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 2023

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 productivity. Skill-intensive, high-paying firms expand the most after winning the H-1B lottery. We find limited evidence of displacement effects on native-born, college-educated workers.

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, Bank of Canada, Wharton Migration and Organizations Conference, SEA Annual Meetings (Ft. Lauderdale), and NBER Immigrants and the US Economy Conference. Any views expressed are those of the authors and not those of the US 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-FY23-P1582-R10450). 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. 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. 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, 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, into the sectors and locations that drive US innovation. 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 evaluate how lottery wins impact employment, worker composition, revenues, and labor productivity.

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 went through a lottery to determine which visa applicants were allowed to work in the United States. We combine a firm-level data set on granted H-1B visas obtained through a Freedom of Information Act (FOIA) request from the United States Citizenship and Immigration Services (USCIS) with information on intended applications for H-1B visas submitted by firms to the Department of Labor—Labor Conditions Applications (LCAs). Using these two sources, we construct our firm-level, lottery-induced shock to immigrant hiring: the number of visas the firm won 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 using both direct matching through Employer Identification Number (EIN) and a fuzzy string match procedure. The economic data primarily consists of the Longitudinal Employer-Household Survey (LEHD), a data set on the universe of employer-employee matches of twenty-five US states between 2002 and 2011, which includes wages, employment histories, and workers' country of birth, among other variables. We complement the LEHD data with firm-level revenue and employment information in the Longitudinal Business Database (LBD), a panel of nearly all non-farm, private sector US firms that can be linked to the LEHD. 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 on firm performance and hiring. This empirical approach allows us to evaluate the plausibility that our lottery measure is exogenous and examine its impact on firm-level outcomes. As a first step, we corroborate there is a "first-stage" by showing that lottery winners employ more young, low-tenure immigrant workers with college degrees after 2007. Importantly, we also provide evidence that our lottery measure is uncorrelated with pre-period changes in other firm outcomes, and with firm-level shocks around the time of the H-1B application.

The increases in likely H-1B immigrant employment lead to increases in total employment, revenues, and revenues per worker. We find that each additional H-1B worker is associated with an increase in total firm employment of one by 2009 and an increase of 0.91 by 2011. These numbers do not point to either crowd out or crowd in of overall employment at the firm in response to lottery wins. Interestingly, we also find a response in the extensive margin, where firms that win all of their lottery applications are 2.4% more likely to remain active than firms that lose all of their applications.

We find limited evidence for impacts of lottery wins on a firm's employment of native workers. We estimate that each lottery win crowds out at most 0.1 native workers with similar experience and education to the H-1B immigrant workers. Meanwhile, we see no evidence of crowd out of native workers with college degrees more generally. Firms also hire marginally more noncollege workers in response to lottery wins. All told, our main results indicate that lottery wins enable firms to scale up without generating large amounts of substitution.

Finally, we explore which firms respond the most to the availability of immigrants. Firms that are skill-intensive, immigrant-intensive, and pay high wages are more likely to expand in terms of employment, and even crowd in native workers after winning the lottery. Low-paying firms, on the other hand, seem to drive the crowd out of native workers with characteristics similar to the H-1B immigrants.

These firm-level responses do not necessarily imply that incumbent employees' career paths are unaltered by lottery results. In ongoing analysis, we use an individual-level event study analysis to evaluate how incumbent worker career paths change in response to firm-level lottery wins, relative to their peers at lottery losers. In particular, we study changes in earnings trajectories and how likely a worker is to switch jobs and industry, among other outcomes. Results from these analyses have not undergone disclosure avoidance review and are thus in preparation for future drafts.

In its current form, our work makes four main contributions. First, we are among the first papers to use the rich, firm-worker 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. Finally, we provide new evidence on how high-skilled immigration restrictions impact firm performance, a key question for policy-making.

A nascent, growing literature studies the impact of immigrant workers on firm performance using employer-level data (see, e.g., Brinatti and Morales, 2021; Clemens and Lewis, 2022; Doran et al., forthcoming; Kerr et al., 2015; Mahajan, 2022; 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.

The most closely related work is Doran et al. (forthcoming), 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 et al. (forthcoming) by analyzing a more representative sample of H-1B employers.[1,2]

Furthermore, the LEHD provides 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. The primary drawback of our work relative to Doran et al. (forthcoming) is that we do not observe applications that lost the lottery. To get around this issue, we construct and validate a proxy measure for firm-level lottery success, as described in Section 4.

Other papers have focused on the impact of the H-1B program on alternative outcomes such as offshoring (Glennon, 2020), multinational activity (Morales, 2021), entrepreneurship (Dimmock et al., 2021), international human capital (Khanna and Morales, 2021), innovation (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).[3] Of

[1]Our firm count has been rounded to the nearest 100 as per the US Census Bureau's disclosure rules.

[2]Doran et al. (forthcoming) 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 8.

[3]The H-1B program also has the potential to impact higher education (Kato and Sparber, 2013). Literature

particular note is Peri et al. (2015a), who also study the 2007 and 2008 lotteries and find evidence of complementarities between H-1B and native workers in computer-related occupations at the city level. 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 individual worker career paths in response to H-1B worker inflows. Clemens (2013) also looks at the H-1B lotteries—but using within firm variation for an Indian multinational—and shows that Indian workers selected in the lottery get sixfold wage increase with respect to comparable workers that lose the lottery. While not specific to the H-1B, Clemens and Lewis (2022) look into the H-2B program, aimed predominantly at hiring temporary, noncollege workers in the service sector. Using the 2020 and 2021 lotteries, their results align well with ours: lottery-winning firms expand more in terms of revenues and with no impact on native employment.

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. The program has undergone a variety of reforms primarily related to its quota, which was initially set at 65,000 per year.[4] Since fiscal year (FY) 2004, quotas 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), have only applied to new employment at for-profit employers, and have been binding in each year since FY 2004.[5]

H-1B visas are allocated on a first-come first-served basis. In 2005, procedures were updated (see *70 FR 23775, May 2005*) 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 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.

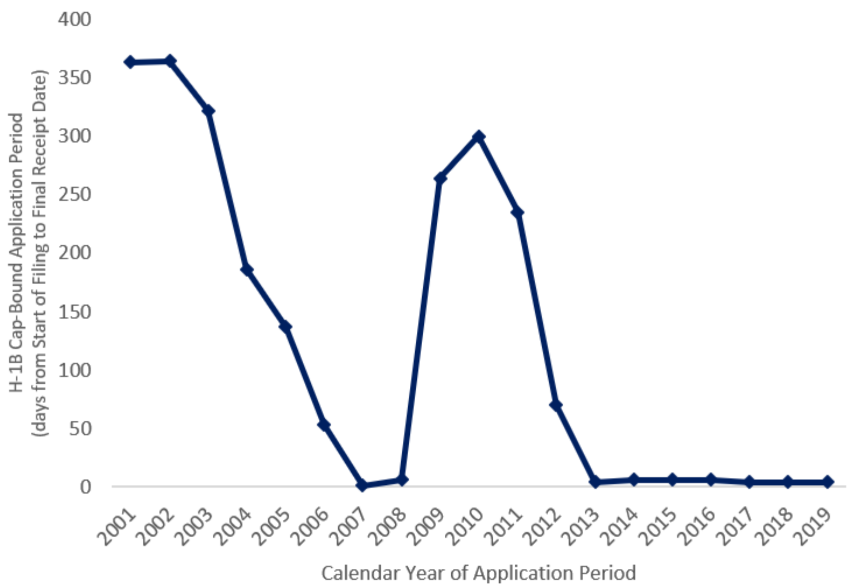
Figure 1 displays the number of days to reach the H-1B cap from the start of the filing period for each application season, denoted by calendar years 2001-2019. From 2003 application period onward, the cap was reached in successively fewer days from the start of filing. The 2005 and 2006 application periods marked the first time USCIS held small lotteries on the final receipt

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).

[4]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).

[5]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.

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.

dates, and were comprised of just under 3,000 firms (Doran et al., forthcoming).[6]

The entire Regular cap of 65,000 was distributed by random lottery for the first time in 2007. By April 2, 2007, the first day of the filing period, the number of applications 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. Lotteries were also held in 2008[7], after which the Great Recession temporarily reduced demand for H-1B workers from 2009-2012. Since 2013, lotteries have been held in each year to distribute the H-1B cap.

Finally, firms with less lottery success had alternatives to the H-1B visa, however these often had shorter work duration or restrictions to specific countries. The Optional Practical Training program (OPT) provided only one to two years of work duration for international students in the years we study.[8] The L-1 visa was limited to multinational firms transferring workers

[6]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.

[7]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.

[8]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

to US offices. Lastly, alternative visas were available for specific nations, but had very low levels of utilization.[9] The H-1B program remains the primary way to hire skilled workers from abroad.

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[10], 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.[11]

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 show up at firms.

---

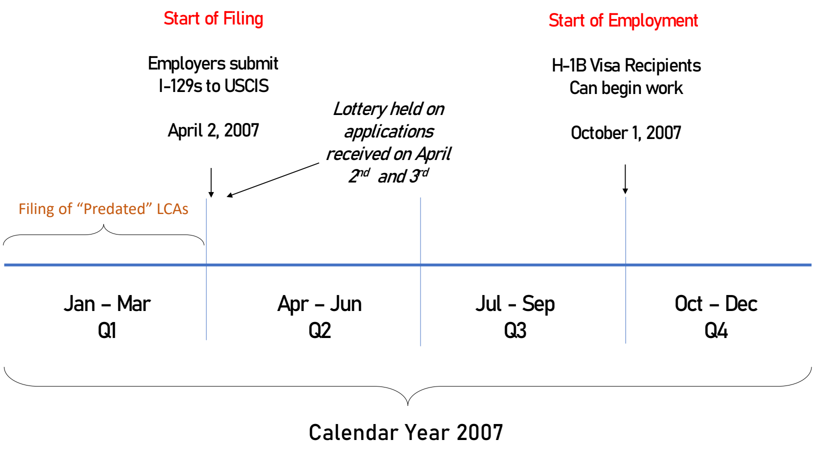
from 12 months to 29 months for individuals graduating in selected STEM fields (Amuedo-Dorantes et al., 2023).

[9] These include the TN visa for Mexicans and Canadians, the E-3 visa for Australians, and the H-1B1 visa for Chileans and Singaporeans. The L-1 and TN programs are less than 15% of the size of the H-1B (Morales, 2021), while the H-1B1 and E-3 visa usage is even smaller.

[10] They must attest that incumbent workers have been informed of their intention to hire H-1B workers, that H-1B workers will be afforded the same benefits and working conditions as other workers in the firm, and that H-1B workers will not be paid below market wages.

[11] 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.). Filing fees range from $2,000 to $10,000, not including attorney fees, which can be significant for those employers needing to retain external legal assistance with the application process.

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

Our main data set consists of 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 the following twenty-five US 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. During our sample period of 2002 to 2011, 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.[12]

The LEHD data is primarily collected from state agencies, then complemented with other administrative sources, surveys, and censuses by the US Census Bureau. The 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

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

details on the LEHD can be found in Vilhuber (2018).

We complement LEHD data by linking with the Longitudinal Business Database (LBD), which consists of the universe of private sector establishments in the United 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 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.

A final note pertains to the timing of when LEHD and LBD data are measured. The LBD provides variables primarily an annual frequency, with employment measured in March of each year. This has implications for partitioning observations pre- and post-lottery. The FY 2008 lottery occurs in April 2007 (Q2 of calendar year 2007). Hence, LBD variables (e.g., employment) measured in 2007 reflect pre-lottery outcomes (i.e., in March 2007) while 2008 and after reflect post-lottery outcomes. The LEHD variables are measured quarterly, and we use Q4 measures when examining LEHD outcomes at an annual frequency, so that Q4 2007 would be the first post-lottery measurement of outcomes.

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 2001-2018. 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 identification number (EIN) that is common in both data sets. Records remaining unmatched 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. The resulting data set has a high rate of matches.[13]

4 Measuring H-1B Visa Lottery Success

Our variation in a firm's ability to hire highly educated immigrant labor comes from the 2007 H-1B lottery. An ideal measure of lottery success would divide the number of successful H-1B applications for a given firm by the total number of H-1B applications for that firm:

$$\text{Ideal Win Rate}_j \equiv \frac{\text{Lottery Wins}_j}{\text{Lottery Applications}_j},$$

where Lottery Wins_j is the number of I-129 petitions (i.e., H-1B applications) filed by firm j that were successful in year t's lottery, and $\text{Lottery Applications}_j$ is the total number of I-129 petitions filed by firm j that were entered into the lottery. Ideal Win Rate_j is then the 2007 lottery win rate for firm j.

There are two limitations in the data that prevent us from computing the ideal measure. First, I-129 administrative records do not contain a variable that exactly identifies lottery winning petitions. Second, while USCIS processed lottery winning applications, losing applications were returned to senders without further processing. This means that our USCIS I-129 administrative records only contain data on lottery winners—data on lottery losers was never entered into record.

Despite not having exact identifiers for lottery winners, our administrative I-129 records contain

[13] The exact match figures have not yet received disclosure approval from the Census.

granular application information that allows us to identify lottery winners with high confidence. We classify winning lottery petitions as those that were received on April 2 and 3 of 2007, that are from nonuniversity, private sector firms, that are for individuals without a master's degree or higher[14], and that requested an employment start date not earlier than October 1, 2007—the first day that winning H-1B workers can legally commence employment.[15] These restrictions result in 64,723 winning I-129 petitions, remarkably close to the actual cap figure of 65,000. Hence, we believe I-129 data provides a rather accurate measurement of the numerator, which we call Likely-Lottery Wins_j.

I-129 data is less useful in measuring Lottery Applications_j, as they do not contain the applications of lottery losers. Instead, we take advantage of data on LCAs to construct a proxy for total applications, which we call Likely-Lottery Applications_j. Our strategy improves upon prior research (Dimmock et al., 2021; Peri et al., 2015a), which uses a simple count of LCAs filed as a proxy. This can overestimate the number of lottery I-129 petitions for two reasons: 1) LCAs are also filed for H-1B renewals and other changes of status; 2) unlike I-129 petitions, LCAs are costless to file (in monetary terms), so firms may file LCAs without a strong commitment to actually submitting an H-1B application.

We make several refinements to reduce measurement error and more accurately target LCAs that were related to true H-1B lottery applications. First, we retain only LCAs filed between March 1 and April 3 that maintain a six-month duration between LCA submission date and employment start date (i.e., employment start dates between August 1 and October 3).[16] Next, for each firm we also calculate the total *nonlottery* I-129 filings (e.g., renewals, change of status, etc.) between March 1 and April 3 and then net this out of the LCA count. Hence, our proxy removes LCAs that were filed for nonlottery H-1B applications to better proxy for the true number of lottery applications.[17]

[14]Note that this restriction is to exclude ADE applications as they did not participate in the Regular-cap lottery in 2007.

[15]We note that 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).

[16]The six-month restriction, as described in Appendix Section A, is to better capture predated LCAs that eventually became H-1B applications. Because firms may not submit LCAs more than six months prior to the employment start date, firms began putting employment start dates prior to October 1. Hence, for example, an LCA filed on March 1 with an employment start date of September 1 (i.e., exactly six months later), is more likely to be related to an H-1B lottery application. An LCA filed on March 1 with an employment start date of April 1, is likely related to nonlottery H-1B applications for renewals or change of status.

[17]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 $\text{Applications}_j \leq 0$ after this step.

Putting these two pieces together results in our win rate proxy:

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

We then take one final step to address remnant mismeasurement of Win Rate$_j$. We identify and remove observations whose application counts Likely-Lottery Applications$_j$ are improbably small or large given the number of Likely-Lottery Wins$_j$ that we observe. This "filtering" procedure takes advantage of two key premises: 1) the true win probability for any application should be $\frac{65{,}000}{123{,}480} = 0.5264$, given public reports from USCIS; 2) we are confident that Likely-Lottery Wins$_j$ approximates Lottery Wins$_j$ well. In this situation, the number of losing applications for a given firm should follow a negative binomial distribution with success probability 0.5264 and Likely-Lottery Wins$_j$ successes. Further details of our filtering strategy can be seen in Appendix Section B.

Table 1 summarizes our measurement of Win Rate$_j$. To start, several key points emerge from Column (1), which is tabulated from our *unfiltered* sample of linked I-129-LCA data, wherein each observation is a unique EIN.[18] Perhaps most importantly, despite our initial restrictions, our proxy for applications (Likely-Lottery Applications$_j$) still overestimates the true number of lottery applications reported by USCIS without filtering (182,951 vs 123,480) in Column (1), resulting in an estimated overall fraction of winning applications of 30% instead of 53%. We also note that the average win rate calculated across firms (46%) is much larger than this implied overall fraction of winners (30%) in Column (1). This introduces the potential for nonrandom measurement error in our win rates, correlated with size.

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 we are more likely capturing lottery variation in the filtered sample. Column (3) similarly displays disclosed statistics from our primary analysis sample after linking to census data and filtering but where the unit of analysis is the firm (firmid).[19] Compared to Column (1), average application size is also smaller, and the average win rate in this filtered sample (41%) is once again nearly identical to the overall

[18]In order to avoid excessive disclosure avoidance review at this juncture, Columns (1) and (2) of Table 1 are constructed at the Employer Identification Number (EIN) level using only the linked I129-LCA data. Column (3) contains disclosed statistics from the analysis data that is linked to the US Census and is at the "firmid" level. The firmid is a variable coded by the Census that identifies firms and is at a higher level of aggregation than EIN.

[19]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: Likely-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 Applications$_j = 1$	0.52	0.55	–
Prop. EINs with Wins$_j = 0$	0.35	0.36	–
Prop. EINs with Wins$_j = 1$	0.44	0.46	–
Prop. EINs with 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 B. 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.

share of wins (32,170/73,180≈40%).[20]

In Panel D, we also take note of the skewness in both wins and applications, which helps contextualize the results that follow in Section 6. 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 EIN entities only file one application, which contributes to large shares of EINs either winning zero H-1B visas or exactly one H-1B visas. The share winning more than one is less than 20%. Given that our regressions are at the firm-year level and not weighted by firm size, much of our identifying variation comes from these firms that apply for only one H-1B worker.

[20]We are unable to report EIN level statistics in Panel D for Column (3) since the unit of analysis is at the firmid level, and many of these have not yet received disclosure approval.

5 Identification and Validity

The goal of our study is to quantify how access to college-educated, immigrant workers affects firm performance and employment composition. Estimating this effect is challenging, as firms that hire immigrants are likely subject to different unobservable productivity shocks than firms that do not. The H-1B lottery provides a unique source of random variation that would allow us, in an ideal scenario, to compare similar firms that had differential success in their hiring of college-educated foreign-born works.

However, as described in Section 4, our win rate proxy is subject to specific concerns that we need to consider when setting up our empirical specification. First, since applying for LCAs is not costly, it is possible that certain firms apply for more LCAs than the actual number of I-129 applications they intend to submit. Similarly, firms might receive a negative shock between the time they submit their LCA and the time they need to submit the I-129 application, even though the time window between the LCA and the I-129 application is generally no more than a few months. In both of these cases, we could mistakenly assign a low win rate to a firm that never participated in the lottery. In the latter case, we would be biased toward finding a spurious impact of our lottery measure on firm performance.

While our various corrections to win rates help reduce the scope of measurement error, they do not remove it completely. We now turn to the implications of mismeasured win rates for empirical identification of the causal impacts of skilled immigration on worker and firm outcomes.

There are two primary concerns with measurement error in win rates that stem primarily from overstating applications for many firms. Consider a simple latent variable decomposition, where measured applications are comprised of true (unobserved applications) and additional "erroneous" applications, that essentially are excess LCAs that never become true H-1B applications:

$$\text{Likely-Lottery Applications}_{jt} = \text{Lottery Applications}_{jt} + \text{Excess LCAs}_{jt}.$$

Further, assume excess LCAs by firm j in any period t, are comprised of a firm specific component (δ_j), time specific factor (δ_t), a firm-time specific component (δ_{jt}), and a random error (ϵ_{jt}):

$$\text{Excess LCAs}_{jt} = \eta_j + \eta_t + \eta_{jt} + \epsilon_{jt}.$$

This decomposition helps illustrate potential threats, which our identification strategy must be explicit in addressing. If all excess LCA filing was purely due to random error (ϵ_{jt}), then measurement error would be of the classical form and lead to attenuation bias. It is much more concerning if the other components of excess LCA filing are relevant.

The firm specific component (η_j) captures fixed features of firms that lead it to always file some amount of excess LCAs. For example, large firms with in-house legal teams may always file LCAs in excess of their true anticipated hiring needs, since they are costless to file. This behavior is time-invariant and would be expected in every filing season. The time specific component (η_t) refers to aggregate shocks that affect all firms applying to the H-1B visa. For example, in economic expansions, all firms may file many excess LCAs, whereas in contractionary ones, firms may be more cautious. This reasoning helps motivate our empirical strategy, which leverages our panel data to net out both firm- and time-specific components in a difference-in-difference approach.

However, remaining firm-time specific shocks embodied by η_{jt} remain a threat to identification. For example, a firm that applies for an LCA with the intention to hire might subsequently suffer a negative labor demand shock which lowers their probability of following through with an H-1B application. This would clearly induce a spurious correlation between win rates and outcomes, as firms with these negative labor demand shocks would have more excess LCAs and thus larger measured applications (relative to the true number). This in turn reduces firm's measured win rate from its true win rate in the lottery. Therefore, firms with lower win rates will present with negative outcomes, solely due to negative labor demand shocks occurring before the H-1B lottery.

To alleviate concerns about such time-varying, firm-specific shocks, we provide two types of empirical tests. First, we perform event-study analyses that compare outcomes of more and less successful firms over time. These dynamic analyses will allow us to test for any differential trends in outcomes in the years leading up to the FY 2008 lottery. To preview event study findings, which are presented alongside results in Section 6, we find no significant pre-trends across the full range of our outcome variables in the years leading up to the 2007 lottery.

Second, we perform more targeted balance checks by regressing pre-lottery changes in outcomes on win rates. These are targeted since we can 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.

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. In support of our identification strategy, there does not appear to be any correlation between these outcomes and win rates, thereby increasing our confidence that endogenous firm

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.

shocks occurring just before the lottery are unlikely to be a significant source of bias.

The aforementioned tests provide compelling substantiation for using our measured win rates as an exogenous source of variation in firm success in the lottery. In Appendix B, we provide even further tests that show our win rate is similar to what one would expect from truly random data.

6 Empirical Analysis

6.1 Difference-in-Difference Approach

Given the results from Section 5, we estimate continuous difference-in-difference event study models specified by Equation (2):

$$\log(y_{jt}) = \sum_{\tau \neq 2006} \beta_\tau \left[\text{Win Rate}_j \times \mathbb{1}(\tau = t)\right] + \Gamma X_{jt} + \alpha_j + \alpha_t + \varepsilon_{jt}, \tag{2}$$

where j denotes the firm, and t denotes a calendar year. Logged outcomes $\log(y_{jt})$ include firm-level variables such as revenues, total employment, and employment of specific subgroups. Controls X_{jt} include a four-digit industry-by-time fixed effect and a firm's log employment in March 2007 (pre-lottery) interacted with time fixed effects. Crucially, α_j plays the role that differencing played in our balance tests.

This approach helps us assess concerns regarding the win rate proxy Win Rate$_j$. First, we can use estimated $\hat{\beta}_\tau$ for $\tau < 2007$ to evaluate whether firms with different win rates were following similar trends before the lottery. Second, the inclusion of X_{jt} allows us to compare firms that are similar along observable characteristics such as size and industry. The size control is particularly important because larger firms are more likely to engage in "over-applying" for LCAs and therefore more likely to have artificially low win rates.

For some secondary outcomes, we also present estimates from a standard continuous difference-in-difference specification:

$$\log(y_{jt}) = \beta\left[\text{Win Rate}_j \times \mathbb{1}(t \geq 2007)\right] + \Gamma X_{jt} + \alpha_j + \alpha_t + \varepsilon_{jt}, \tag{3}$$

where the coefficient of interest, β, measures the impact of the win rate on firm-level outcomes for the time period 2007-2011, relative to 2002-2006. 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. 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 nonoutlier 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 nonoutlier 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 when analyzing employment count outcome variables, as values are changed from missing to 0 if the firm is inactive (has zero payroll for the year). As suggested in Roth and Chen (2022), we replace values of $\log(y_{jt})$ to a negative constant (-2) when $y_{jt} = 0$. 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.[21] Analyses with outcomes that are either transformations of revenues, on a per worker basis, or both are conditional on the firm being active and nonmissingness of both the numerator and denominator.[22]

6.2 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 increase observed I-129 approvals for new employment during 2007. Similarly, 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

[21]The Davis-Haltiwanger-Schuh growth rate in x between $t = 0$ and $t = 1$ is $\frac{2(x_1 - x_0)}{x_1 + x_0}$. In future drafts, we intend to show that our results are robust to the specific choice of -2 as the constant in this procedure. Note that $e^{-2} < 1$, so for any outcome relating to employment counts, the support of y_{jt} does not include $(0, e^{-2})$.

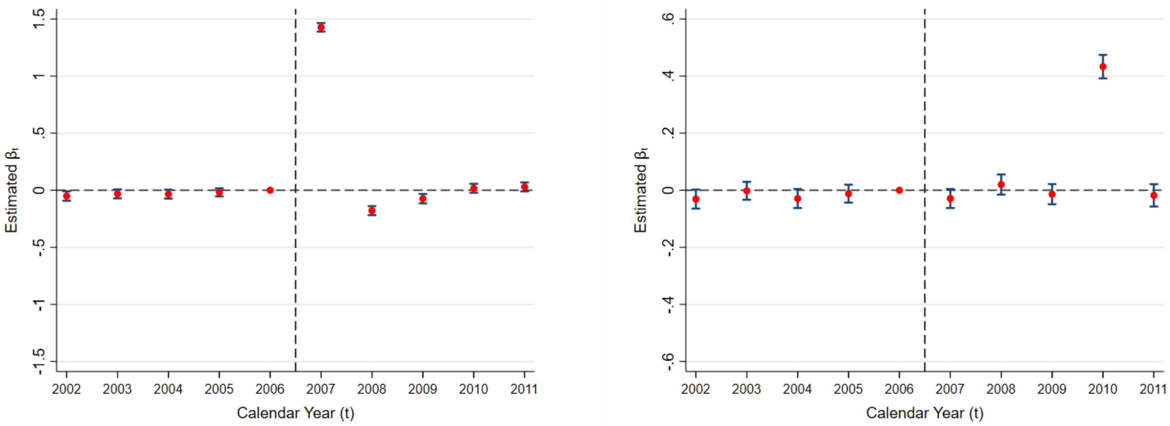
[22]We do not change revenues to 0 when missing because our revenue data originates from a different data source (BRFIRM_REV) than the LBD that is only available for a subset of firms that changes from year to year. See Section 3 for more.

see significant estimates of $\hat{\beta}_\tau$ for new employment or renewals prior to 2007.

Figure 3 presents estimates from Equation (2) for I-129 approvals and provides several pieces of validation for our approach. First, Figure 3a shows no visible 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 firms that lost the 2007 lottery 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 this as another piece of evidence that we have isolated lottery-induced variation in the ability to hire H-1B workers.

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

(a) New Employment (b) Renewals



Note. See Equation (2) for specification. 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 effects). 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 variable in subfigure 3a is the log number of I-129 petitions for new employment in year t. The dependent variable in subfigure 3b is the log number of I-129 petitions for renewal of employment in year t. These data are from the USCIS. Regressions include firm fixed effects, industry-time fixed effects, and log employment in March 2007 interacted with time fixed effects.

We proceed by asking whether the provision of H-1B visas found in Figure 3 translates into utilization of workers with characteristics of H-1B visa holders in the LEHD data.[23] This basic test 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., forthcoming; 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

[23]Recall that visa status is not included in the LEHD data.

to first ask whether a larger immigrant college workforce is an important mechanism through which lottery success affects firm-level outcomes.

We define H-1B-like immigrants as those workers in the LEHD 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.[24] We also test the effect of lottery wins on a firm's total immigrant college workforce—the number of foreign-born workers classified as having a college degree in the LEHD.

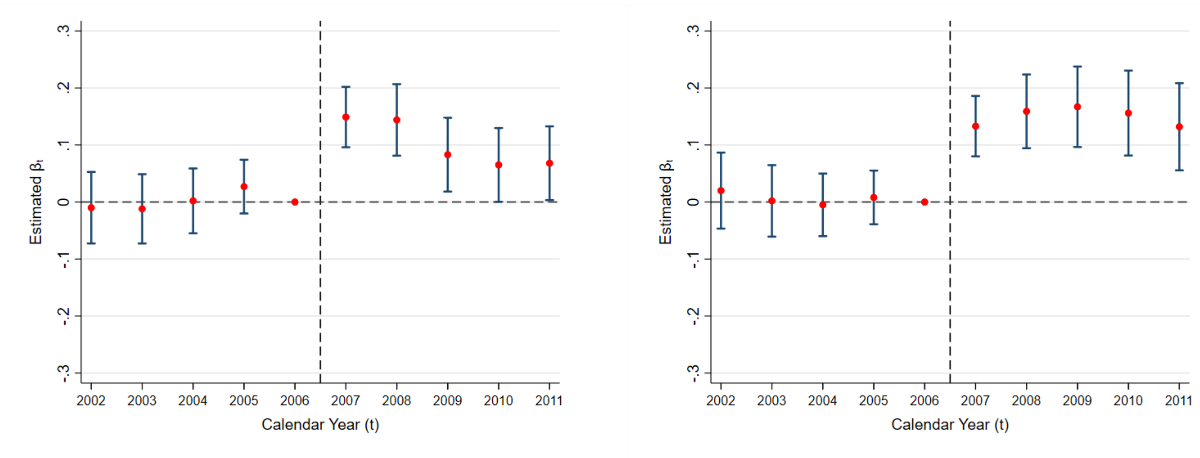
Figure 4 shows that lottery-winning firms increase their use of workers in each category. First, Figure 4a shows a clear increase in winners' employment of H-1B-like immigrant workers immediately after the 2007 lottery. Given that this group has less than three years of tenure at the firm, the hump-shaped response displayed in Figure 4a is specifically consistent with increased hiring of H-1B workers in 2007, as seen in Figure 3a. Figure 4b further shows a permanent increase in the employment of foreign-born workers with college degrees during our study period. Our estimate of $\hat{\beta}_{2009}$ implies that a firm that wins all of its lottery applications ($\text{Win Rate}_j = 1$) increases employment of immigrant college workers by 16.7% relative to a firm that loses all of its lottery applications ($\text{Win Rate}_j = 0$). In total, Figure 4 demonstrates that the H-1B lotteries play an important role in determining whether or not lottery participants have access to college-educated immigrant workers.

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 3, we separately estimate Equation (3) for Indian H-1B-like workers, Canadian/Mexican H-1B-like workers, and foreign workers from all other 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 3 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.

[24]Some reasons why this measure is likely broader than the true H-1B workers at the firm is that the LEHD likely H-1B immigrants might be working on an L-1 status, OPT, or green card. 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. This measure may also miss H-1B workers 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 of Immigrant Workers with College Degrees

(a) H-1B-Like Immigrants (b) All College Immigrants



Note. See Equation (2) for specification. We plot the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. Vertical dashed line separates pre-lottery years (placebo tests) from post-lottery years (treatment effects). Omitted period is 2006. Number of firms is roughly 13,500 (rounded as per the US Census Bureau's disclosure rules). Standard errors clustered at the firm level. The dependent variable in subfigure 4a is the log number of foreign-born workers at the firm who are 25-40 years old, with less than three years of tenure, and with a college degree in the fourth quarter of year t. The dependent variable in subfigure 4b is the log number of all immigrant workers at the firm who had a college degree in the fourth quarter of year t. These outcomes are measured using the LEHD. Regressions include firm fixed effects, industry-time fixed effects, and log employment in March 2007 interacted with time fixed effects.

Table 3: Effect of Lottery Win Rate on Employment of H-1B-Like Workers, by Origin

	Outcome: log H-1B-like immigrant workers			
Origin Region:	All	India	Mexico/Canada	Other
Win Rate$_j$ $\times$ $\mathbb{1}$(Year $\geq$ 2007)	0.100***	0.072***	-0.020**	0.054**
	(0.024)	(0.015)	(0.010)	(0.021)
Firms	13,500	13,500	13,500	13,500
Observations	137,000	137,000	137,000	137,000

Note. See Equation (3) for specification. Standard errors clustered at the firm level. The dependent variable is the log number 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 log employment in March 2007 interacted with time fixed effects. Firm and observation counts rounded as per the US Census Bureau's disclosure rules. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

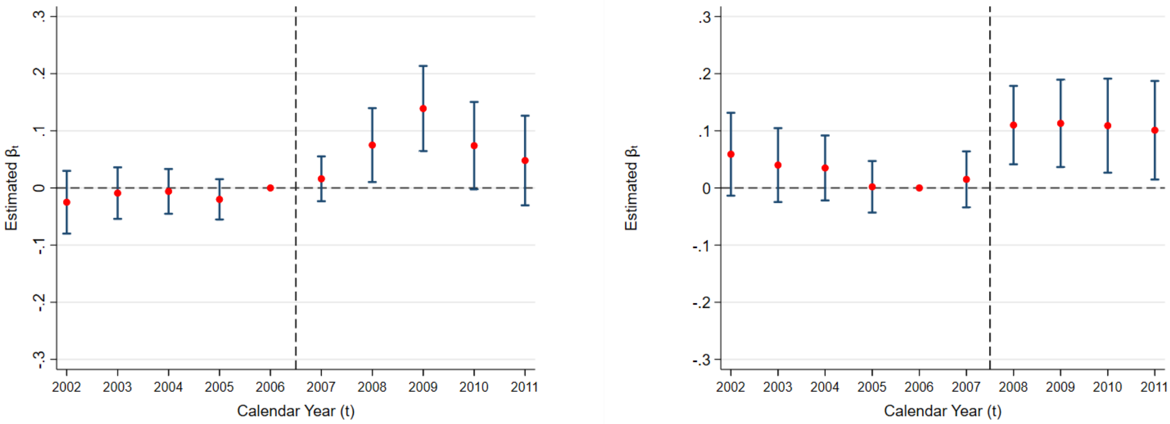
6.3 Firm Performance

Having validated that lottery wins generate a "first stage"—an effect on both H-1B visa approvals and LEHD-measured, likely H-1B hiring—we next turn to the effect of lottery wins on firm performance. These analyses use outcomes measured in the LBD and are available for the full universe of firms.

We start by analyzing impacts on the scale of firm operation. Figure 5 presents estimates from Equation (2) with log of annual revenues and log of Q1 employment as outcomes. Both measures of size increase in the post-lottery period. Figure 5a shows effects of lottery wins on revenues, conditional on survival, that peak in 2009 at a 13.9% increase for firms that win all of their lottery applications relative to firms that lose all of their lottery applications. Figure 5b, meanwhile, shows that employment appears to rise permanently in our study period, by around 11% under the same comparison. For context, our back-of-the-envelope calculation suggests that each i129 approval for a new H-1B worker in 2007 translates to 1.02 total employees (including the H-1B worker themselves) at the firm in March 2009 and 0.91 total employees at the firm by March 2011, inclusive of extensive margin responses.[25] We do not find strong evidence of either crowd out or crowd in of employment at the firm level.

Figure 5: Effect of Lottery Win Rate on Firm Scale

(a) Revenues (b) Q1 Employment



Note. See Equation (2) for specification. We plot the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. Vertical dashed line separates pre-lottery years (placebo tests) from post-lottery years (treatment effects). 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 variable in Subfigure 5a is the log of firm revenues in year t. The dependent variable in Subfigure 5b is the log number of all workers at the firm on March 12 of year t. Note that revenues are measured conditional on firm survival, whereas worker counts are set to 0 when a firm is inactive, so Subfigure 5a does not include extensive margin responses but Subfigure 5b does. These outcomes are measured using the LBD. Regressions include firm fixed effects, industry-time fixed effects, and log employment in March 2007 interacted with time fixed effects.

Table 4 adds three estimates from Equation (3) that provide key, additional context. First, and most importantly, there are large extensive margin responses to lottery wins. Firms that win all of their lottery applications are 2.4 percentage points more likely to have positive, nonmissing 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

[25] We conduct these back-of-the-envelope calculations as follows. Let $\hat{\beta}^y_\tau$ be the estimated effect from Equation (2) with outcome $\log(y_{jt})$. Then, our back-of-the-envelope estimate of the number of additional y in year τ per I-129 lottery-induced approvals for new H-1B workers is given by $\frac{\hat{\beta}^y_\tau}{\hat{\beta}^{\text{I-129}}_{2007}} \frac{Med(y)}{Med(\text{I-129})}$, where $y = $ I-129 refers the outcome log of I-129 approvals for new H-1B employment. The rationale behind this estimate is that $\frac{\hat{\beta}^y_\tau}{\hat{\beta}^{\text{I-129}}_{2007}} \approx \frac{d\log(y)}{d\log(\text{I-129})} = \frac{dy}{d(\text{I-129})}\frac{\text{I-129}}{y}$, and we are after $\frac{dy}{d\text{I-129}}$. $Med(\text{I-129}) = 1$, and we use pseudo-medians calculated from our census data for $Med(y)$.

workers for survival. These extensive margin responses also help reconcile the different-shaped responses found in Figure 5. Recall that missing values of employment are set to 0 when a firm is inactive, whereas missing values of revenues are still treated as missing. This means that Figure 5b includes the sizable extensive margin responses found in Table 4, whereas Figure 5a does not. Second, Table 4 suggests that labor productivity increases 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.

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

		Logged	
	Active	Revenues per Worker	Average Wage
Win Rate$_j$ $\times$ $\mathbb{1}$(Year $\geq$ 2007)	0.024***	0.022*	0.012*
	(0.006)	(0.013)	(0.007)
Firms	20,000	20,000	20,000
Observations	199,000	199,000	199,000

Note. See Equation (3) for specification. Standard errors clustered at the firm level. The dependent variable in the first column is an indicator for whether or not the firm had positive payroll in year t. The dependent variables in the second and third columns are the log of revenues per worker and the log of total payroll per worker in year t, respectively. These outcomes are measured using the LBD. Regressions include firm fixed effects, industry-time fixed effects, and log employment in March 2007 interacted with time fixed effects. Firm and observation counts rounded as per the US Census Bureau's disclosure rules. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In total, the results in Figure 5 and 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 revenues per worker coupled with *increases* in average wages strongly suggest that firms relied on H-1B workers to increase productive capacity.

6.4 Workforce Composition

We next ask how the hiring of H-1B workers alters the worker composition of a firm, utilizing the unique data contained in the LEHD. Of first order importance is the canonical question of whether and how much immigrant workers substitute for native counterparts with similar characteristics, studied here in the context of the H-1B program. We are also interested in whether hiring H-1B workers leads to complementary increases of other worker types.

To study these questions, we use worker counts from different groups as outcomes in Equation (2) and plot the results in Figure 6. To start, Figure 6a simply validates that overall expansions in Q4 firm employment—measured in our limited LEHD sample—mimic the expansions seen in Figure 5b—measured in our full LBD sample—both in sign and magnitude.[26] Next, Figure

[26]Recall that H-1B workers hired through the 2007 lottery started working in Q4 of 2007. So, we expect

6b shows suggestive but imprecise evidence of expansions in the noncollege workforce at lottery winners relative to lottery losers, consistent with the scale and extensive margin responses reported in Section 6.3.[27]

We next turn to our key estimates on immigrant-native substitutability in Figures 6c and 6d, which can be compared directly to Figure 4a and 4b. Figure 6c presents evidence of limited substitution away from native workers who we consider most substitutable with H-1B workers—those who are 25-40 years old, with less than three years of tenure at the firm, and with a college degree.[28] For context, the statistically significant decline in "H-1B-like native" employment seen in 2009 translates to a drop in employment of less than 0.1 per I-129 approval according to our back-of-the-envelope calculations. Meanwhile, Figure 6d does not show any evidence of a decline in the native college workforce more broadly.

All told, Section 6 finds that lottery wins in 2007 increase the scale of firm operation through increased hiring of college-educated immigrant workers. As a result of this increased scale, most worker groups are insulated from crowd out. The lone exception is the group of most-substitutable natives, who experience slight reductions in employment at lottery winners.

7 Heterogeneity Analysis

As a final step, we investigate whether certain firm characteristics drive our results. We focus our analysis on two main outcomes: total firm employment and native H-1B-like employment. To estimate differential response to winning the H-1B lottery across firms, we set up a triple difference approach as in Equation (4):

$$\begin{aligned}\log(y_{jt}) =& \beta\left[\text{Win Rate}_j \times \mathbb{1}\left(t \geq 2007\right)\right] + \gamma\left[\mathbb{1}\left(t \geq 2007\right) \times Z_j\right] + \\ & \delta\left[\text{Win Rate}_j \times \mathbb{1}\left(t \geq 2007\right) \times Z_j\right] + \Gamma X_{jt} + \alpha_j + \alpha_t + \varepsilon_{jt}\ ,\end{aligned} \tag{4}$$

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 SD above the mean in characteristic Z_j relative

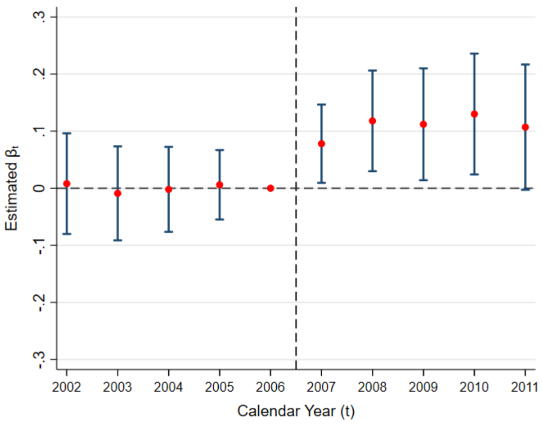
to see a response in 2007 in Q4 LEHD employment but not in 2007 LBD employment, which is measured on March 12, 2007.

[27] The difference-in-difference estimate for noncollege workers is a marginally statistically significant 0.060. See Appendix Section C.

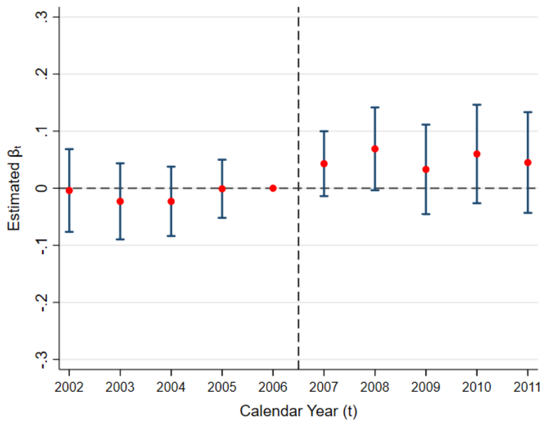
[28] This exactly mimics our definition of "H-1B-like immigrant" workers except for nativity.

Figure 6: Effect of Lottery Win Rate on Employment of Selected Subgroups

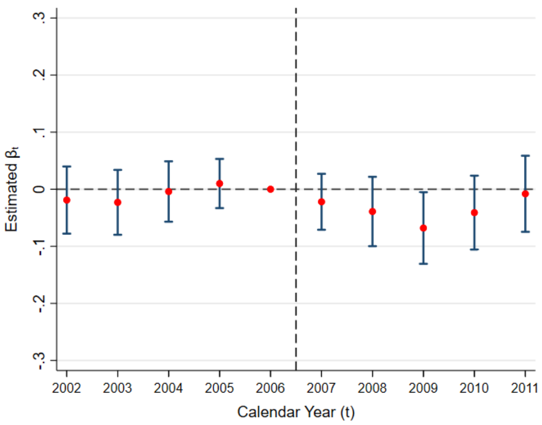
(a) Q4 Employment



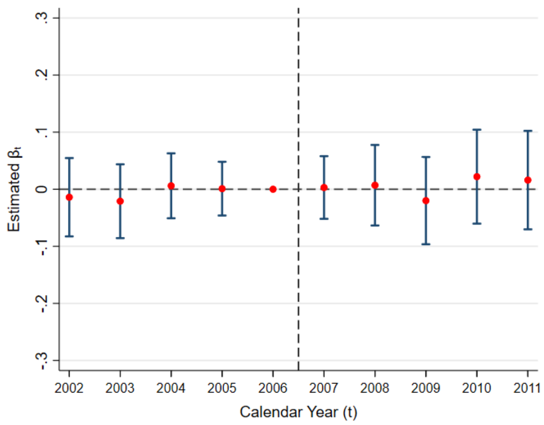
(b) Noncollege



(c) H-1B-Like Natives



(d) All College Natives



Note. See Equation (2) for specification. We plot the estimated coefficients $\hat{\beta}_\tau$ and their 95% confidence intervals. Vertical dashed line separates pre-lottery years (placebo tests) from post-lottery years (treatment effects). Omitted period is 2006. Number of firms is roughly 13,500 (rounded as per the US Census Bureau's disclosure rules). Standard errors clustered at the firm level. The dependent variable in subfigure 6a is the log number of workers at the firm in the fourth quarter of year t. The dependent variable in subfigure 6b is the log number of workers at the firm who did not have a college degree in the fourth quarter of year t. The dependent variable in subfigure 6c is the log number of native workers at the firm who are 25-40 years-old and with fewer than three years of tenure in the fourth quarter of year t. The dependent variable in subfigure 6d is the log number of native workers at the firm who had a college degree in the fourth quarter of year t. These outcomes are measured using the LEHD. Regressions include firm fixed effects, industry-time fixed effects, and log employment in March 2007 interacted with time fixed effects.

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 look at firm size measured as total firm employment. Second, we compute the number of college graduates as a share of firm employment to proxy for the skill intensity of the firm. Third, we compute the number of immigrant college graduates as a share of total college graduates, as a measure of immigrant intensity for college labor. Finally, we compute average wage as a measure of overall worker quality and skill.

As shown in Table 5, there are significant differences across some of these characteristics. The triple interaction term when looking at firm size is not significant, meaning that there is not a differential response of winning the lottery between firms with average size and firms with one SD above average size. However, our measures of skill-intensity and immigrant-intensity present similar pictures. In terms of total employment, firms with a higher college share or immigrant share expand by almost double compared to firms with average value of these characteristics. Similarly, the additional immigrants hired on H-1B visas are less likely to crowd out similar natives in these firms. A possible interpretation for this result is that firms that are skill-intensive and immigrant-intensive, particularly value the H-1B immigrants as a distinct input in production. Hence, the availability of immigrants is crucial for these firms where immigrants are less substitutable with natives.

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

	Outcome: Q4 Employment			
Characteristic (Z_j):	Firm Size	College Share	Immigrant Share	Average Wage
Win Rate$_j$ × $\mathbb{1}$(Year ≥ 2007)	0.110**	0.104**	0.097**	0.060
	(0.045)	(0.046)	(0.047)	(0.047)
Z_j × $\mathbb{1}$(year ≥ 2007)	-0.051**	0.046	0.064**	0.059**
	(0.026)	(0.029)	(0.03)	(0.03)
Win Rate$_j$ × Z_j × $\mathbb{1}$(Year ≥ 2007)	-0.016	0.088*	0.125***	0.142***
	(0.039)	(0.046)	(0.045)	(0.053)
	Outcome: Native H-1B Like Employment			
Characteristic (Z_j):	Firm Size	College Share	Immigrant Share	Average Wage
Win Rate$_j$ × $\mathbb{1}$(year ≥ 2007)	-0.027	-0.032	-0.024	-0.044*
	(0.025)	(0.026)	(0.026)	(0.026)
Z_j × $\mathbb{1}$(Year ≥ 2007)	-0.037**	-0.018	0.110***	0.017
	(0.017)	(0.014)	(0.016)	(0.013)
Win Rate$_j$ × Z_j × $\mathbb{1}$(Year ≥ 2007)	-0.004	0.038*	0.057**	0.068**
	(0.026)	(0.023)	(0.023)	(0.027)
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 total employment, taken from the LEHD. The bottom panel uses as outcome the total employment of H-1B-like natives, also computed through the LEHD. The characteristics Z_j and standardized and computed in year 2006.

Finally, the last column of Table 5 looks at heterogeneity by average wage, and two main results stand out. First, as shown in the top panel, the larger employment response driven by winning the lottery seems to be fully driven by firms that pay higher wages. Going from zero to one win rate increases employment by 20.2% (6% + 14.2%) for firms with one SD above the mean in terms of average wage, while only increasing by a nonstatistically significant 6% for firms with the mean value of average wage. Second, the larger increase in employment for high-wage firms is partially driven by a lower native crowd out. As shown in the bottom panel, firms at the mean value of average wages have a statistically significant decrease in the number of H-1B-like natives employed at the firm, whereas H-1B wins appear to crowd in natives at firms that pay higher wages.

8 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 et al. (forthcoming), 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, Doran et al. (forthcoming) focus predominantly on the lotteries of 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 one.

The main reason that these results diverge is likely that the sample of firms that applied for visas on the last day of filing for 2005 and 2006 (Doran et al., forthcoming) are different than the full sample of H-1B applicants that applied in 2007 (as in our paper). As shown in Section 7, firms that expand the most after winning the lottery are high-paying firms that are skill-intensive and immigrant-intensive. Hence, firms that really need the immigrants for production are likely to submit their applications as soon as they can, while firms where immigrants are not as critical to production may apply with less urgency.

Closer to our estimation strategy, Doran et al. (forthcoming) run a robustness check using the full sample of applicants in 2007, proxying the number of applications with the number of LCAs. In this case, they find that one additional H-1B win increases firm employment by just 0.36 employees (crowd out of 0.64).[29] While this estimate is lower than our estimate of one-for-one increase, we note that our estimates are contained within their confidence intervals.

We also want to emphasize two key improvements to the Doran et al. (forthcoming) approach with the 2007 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

[29] 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.

likely submitted for purposes other than new employment, and firm-level application counts that were extreme outliers. Second, as discussed in Section 5, 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.

9 Conclusion and Next Steps

All told, this paper provides several new, well-identified estimates to the empirical literature on the impact of high-skill immigration on US firms. A key finding is that firms expand their employment, revenues, and productivity when hiring high-skill immigrants through their H-1B program. While there is some crowd out of natives who are similar to the H-1B workers, it is small in magnitude. One additional H-1B worker increases total firm employment by one additional employee in 2009, and about 0.9 additional employees in 2011. High-wage, college-intensive, and immigrant-intensive firms expand more and even crowd in natives when winning the lottery.

Our findings have implications for immigration policy design and the broader welfare effects of immigration to advanced economies. We aim to build on our current set of analyses in two main ways. First, we will add a theoretical framework to rationalize our results and corroborate they are consistent with standard models of immigration.

Second, we fill an important gap in the literature by analyzing how individual workers absorb the shock of being exposed to H-1B workers. We focus on those employees who worked at the lottery subject firms in 2006, and trace their employment and wage outcomes over time using an event study analysis. In ongoing analysis, we accomplish this by adapting our baseline regression to the individual-level as in equation 5:

$$y_{it} = \sum_{\tau \neq 2006} \beta_\tau \left[\text{Win Rate}_{j(i)} \times \mathbb{1}(\tau = t)\right] + \Gamma X_{it} + \alpha_i + \alpha_t + \varepsilon_{it}, \tag{5}$$

where $j(i)$ indexes the FY2008-lottery-participating firm that employed individual i in 2006.

These models provide the first direct evidence of how incumbent workers' career paths are altered when exposed to H-1B workers. We estimate Equation (5)—split into education-by-tenure-by-nativity groups—with several outcome variables, including earnings, an indicator for staying at firm $j(i)$, and an indicator for staying in $j(i)$'s industry. These results shed new light on 1) whether and which incumbent workers benefit or suffer from H-1B hiring and 2) the role of mobility across firms and industries in mediating these effects. Of particular note, the ability to follow individuals across firms may help reconcile the large divergence between estimated market-wide and firm-level impacts of the H-1B program in prior literature (e.g., Doran et al., forthcoming; Kerr and Lincoln, 2010; Peri et al., 2015b).

References

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

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.

Brinatti, Agostina and Nicolas Morales, "Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments," 2021. Working Paper.

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).

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).

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.

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*, forthcoming.

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.

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 U.S. Firms," CESifo Working Paper 9874, CESifo 2022.

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.," 2021. Working Paper.

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.

A 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.[30] 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 8 shows the total number of approved LCA applications by week, for each H-1B filing 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.

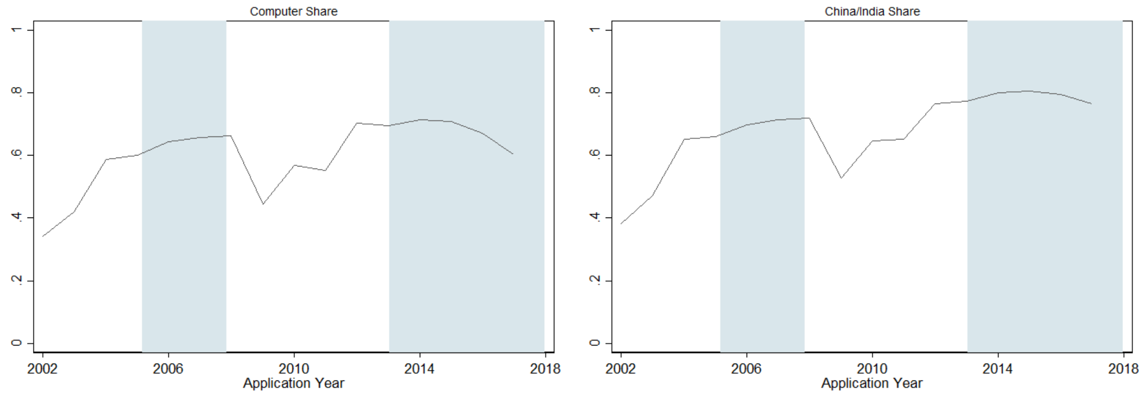
Such predating behavior presents at least two issues. First, this implies particular selection of firms into the lottery. In its simplest form, this means firms that could anticipate high demand and a lottery, ended up predating LCA applications to ensure they could submit a complete H-1B application 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 7. 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.

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

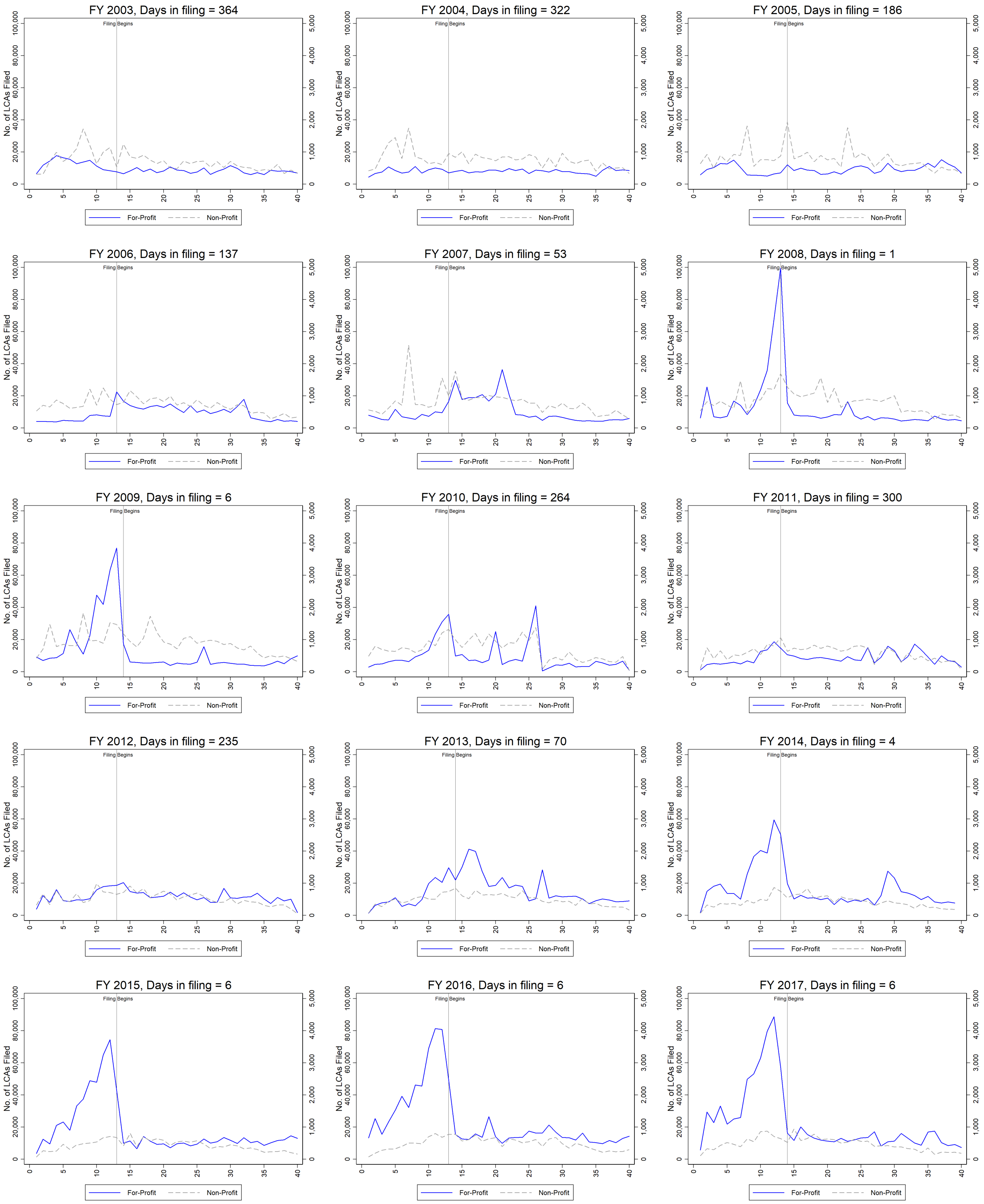
A second issue pertains to the internal validity of this natural experiment. Firms anticipating the lottery may be able to partially mitigate its impact. For example, a firm that anticipates a lottery and believes the odds to be equivalent to a coin flip can mitigate the lottery rationing by applying 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.

Figure 7: 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.

Figure 8: 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).

B 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.[31] 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 9 illustrates the utility of filtering out probabilistic outliers visually. Figures 9a and 9b 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 9c 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 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 9c is confirmed by regression analysis presented below, in Table 6, 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 and 6.1.

Table 6: 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

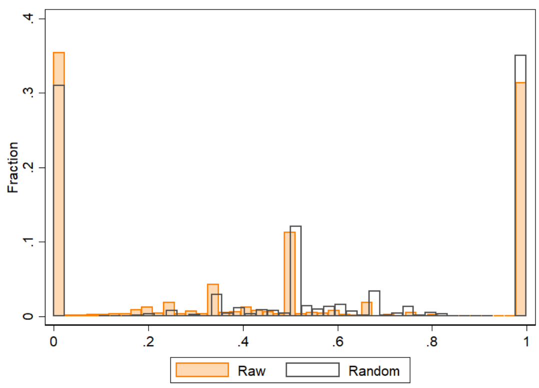
Note. Robust standard errors in parentheses. Both "Filtered" and "Raw" refer to our measure Win Rate$_j$, but "Filtered" refers to the measure after outliers have been eliminated by the procedure described above, in Section B. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[31] 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.

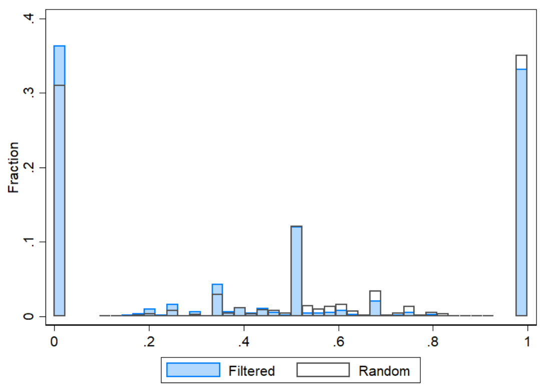
In total, our filtering method has the benefit of retaining only firms for whom we believe Win $\text{Rate}_j \approx$ Ideal Win Rate_j but at the cost of only analyzing 73,180 out of 123,480 total lottery applications.

Figure 9: Filtered and Raw Win Rates

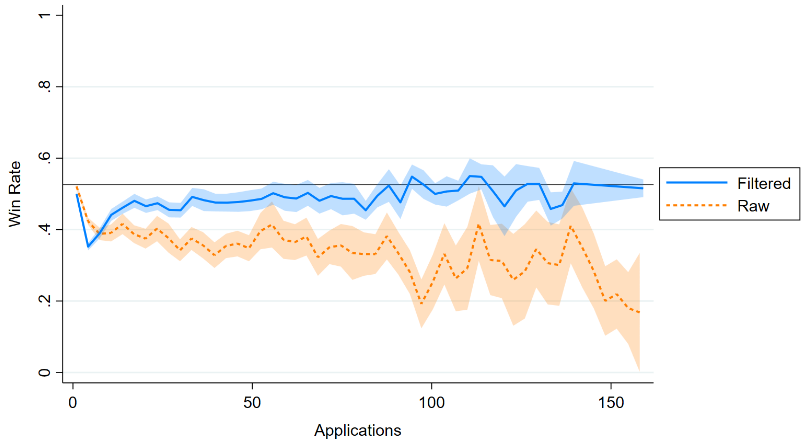
(a) Raw vs. Random Win Rate



(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 B. There are 20,072 EINs represented in "Raw" win rates, and 18,963 EINs represented in the "Filtered" win rates.

C Additional Difference-in-Difference Results

Table 7 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. Event study Figures 5b and 6a find very similar results across the two data sets, with a roughly stable 10-11% increase in employment for firms that won all of their lottery applications relative to firms that lost all of their lottery applications in the post-lottery period. Yet, Table 7 shows a substiantially lower estimate for LBD employment. This is an unfortunate consequence of timing: LBD employment is measured in Q1, which means that we do not expect responses in 2007 (as seen in Figure 5b). Thus, the indicator $\mathbb{1}(\text{Year} \geq 2007)$ includes a year of nonresponse that it should not for LBD-measured employment. We plan to fix this issue by using $\mathbb{1}(\text{Year} \geq 2008)$ as the "after" period in difference-in-difference estimates when LBD-measured employment is the outcome in future drafts.

Table 7: Difference in Difference Estimates

	Panel A: Firm Scale			
	LBD Employment	LEHD Employment	Revenues	Noncollege Workers
Win Rate$_j$ × $\mathbb{1}(\text{Year} \geq 2007)$	0.063*	0.108**	0.079***	0.060*
	(0.035)	(0.045)	(0.023)	(0.036)
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	Immigrant H-1B-Like	Native College	Native H-1B-Like
Win Rate$_j$ × $\mathbb{1}(\text{Year} \geq 2007)$	0.145***	0.100***	0.011	-0.028
	(0.032)	(0.024)	(0.035)	(0.025)
Firms	13,500	13,500	13,500	13,500
Observations	137,000	137,000	137,000	137,000

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