probability less than 0.01.[29] We refer to samples that drop such outliers from the data set as "filtered." This helps ensure that the remaining set of firms have more or less "reasonable" application counts.

Figure A3 illustrates the utility of filtering out probabilistic outliers visually. Figures A3a and A3b compare the distributions of the raw win rate (without dropping probabilistic outliers) and the filtered win rate (that drops outliers), to a truly random win rate where we randomly draw 52.64% of observed applications in our data. Due to LCAs likely being an overcount of the true number of applications, the raw win rate has greater mass in the left tail of the distribution toward smaller win rates, relative to the truly random distribution. The filtered win rate distribution still has some extra mass in lower win rates but is closer to what would be expected in truly random data.

Figure A3c illustrates a local polynomial smoothed plot of win rates against applications. Ideally, in truly random data, there should be no correlation between win rates and applications. The raw win rates show a strong negative correlation, where average win rates decline as application size grows—a result of our denominator especially overstating the true number of applications for large applicants. After removing probabilistic outliers, the relationship is much closer to what one expects from truly random data: there is a much less pronounced

[29] For example, a firm that is observed to have 1 win in I-129 data with a 52.64% chance of winning any given application has a 0.005 probability of having submitted 7 or more lottery applications and a 0.011 probability of having submitted 6 or more lottery applications. We thus drop any firms with 1 estimated win and 7 or more estimated lottery applications. Similarly, the probability that a firm with 20 observed wins submitted 27 or fewer applications is 0.019, and the probability that a firm with 20 observed wins submitted 26 of fewer applications is 0.0098. So, we filter out any firms with 20 estimated wins and 26 or fewer applications. Note that this procedure embeds our confidence in observing I-129 lottery wins relative to I-129 lottery applications.

correlation between Win Rate$_j$ and Likely-Lottery Application$_j$. Furthermore, as expected under the law of large numbers under a true lottery, Win Rate$_j$ converges to 0.5264 among EINs with more applications.

The visual evidence in Figure A3c is confirmed by regression analysis presented below, in Table A1, which shows a drastic decline in the correlation between Win Rate$_j$ and Likely-Lottery Application$_j$. We address the remnant correlation with our firm fixed effects, and our firm size control, as described in Sections 5.2 and 5.1.

Table A1: Bivariate Regressions of Win Rate$_j$ on Likely-Lottery Applications$_j$

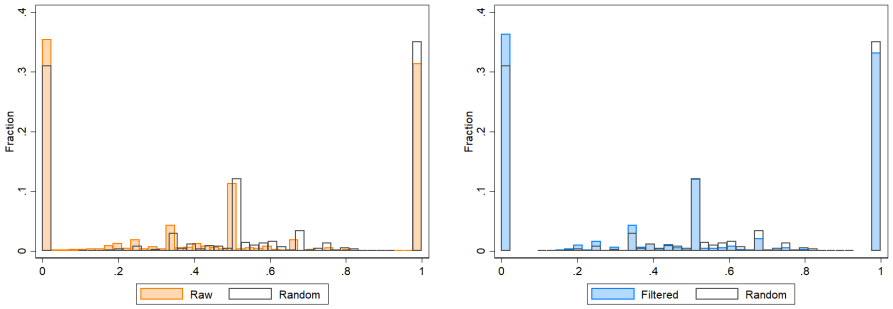
Sample:	Unfiltered (Raw)	Filtered
Win Rate$_j$	-7.946***	-0.286***
	(1.310)	(0.080)
EINs (Observations)	20,072	18,963

Note. Robust standard errors in parentheses. Both "Filtered" and "Raw" refer to our measure Win Rate$_j$, but "Filtered" refers to the measure after outliers have been eliminated by the procedure described above, in Section C. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

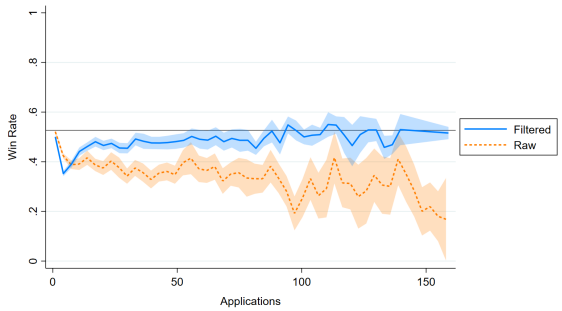
In total, our filtering method has the benefit of retaining only firms for whom we believe Win Rate$_j$ $\approx$ Ideal Win Rate$_j$ but at the cost of only analyzing 73,180 out of 123,480 total lottery applications.

Figure A3: Filtered and Raw Win Rates

(a) Raw vs. Random Win Rate (b) Filtered vs. Random Win Rate



(c) Local Polynomial Smooth Plot of Win Rates vs. Applications



Note. All figures are constructed using I-129 data, at the EIN level. Both "Filtered" and "Raw" refer to our measure Win Rate$_j$, but "Filtered" refers to the measure after outliers have been eliminated by the procedure described above, in Section C. There are 20,072 EINs represented in "Raw" win rates, and 18,963 EINs represented in the "Filtered" win rates.

D Robustness and Supplemental Results

D.1 Robustness

Here, we probe the robustness of our key firm-level scale and composition results along several dimensions. Our outcomes of interest are (percent) changes to total employment (Panel A of Table A2, the employment of college immigrants (Panel B of Table A2, and the employment of college natives at the firm level (Panel C of Table A2, and our estimating equation is Equation (3). In each panel, Column (3) is our preferred specification. Column (1) omits control vector X_{jt}. Column (2) includes the initial employment control from X_{jt} but omits industry-by-year fixed effects. Column (4) includes an additional control for the inverse hyperbolic sine of Likely-Lottery LCAs$_j$ interacted with year fixed effects. Column (5) uses a binary independent variable based on whether the win rate is more than 50%. Finally, Column (6) replaces the inverse hyperbolic sine transformation with one of the recommended replacements in Roth and Chen (2022). Specifically, we define $\text{extlog}(y_{jt}) \equiv \mathbb{1}\{y_{jt} = 0\}(-2) + \mathbb{1}\{y_{jt} > 0\}\log(y_{jt})$. That is, our outcome is now the log of our employment count, but where we replace the undefined log(0) with a value of -2. This choice implies that we consider going from 0 to 1 a growth rate of 200%. This is in keeping with the Davis et al. (1996) growth rates that are common in the firm dynamics literature.[30]

We also note that a previous draft of this paper—available upon request—used the extlog transformation throughout. Across outcomes that were used in both papers, results are unchanged qualitatively.

D.2 Likely H-1B Hiring by Origin

To further explore whether our lottery measure is capturing likely H-1B hiring, we split H-1B-like immigrants by nationality. As shown in Table A3, we separately estimate Equation (3) for Indian H-1B-like workers, Canadian/Mexican H-1B-like workers, and H-1B-like workers from all other foreign nationalities. Indians account for 48.3% of all I-129 records for new employment, while Canadians account for only 3.85% and Mexicans for 1.28%. Besides the H-1B program, college-educated Mexicans and Canadians can work in the United States under a TN employment status, which is a specific visa category established as part of the North American Free Trade Agreement (NAFTA). Therefore, we should expect the hiring of Indians and other non-North-Americans to exhibit larger responses to the H-1B lottery relative to the hiring of Canadians and Mexicans. Table A3 illustrates that this is the case and also implies some potential for substitution toward Canadian and Mexican workers for lottery losers since those workers can be pursued through alternative channels.

D.3 Additional Difference-in-Difference Results

Table A4 presents additional results from estimating Equation (3) not provided in the main text. Results support our broad conclusions that H-1B lottery wins enable firms to expand production without generating large substitution effects on native workers.

One additional note merits mentioning regarding the comparison between estimates for LBD Employment and LEHD Employment. In Figure A4, we show that our event study estimates are similar across the two data sets. Yet, Table A4 shows a substiantially lower estimate for LBD employment. The event studies help us show that this is due in part to a difference in pre-treatment dynamics, even though both are consistent with the identifying assumption of parallel trends.

[30] The Davis-Haltiwanger-Schuh growth rate in x between $t = 0$ and $t = 1$ is $\frac{2(x_1 - x_0)}{x_1 + x_0}$.

Table A2: Robustness

	Panel A: y = Total Employment					
	(1) asinh(y)	(2) asinh(y)	(3) asinh(y)	(4) asinh(y)	(5) asinh(y)	(6) extlog(y)
Win Rate$_j$ × Post$_t$	0.076**	0.085**	0.086**	0.096***		0.126***
	(0.035)	(0.035)	(0.035)	(0.035)		(0.045)
$\mathbb{1}\{\text{Win Rate}_j \geq 0.5\} \times \text{Post}_t$					0.077**	
					(0.031)	
Employment Control		✓	✓	✓	✓	✓
Industry-Year FE			✓	✓	✓	✓
LCA control				✓		
	Panel B: y = College Immigrants					
	(1) asinh(y)	(2) asinh(y)	(3) asinh(y)	(4) asinh(y)	(5) asinh(y)	(6) extlog(y)
Win Rate$_j$ × Post$_t$	0.099***	0.100***	0.093***	0.104***		0.159***
	(0.022)	(0.022)	(0.022)	(0.022)		(0.032)
$\mathbb{1}\{\text{Win Rate}_j \geq 0.5\} \times \text{Post}_t$					0.080***	
					(0.021)	
Employment Control		✓	✓	✓	✓	✓
Industry-Year FE			✓	✓	✓	✓
LCA control				✓		
	Panel C: y = College Natives					
	(1) asinh(y)	(2) asinh(y)	(3) asinh(y)	(4) asinh(y)	(5) asinh(y)	(6) extlog(y)
Win Rate$_j$ × Post$_t$	0.012	0.008	0.010	0.011		0.024
	(0.026)	(0.026)	(0.026)	(0.026)		(0.035)
$\mathbb{1}\{\text{Win Rate}_j \geq 0.5\} \times \text{Post}_t$					0.013	
					(0.022)	
Employment Control		✓	✓	✓	✓	✓
Industry-Year FE			✓	✓	✓	✓
LCA control				✓		

Note. See Equation (3) for specification, with t indexing a calendar year and omitted period $b = 2006$. Standard errors clustered at the firm level. Dependent variables are all measured using the LEHD in calendary Q4. "Employment control" corresponds to the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects. "Industry-Year FE" corresponds to 4-digit industry by year fixed effects. "LCA control" corresponds to the inverse hyperbolic sine of Likely-Lottery LCAs$_j$ interacted with year fixed effects. asinh() refers to the inverse hyperbolic sine transformation. $\text{extlog}(y) \equiv \mathbb{1}\{y = 0\}(-2) + \mathbb{1}\{y > 0\}\log(y)$. Firm and observation counts rounded as per the US Census Bureau's disclosure rules. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

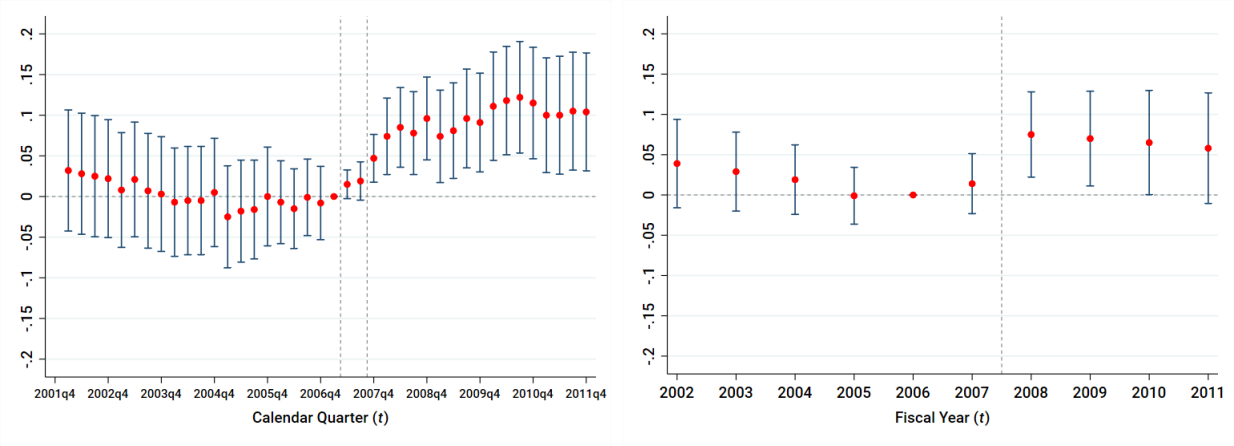
Table A3: Effect of Lottery Win Rate on Employment of H-1B-Like Workers, by Origin

	Outcome: IHS H-1B-like immigrant workers			
Origin Region:	All	India	Mexico/Canada	Other
Win Rate$_j$ $\times$ $\mathbb{1}$(Year $\geq$ 2007)	0.064***	0.047***	-0.011**	0.030**
	(0.015)	(0.009)	(0.005)	(0.013)
Firms	13,500	13,500	13,500	13,500
Observations	137,000	137,000	137,000	137,000

Note. See Equation (3) for specification, where t indexes a calendar year and $b = 2006$. Standard errors clustered at the firm level. The dependent variable is the inverse hyperbolic sine of the count of H-1B-like immigrant workers from a given origin at firm j in the fourth quarter of year t. These outcomes are measured using the LEHD. Regressions include firm fixed effects, industry-time fixed effects, and the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects. Firm and observation counts rounded as per the US Census Bureau's disclosure rules. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure A4: LBD vs. LEHD Employment as Outcome

(a) LEHD (b) LBD (March of year t)



Note. See Equation (2) for specification. In Subfigure A4a, t indexes a calendar quarter, with omitted period $b = 2007Q1$. In Subfigure A4b, t indexes a calendar year, with omitted period $b = 2006$. Each subfigure plots the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. In Subfigure A4a, first vertical dashed line separates pre-lottery quarters (placebo tests) from quarters in which lotteries may have affected behavior but before 2007 lottery H-1B workers could be hired (anticipation tests), whereas second vertical dashed line separates pre-lottery-hire quarters from post-lottery quarters (treatment effect estimates). In Subfigure A4b, the outcome is measured in the LEHD and the number of firms is roughly 13,500 (rounded as per the US Census Bureau's disclosure rules). In Subfigure A4b, vertical dashed line separates pre-lottery years (placebo tests) from post-lottery years (treatment effects). In Subfigure A4b, the outcome is measured in the LBD and number of firms is roughly 20,000 (rounded as per the US Census Bureau's disclosure rules). Standard errors clustered at the firm level in both figures. The dependent variable in each figure is the inverse hyperbolic sine of the given labeled outcome. Regressions across each subfigure include firm fixed effects, industry-time fixed effects, and log employment in March 2007 interacted with time fixed effects.

Table A4: Difference in Difference Estimates

Panel A: Firm Scale

	LBD Employment	LEHD Employment	Revenues	Noncollege Workers
Win Rate$_j$ × Post$_t$	0.050*	0.086**	0.266***	0.053*
	(0.028)	(0.035)	(0.068)	(0.031)
Firms	20,000	13,500	20,000	13,500
Observations	199,000	137,000	199,000	137,000

Panel B: Immigrant-Native Substitutability

	Immigrant College Employment	Immigrant H-1B-Like Employment	Native College Employment	Native H-1B-Like Employment
Win Rate$_j$ × Post$_t$	0.093***	0.064***	0.010	-0.015
	(0.022)	(0.015)	(0.026)	(0.017)
Firms	13,500	13,500	13,500	13,500
Observations	137,000	137,000	137,000	137,000

Panel C: Payroll and Mean Wages

	Payroll	Overall Mean Wage	Imm. College Mean Wage	Nat. College Mean Wage
Win Rate$_j$ × Post$_t$	0.215***	0.001	0.020*	0.004
	(0.058)	(0.009)	(0.012)	(0.012)
Firms	20,000	13,500	13,500	13,500
Observations	199,000	137,000	137,000	137,000

Note. See Equation (3) for specification. Standard errors clustered at the firm level. Dependent variables are all transformed using the inverse hyperbolic sine. All variables are all measured using the LEHD, except for LBD Employment, Payroll, and Revenues. Regressions include firm fixed effects, industry-time fixed effects, and the inverse hyperbolic sine of employment in March 2007 interacted with time fixed effects. Firm and observation counts rounded as per the US Census Bureau's disclosure rules. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

D.4 Theory appendix

In this section we provide additional details on the derivations of the theoretical framework in Section 5.6.

The firm's objective is to maximize profits,

$$\pi_j = PY^{\frac{1}{\varepsilon}}\psi_j^{\frac{\varepsilon-1}{\varepsilon}}L_j^{\delta\frac{\varepsilon-1}{\varepsilon}}\left(\alpha I_j^{\frac{\sigma-1}{\sigma}} + (1-\alpha)N_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}(1-\delta)\frac{\varepsilon-1}{\varepsilon}} - w_L L_j - w_I I_j - w_N N_j$$

First-order conditions are:

$$\frac{\varepsilon-1}{\varepsilon}\delta PY^{\frac{1}{\varepsilon}}\psi_j^{\frac{\varepsilon-1}{\varepsilon}}L_j^{\delta\frac{\varepsilon-1}{\varepsilon}-1}H_j^{(1-\delta)\frac{\varepsilon-1}{\varepsilon}} = w_L \tag{10}$$

$$\frac{\sigma}{\sigma-1}\frac{\varepsilon-1}{\varepsilon}(1-\delta)PY^{\frac{1}{\varepsilon}}\psi_j^{\frac{\varepsilon-1}{\varepsilon}}L_j^{\delta\frac{\varepsilon-1}{\varepsilon}}H_j^{(1-\delta)\frac{\varepsilon-1}{\varepsilon}-1}H_j^{\frac{1}{\sigma}}(1-\alpha)N_j^{-\frac{1}{\sigma}} = w_N \tag{11}$$

$$\frac{\sigma}{\sigma-1}\frac{\varepsilon-1}{\varepsilon}(1-\delta)PY^{\frac{1}{\varepsilon}}\psi_j^{\frac{\varepsilon-1}{\varepsilon}}L_j^{\delta\frac{\varepsilon-1}{\varepsilon}}H_j^{(1-\delta)\frac{\varepsilon-1}{\varepsilon}-1}H_j^{\frac{1}{\sigma}}\alpha I_j^{-\frac{1}{\sigma}} = w_I \tag{12}$$

Totally differentiating the first-order condition for skilled natives, and setting $\text{dlog}w_N = 0$ yields:

$$\text{dlogN} = s_I\left[\frac{\sigma(1-\delta)(\varepsilon-1)-\varepsilon(\sigma-1)(1-s_L)}{(1-\delta)(\varepsilon-1)(1-s_L)-s_N\left(\sigma(1-\delta)(\varepsilon-1)-\varepsilon(\sigma-1)(1-s_L)\right)}\right]\text{dlogI} \tag{13}$$

, where s_x represents the share of factor x in revenue.

While less straightforward than the simple two-factor case, the impact on native employment still depends on the relationship between the output demand elasticity and the elasticity of substitution. The numerator is less than zero when σ is sufficiently larger than ε: $\frac{\frac{\varepsilon-1}{\varepsilon}}{\frac{\sigma-1}{\sigma}} < \frac{1-s_L}{1-\delta}$. The denominator is difficult to sign, but is positive for admissible ranges of parameter values.

We can similarly derive how less-skilled workers change in response to exogenous, lottery-induced reductions in skilled immigrants. Interestingly, the model predicts that both revenues and less-skilled workers will exhibit the same proportional response to skilled immigration:

$$\text{dlogR} = \text{dlogL} = \left(\frac{s_I}{1-s_L} + \frac{s_N}{1-s_L}\left[s_I\frac{\sigma(1-\delta)(\varepsilon-1)-\varepsilon(\sigma-1)(1-s_L)}{(1-\delta)(\varepsilon-1)(1-s_L)-s_N\left(\sigma(1-\delta)(\varepsilon-1)-\varepsilon(\sigma-1)(1-s_L)\right)}\right]\right)\text{dlnI} \tag{14}$$

, where the term in brackets is the response of native skilled workers (dlogN/dlogI). Revenues must weakly be larger due to an exogenous increase in skilled immigrants. Hence, non-college workers should also rise in response.

E Comparison with Prior Literature

Our results provide new evidence on the impact of immigration on firm-level outcomes. To put our findings in context, we next compare them with the findings of Doran, Gelber and Isen (2022), henceforth DGI, who also use H-1B lottery variation and study the impact of winning the lottery on total firm employment.

Using IRS data on total employment and profits, DGI focus predominantly on the H-1B lotteries that occurred in 2005 and 2006, where only the subset of firms that applied for H-1B employment on the day that the cap filled were included. For these smaller lotteries, they have the ideal data since they observe both actual applications and lottery winners. They find that one additional H-1B win reduces the total employment at the

firm by 0.5, while in our results, we find that one additional H-1B win increases total firm employment by 0.83 workers.

In our view, the main reason that our results diverge from DGI likely owes to a difference in samples across the two papers. Specifically, firms that applied for visas on the last day of filing for 2005 and 2006 (used by DGI) likely differed in their need for H-1B workers than the full sample of H-1B applicants that applied in 2007 (as in our paper). We arrive at this conclusion by using our FOIA USCIS data to closely match the DGI 2005 and 2006 sample, as seen in Table A5. We then use this data to highlight some key differences across the main sample used in DGI and the main sample used here.

Table A5: Doran, Gelber and Isen (2022) Sample

	As reported in Doran, Gelber and Isen (2022)	Author's Calculations
Total CY 2005-2006 (FY2006-2007) Lottery EINs	2,750	2,840
EINs participating in lottery in both years	300	311
2005 Regular Lottery Win Rate	0.04	0.06
2005 ADE Lottery Win Rate	0.17	0.16
2006 Regular Lottery Win Rate	0.98	0.94
2006 ADE Lottery Win Rate	0.55	0.58

Note. Left column: taken from text of Doran, Gelber and Isen (2022). Right column: based on authors' calculations of USCIS FOIA data.

A key difference between the 2005 and 2006 lotteries (DGI) and the 2007 lottery (here) is that firms in 2005 and 2006 had several days to file applications for H-1B workers that were not subject to a lottery. In Table A6, we show that the majority of firms in the 2005 and 2006 lotteries had done so, with success. For example, in the 2005 regular lottery, firms that were subject to the last-filing-date lottery had already procured a mean of 18.8 H-1B workers and a median of 2 H-1B workers between April 1, 2005 and August 9, 2005 (the last day of filing was August 10, 2005). 63% of firms participating in the 2005 regular lottery had already succeeded in getting a USCIS approval for and H-1B worker before the last day of filing. Numbers are similar for the 2006 regular and 2006 ADE lottery and more modest for the small 2005 ADE lottery. These numbers are zero in the 2007 regular lottery that we study in this paper as all applications were subject to the lottery.

We therefore believe it is fair to posit that the effects we find in this paper combine the intensive and extensive margin—the ability for a firm to get *any* H-1B workers with the ability for a firm to meet its exact demand for H-1B workers—whereas DGI primarily identify intensive margin effects—the effect of getting the marginal H-1B worker into a firm. In our view, *both* of these effects are of substantial policy interest. Policies that allow heavy users of the H-1B program to further increase usage may be more likely to feature crowd-out, as found in DGI, whereas policies that expand the availability of H-1B workers to a wider swath of firms may be more likely to feature scale expansion and limited crowd-out, as found here.

Moving beyond comparisons *across* lotteries, DGI do also present an estimate of crowd-out from the 2007 lottery as a robustness check. They also proxy the number of applications with the number of LCAs and find that additional H-1B win increases firm employment by just 0.36 employees (crowd out of 0.64).[31] While this estimate is lower than our estimated crowd-out estimate of 0.17, we note that our estimates are contained within their confidence intervals.

We also want to emphasize two key improvements to the DGI approach *within* the study of the 2007 lottery.

[31] Using the sample of all firms, they estimate that one H-1B win generates a decrease in employment of *other* workers by 0.64 employees, after subtracting one to total firm employment. This is consistent with an increase in total firm employment of 1-0.64=0.36 workers.

Table A6: Cap-Subject H-1B Approvals Prior to Lottery

		Mean	Median	Prop. ≥ 1	Mean Filled Demand	Number of EINs
Doran, Gelber and Isen (2022)	2005 Regular Lottery	18.8	2	0.63	0.37	1,148
	2005 ADE Lottery	2.31	0	0.39	0.20	244
	2006 Regular Lottery	15.7	1	0.59	0.32	1,612
	2006 ADE Lottery	31.3	2	0.64	0.39	232
This paper	2007 Regular Lottery	0	0	0	0	$\approx 18,963$

Note. Calculations based on our identification of the Doran, Gelber and Isen (2022) sample in our USCIS FOIA data. See Table A5. "Filled demand" is defined as the number of cap-subject H-1B approvals prior to the lottery divded by the sum of the number of cap-subject H-1B applications prior to the lottery and the number of cap-subject H-1B applications in the lottery.

First, as discussed in Section 4, we improve the measurement of the lottery by removing the applications that were part of the ADE Cap, applications that were likely submitted for purposes other than new employment, and firms that featured extreme outlier application counts. Second, as discussed in Section 5.2, when using LCAs to proxy for applications, we believe it is prudent to go beyond cross-sectional regressions and implement a difference-in-difference approach that controls for firm, industry-time, and size-time fixed effects in order to isolate the random component of the lottery. Finally, we believe our filtering strategy helps remove bias from our estimates.

In sum, we believe this paper provides an important complement to the well-identified estimates contained in DGI regarding the last-day-of-filing lotteries conducted in 2005 and 2006. Our lottery affected a larger number of firms, and losses in our lottery froze more firms out of the H-1B procurement process altogether. In our view, it is therefore natural that the two papers would find different crowd-out estimates. The "right" estimate of crowd-out owing to H-1B hiring will therefore depend on the context and policy in question.