The impact of immigration on firms and workers: insights from the H-1B lottery and US employer-employee data*

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

December 2022

Preliminary and incomplete. Please do not distribute.

Abstract

We study how random variation in the availability of highly educated immigrants impacts firm performance and recruitment behavior. To do so we combine two rich data sources: 1) the Longitudinal Employer-Household Dynamics (LEHD), an employer-employee matched data set from the US Census Bureau, and 2) firm-level information on the large-scale 2007 and 2008 H-1B visa lotteries. Using an event-study approach, we find that lottery wins induce temporary increases in a firm's college-educated immigrant workforce but permanent increases in revenues and labor productivity. We do not find evidence that H-1B immigrants crowd out workers at the firm level.

JEL: F22, J61
Keywords: Immigration, firm dynamics, productivity, H-1B visa

*Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2104. (CBDRB-FY22-P2104-R9930). 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.

†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 largely determined by the recruitment decisions and subsequent performance of individual firms. Yet, two major barriers have limited prior literature from studying the consequences of immigration through the lens of the firm. 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 through the number of immigrants in the labor market, ignoring within-firm adjustments and heterogeneity across firms within the labor market.

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 U.S. productivity growth. We study the impact of a plausibly exogenous shock to a firm's ability to hire new immigrants generated by the 2007 and 2008 H-1B visa lotteries. 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.

Our analysis proceeds in three steps. We begin by constructing a proxy of a firm's success in the 2007 and 2008 H-1B lotteries. In 2007 and 2008, as applications for H-1B visas immediately exceeded the legal cap, *all* applications went through a lottery to determine which visa applicants were allowed to work in the US. 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 this lottery measure to restricted-access US Census Bureau economic data using the a fuzzy string matching procedure. The economic data primarily consists of the Longitudinal Employer-Household Survey (LEHD), a data set on the universe of employer-employee matches of 25 US states for the period between 2000 and 2014, which includes wages, employment histories, and workers' country of birth among other variables. We complement these data with firm-level revenue information attached to 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 US respond to immigration restrictions.

Finally, to understand the impact of the lottery on firm performance and hiring, we set up an

event study analysis in which we can both evaluate the plausibility that our lottery measure is exogenous and examine its impact on firm-level outcomes. We show that lottery winners employ more young, low-tenure immigrant workers with college degrees in 2008 and 2009, the fiscal years for which the lottery applied. We then show that these temporary increases in likely H-1B immigrant employment lead to longer-run, lasting increases in revenues and revenues per worker. Meanwhile, we find that each additional H-1B worker is associated with 1.38 additional workers overall, implying crowd-in employment at the firm level. We do not find statistically significant evidence for crowd-out of native workers with similar experience and education to H-1B immigrant workers, but our point estimates do suggest the possibility of a small amount of native crowd-out among this finer grouping.

Lack of crowd-out at the firm level, in aggregate, does 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. Outcomes include tenure at the original, lottery-participant firm, probability of job-switching and industry-switching, and earnings. 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, to our knowledge, we are the only paper other than Kerr et al. (2015) 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 lotteries. 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 small but growing literature studies the impact of immigrant workers on firm performance using employer-level data (see, e.g., Brinatti and Morales, 2021; Kerr et al., 2015; Mahajan, 2022; Mayda et al., 2020; Mitaritonna et al., 2017). The most closely related work is Doran et al. (forth.), who uses 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 who applied for an H-1B visa. In contrast, we focus on the lotteries of 2007 and 2008, where H-1B visa applications exceeded the cap within the first week and were therefore *all* placed into the lottery. Our analysis covers 14,500 LEHD-enumerated firms applying for cap-subject H-1B visas in 2007 and 2008, and we believe that our work thus complements Doran et al. (forth.) by analyzing a more representative and powerful sample of H-1B employers.[1,2]

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

[2] Doran et al. (forth.) also analyze the effect of 2007 lottery wins on employment and patents as a robustness

The LEHD also 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. (forth.) is that we do not observe applications that lost the lottery. To get around this issue, we construct and validate a proxy measure for lottery success 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 (Kerr and Lincoln, 2010), local employment opportunities in H-1B-related occupations (Peri et al., 2015a), and local productivity (Peri et al., 2015b). Of 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.

The rest of this paper is organized as follows. Section 2 provides an overview of the institutional details of the H-1B program. Section 3 describes the data used in the analysis. Section 4 describes how we use the data to construct our win rate proxy among lottery participants in the 2007 and 2008 H-1B lotteries. Section 5 describes our analysis and results. Section 6 describes next steps and concludes.

2 Overview of the H-1B Program

The H-1B visa was created within the Immigration Act of 1990, which replaced the preexisting H-1 visa for temporary workers with the H-1A visa for nurses and the H-1B visa for workers in specialty occupations. The law defined requirements on both the type of occupation—"*requires theoretical and practical applications of a body of highly specialized knowledge*"—and the educational attainment required—"*a bachelor's or higher degree in the specific speciality (or its equivalent)*"—for a potential H-1B visa holder. Additionally, the law also placed a numerical quota of 65,000 new visas per year.

Since its inception, the H-1B visa has undergone a variety of reforms primarily related to the quota. In 1998, Congress authorized an increase in the quota from 65,000 visas per year to 115,000 for FYs 1999 and 2000. Congress then authorized another increase in the quota to 195,000, which lasted from FYs 2001-2003. The quota fell back to its initial level of 65,000 in

check on their 2005-06, using a methodology that follows the broad contours of ours but with some important differences. They find a strong negative effect on employment for the 2007 lottery, which is not consistent with our results. Future drafts will fully elucidate whether and why our results may differ from theirs.

FY 2004.[3]

In 2000, U.S. 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. Additionally, in 2004 an additional 20,000 visas were reserved for individuals with a Master's degree or higher from a U.S. educational institution, referred to as the "Advanced Degree Exemption" (ADE) cap. Hence, the total cap since FY 2004 has remained constant at 85,000–65,000 visas under the regular quota plus the 20,000 under the ADE quota.

Applications for H-1B visas are submitted by employers on behalf of workers they wish to hire. There are several steps involved in order to submit an H-1B petition. A firm wishing to hire an H-1B worker must first file a Labor Conditions Application (LCA) with the Department of Labor. Only after receiving LCA approval, can a firm then apply for the H-1B visa.

The LCA outlines information about the job, the number of positions a firm wishes to fill, the work start and end dates, the work location, and the associated salary. In filing the LCA, firms must attest that they have informed incumbent workers of their intention to hire immigrants on H-1B visas, and that H-1B workers will be afforded the same benefits and working conditions as other workers in the firm, and will not be paid below market wages.

LCA filing is governed by a unique rule that limits applications to a maximum 6-month "filing window" prior to the employment start date. Specifically, LCAs cannot be submitted earlier than 180 days (6 months) prior to the employment start date listed on the application. Hence, an employer that completes an LCA with a work start date of October 1st may not actually submit the LCA to the Department of Labor for processing *before* April 1st of that year.

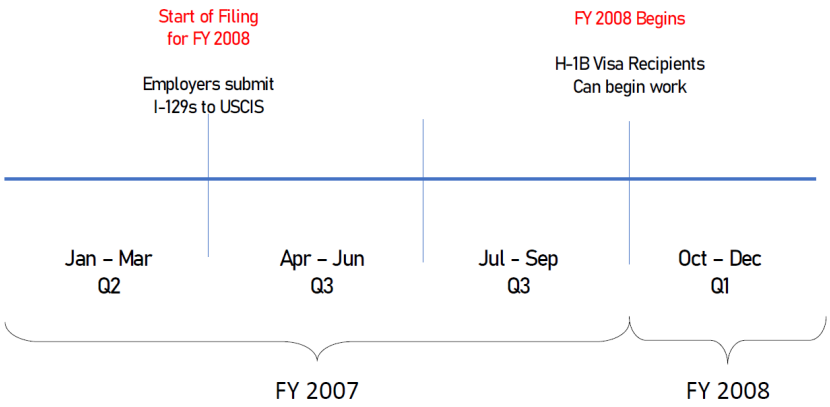
Following LCA approval, employers may then submit an H-1B visa application. The primary document of an H-1B application is the I-129 form, which provides information about the employer, demographic and other personal information about the worker, the occupation, the work start and end dates, and the wage/salary to be paid. In addition to the I-129 form, a completed 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.

As depicted in Figure 1, USCIS begins accepting H-1B applications for each fiscal year on the first business day in April (usually April 1st or 2nd) of the prior calendar year. The earliest date H-1B visa recipients can begin working is the start of the fiscal year, which occurs on October 1st of the prior calendar year. Figure 1 illustrates the timeline, using the application season for FY 2008 as an example. Applications traditionally have been processed on a first-come

[3] At times prior to 2000, the number of visas issued exceeded the quota as the result of processing errors where too many visas were awarded.

first-served basis, until the cap is reached.

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



Note. Figure illustrates the application timeline using calendar year 2007 and its lottery for start-of-fiscal-year 2008 employment as an example. Quarters are labeled as the correspond to fiscal year.

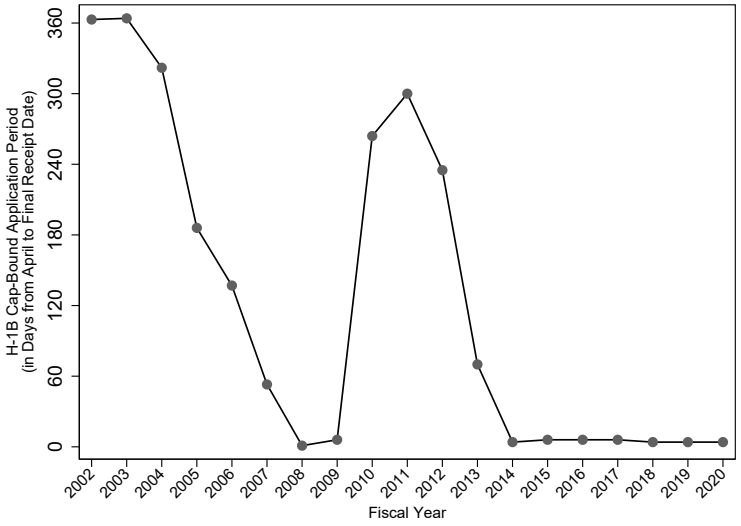
In 2005, USCIS updated their procedures (see *70 FR 23775, May 2005*) to allow for randomized selection of petitions under two scenarios. First, petitions received on the "final receipt date"—defined as they date on which USCIS received enough petitions to reach the cap—will be subject to a randomized lottery. Second, if the final receipt date is the first day of the application season, the entire cap will be randomly allocated across applications received on the first two days of the filing period. In the application seasons in 2005 and 2006 for FY 2006 and FY 2007 employment, USCIS held small lotteries on the final receipt dates in those two years.[4] These smaller lotteries were the studied by Doran et al. (forth.).

As seen in Figure 2, since FY 2004, demand for visas grew until the Great Recession, and the cap started to be exhausted much more quickly. USCIS began accepting H-1B applications for FY 2008 on April 2nd, 2007. Adjudicators quickly realized that the number of applications received on April 2nd already exceeded the cap, and thus informed applicants that a lottery would be held to distribute visas. USCIS processed 123,400 received on April 2nd and 3rd of 2007, tagging them with numerical identifiers for random selection by computer. The lottery was only held on the Regular cap for 65,000, and not the ADE cap as the volume of applications did not exceed the ADE cap by the first day of April. Hence, according to USCIS published memos, 65,000 petitions were drawn from a pool of 123,480 applications, yielding an overall win-rate of close to 50%. A similar procedure was followed, with 163,000 applications received for 85,000 cap visas between April 1st and April 8th of 2008, implying a similar win rate.[5]

[4]The final receipt dates for the FY2006 Regular visa, FY2006 ADE visa, FY2007 Regular visa, and the FY2007 ADE visa, were August 10, 2005, January 17, 2006, May 26, 2006, and July 26, 2006, respectively

[5]ADE visas were included in the 2008 lottery.

Figure 2: Days-in-Filing for the H-1B Visa Cap, FY 2002-2020



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.

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 U.S. Census Bureau's, the Longitudinal Employer-Household Dynamics (LEHD). The LEHD is a panel on the universe of individual worker histories and their associated firms for 25 U.S. states. In our case, we obtained information on Alabama, Arizona, Arkansas, California, Colorado, District of Columbia, Delaware, Idaho, Illinois, Indiana, Iowa, Kansas, Maryland, Michigan, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Tennessee, and Utah. For the period 2002-2012, our states account for 42% of the I-129 petitions for new employment H-1B visas, 47.5% of total college educated workers, and 46.5% of total college educated immigrants in the United States.[6]

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 Census using observed education for individuals that participate in the American Community Survey. When constructing

[6]Calculations done with I-129 records and Public Use American Community Survey.

our employment outcomes using the LEHD, we focus on the observed employment and earnings in the first quarter of each year. More details on the LEHD can be found in Vilhuber (2018).

The LEHD can be linked to the Longitudinal Business Database (LBD), which consists of the universe of private-sector establishments in the US, with information on employment, payroll, revenues, firm age, industry, and exact location. Furthermore, around 75% of LBD firms can be matched to firm-level revenues on an annual basis. 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, the LEHD and LBD allow us to look at the near-universe of firms that participate in the program.

3.2 H-1B Data

Our H-1B visa information comes from two different sources. First, we get data on I-129 records obtained by a Freedom of Information Act request from the U.S. Citizenship and Immigration Services (USCIS) for the period spanning 2001 to 2018. Our records cover FYs 2002-2019 and 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 and location. Information about the job position include the occupation, salary, and work start and end dates. As noted before, for those I-129 petitions that are cap subject and go through the H-1B lottery, USCIS only keeps information of those applications that win the lottery. Hence, we can use the I-129 records to infer the winners of the H-1B lottery 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 FYs 2002 through 2018. 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.

Unfortunately, there is no common identifier that allows us to merge the Census administrative data with the I-129 and LCA records. We proceed by using a fuzzy matching approach to merge the I-129 and the LCA records with the US Census by firm name, city and state as developed by Flaeen and Wasi (2015). We run the matching procedure in two steps. First, we construct a dataset with firm identifiers, firm names, address, city, and state for the universe of companies available in the LBD. Second, we run the matching algorithm between the Census data and the I-129/LCA records matching by name, city and state. If an observation is still

unmatched, we repeat the process by matching only by name and state. Once the process is done, we manually inspected the list of companies, removing poor matches, to ensure that the quality of the match is high.

4 H-1B Visa Lottery Construction

Our variation in a firm's ability to hire highly educated immigrant labor comes from the 2007 and 2008 H-1B lotteries. The ideal measure of lottery success would divide the number of successful I-129 H-1B petitions for a given firm by the total number of I-129 H-1B petitions for that firm:

$$wr_j \equiv \frac{\sum_{t=2007,2008} \text{Wins}_{jt}}{\sum_{t=2007,2008} \text{Applications}_{jt}}$$

where Wins_{jt} is the number of I-129 H-1B applications filed by firm j that were successful in year t's lottery and Applications_{jt} is the total number of I-129 H-1B applications filed by firm j that were entered into year t's lottery. wr_j is then lottery win rate for firm j across the 2007 and 2008 lotteries.

There are two limitations in the data that prevent us from computing the ideal measure. First, while USCIS processed lottery winning applications in 2007 and 2008, 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. Second, the I-129 data set does not specify if the H-1B application is cap-subject or not.

To overcome these limitations, we start by generating $\widehat{\text{Wins}}_{jt}$ using the FOIA USCIS I-129 data described in Section 3.2, which contains information on successful I-129 applications. We classify successful petitions as lottery-subject if they were received between the first business day of April and the final receipt date,[7] came from non-university, private sector firms, and requested employment start date on October 1st or later—the first day of the following fiscal year. As seen in Table 1, these restrictions result in 79,231 identified wins in 2007 and 78,177 identified wins in 2008, each of which closely correspond to the cap figure of 85,000.[8]

Without information on applications that lost the lottery, we cannot construct as direct of a proxy for Applications_{jt} as we can for Wins_{jt}. Instead, we take advantage of data from the LCA step in the H-1B worker procurement process to generate $\widehat{\text{Applications}}_{jt}$. Naive use of all LCAs filed for employment in a given fiscal year would include far more than the actual

[7] April 3rd in 2007 and April 8th in 2008.

[8] Because the USCIS data used for this draft does not distinguish applications for workers with an advanced degree with applications for workers with a bachelor's degree, we implicitly include ADE in our measures even though the ADE cap was not subject to the lottery in 2007, and was subject to a different lottery in 2008. In a future draft, we will be able to separate ADE visas from non-ADE visas thanks to a new version of the data that contains educational attainment of the prospective H-1B visa holder.

number of I-129 applications eventually filed by firms for two main reasons: 1) LCAs are also filed for H-1B renewals and other changes of status, not just for new hires; 2) LCAs are costless to file (in monetary terms) while I-129 applications carry a fee, so firms may file LCAs without a strong commitment to eventually filing an I-129 application. In order to hone in on LCAs that eventually led to I-129 applications for new H-1B visa holders, we retain LCA applications for 3-year work duration, filed between February and the final I-129 application receipt date, and that maintain a 6-month duration between LCA submission date and employment start date for $\widehat{\text{Applications}}_{jt}$.

The 6-month restriction bears further explanation because it reflects additional nuance in the H-1B application process. 6 months is the maximum duration permitted between LCA filing and employment start date indicated on an LCA application. However, the first day a new H-1B visa holder can actually work is October 1, the start of the new fiscal year—as reflected in *I-129 applications* for new H-1B workers. So, firms that filed LCAs in lottery years were balancing getting LCA applications in early to ensure approval by the start of the I-129 application period against loss of actual work duration if they indicated a start date prior to October 1, since maximum work duration is 3 years. For example, a firm that filed an LCA on February 1, 2007 for a new, 3-year, cap-subject H-1B visa worker would indicate employment start on August 1, 2007 but would only be able to employ that worker from October 1, 2007 through August 1, 2010, losing the ability to employ the worker in August and September of 2010. Peri et al. (2015a) provide evidence that such "pre-dating" was common. In Appendix A, we track the pattern of LCA filing in for each fiscal year and further document the phenomenon of pre-dating.

Finally, returning to our construction of $\widehat{\text{Applications}}_{jt}$, we also net out successful I-129 *non-lottery* filings filed between February and April of a given lottery year from our retained LCA applications. This decreases our chances of counting non-lottery LCAs.[9]

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

$$\widehat{wr}_j \equiv \frac{\sum_{t=2007,2008} \widehat{\text{Wins}}_{jt}}{\sum_{t=2007,2008} \widehat{\text{Applications}}_{jt}} \tag{1}$$

where we eliminate a small number of outlier observations with $\widehat{wr}_j < 0$ or $\widehat{wr}_j > 1$ since $wr_j \in [0, 1]$.

Table 1 provides further details and context around $\widehat{wr}_j$.[10] Several key points emerge. Perhaps most importantly, despite our restrictions, $\widehat{\text{Applications}}_{jt}$ still overestimates Applications_{jt} as

[9] Recall that each I-129 requires a previously successful, corresponding LCA application.

[10] In order to avoid excessive disclosure avoidance review at this juncture, Table 1 is constructed at the Employer Identification Number (EIN) level and data from a different, newer FOIA request relative to the one described in Section 3.2. EINs were only included in the newer request. As described in Section 6, this new request will help us make improvements to our lottery measure in future drafts.

of this draft. For example, the USCIS reported receiving 123,480 applications for 65,000 non-ADE, cap-subject visas in 2007, implying an overall win rate of 0.53. Our analogous figure for 2007 is 0.43 despite including ADE H-1B applications. In 2008, the true overall win rate was 0.52, while our analogous figure is 0.41. A key concern is that the source of these discrepancies are firms that apply for an LCA but then face a negative labor demand shock that lowers their probability of following through with an I-129 application. In future drafts, we plan to exploit the quarterly frequency of the LEHD to assess whether firms with $wr_j = 0$ were more likely to be experiencing negative shocks during the first quarter of a calendar lottery year. In general, honing our measure of $\widehat{\text{Applications}}_{jt}$ will be a key focus of future drafts.

Table 1 also helps contextualize the results that follow in Section 5. While there is a long right tail in the application (and therefore win) distribution, the majority of firms ($\approx 62\%$) apply for just one visa worker and then have a roughly 50% chance of obtaining that worker in the lottery. Given that our regressions are at the firm-year level and not weighted by firm size, the majority of our identifying variation comes from these firms.

Table 1: Estimated Lottery Characteristics

	Lottery		
	2007	2008	Combined
Panel A: $\widehat{\text{Wins}}_{jt}$			
Mean	2.66	2.67	3.27
Std. Dev.	(30.65)	(33.12)	(47.71)
Total	79,231	78,177	157,408
Panel B: $\widehat{\text{Applications}}_{jt}$			
Mean	6.23	6.55	7.82
Std. Dev.	(57.16)	(93.92)	(103.98)
Median	1	1	1
Total	185,225	191,948	376,893
Panel C: Estimated Win Rate, $\widehat{wr}_j$			
Mean	0.41	0.40	0.41
Std. Dev.	(0.43)	(0.42)	(0.42)
Panel D: Other Lottery Characteristics			
EIN Count	29,731	29,305	48,196
Prop. EINs with $\widehat{\text{Applications}}_{jt} = 1$	0.62	0.62	0.60
Prop. EINs with $\widehat{\text{Wins}}_{jt} = 0$	0.39	0.41	0.37
Prop. EINs with $\widehat{\text{Wins}}_{jt} = 1$	0.40	0.38	0.41
Prop. EINs with $\widehat{\text{Wins}}_{jt} > 1$	0.21	0.22	0.22

Note. EIN stands for Employer Identification Number. 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 generally what we mean when we reference firm j.

All told, $\widehat{wr}_j$ highlights some of the key benefits and drawbacks of our study, particularly in relation to Doran et al. (forth.). Because $\widehat{wr}_j$ covers the 2007 and 2008 lotteries, it reflects potentially random variation in immigrant hiring for a wider, more policy-relevant swath of firms than the lotteries from previous years. However, differences between $\widehat{wr}_j$ and wr_j, primarily due to differences between $\widehat{\text{Applications}}_{jt}$ and Applications_{jt}, raise questions about internal validity.

Our hope is that validating exercises presented in Section 5 and ongoing improvements to $\widehat{wr}_j$ thanks to improved data collection solidify the case that we can isolate random variation in access to H-1B workers from the larger 2007 and 2008 lotteries.

5 Empirical Analysis

The goal of our study is to to quantify how access to immigrant college graduate labor affects the performance of firms and their 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 hire immigrants. 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 different success in their H-1B hiring and estimate the true impact of immigrant hiring on firms.

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 towards finding a spurious impact of our lottery measure on firm performance. A second concern is that we cannot distinguish which petitions qualified for the special ADE cap. In 2007, for example, there was no lottery for the applications that were under ADE. Hence, we might be assigning a lottery win to some I-129 records that did not go through the lottery.

With these concerns in mind, we do not rely solely on the randomness of our win rate proxy to identify our coefficients of interest. Instead, we limit comparisons to firms that were likely subject to the same LCA application behavior, who had similar productivity shocks, and similar labor demand to isolate the true randomness of the lottery.

5.1 Identification Strategy

We begin by setting up an event-study approach described in Equation (2):

$$\log(y_{j,t+1}) = \sum_{\tau \neq 2006} \beta_\tau \left[\widehat{wr}_j \times \mathbb{1}(\tau = t)\right] + \Gamma X_{j,t} + \alpha_j + \alpha_{k(j),t} + \varepsilon_{j,t+1} \tag{2}$$

where j denotes the firm, t denotes a calendar year, and $k(j)$ denotes firm j's industry. Outcomes $\log(y_{j,t+1})$ include firm-level variables such as revenues, total employment, and employment of specific subgroups. We take logs and always add 1 to the outcome to ensure we keep

values where the outcome is zero. Outcomes are forwarded to reflect the lag between a calendar year's lottery and the fiscal year for which it is held: for example, winning applications from the 2007 would not manifest as new H-1B employment in our data set until 2008. The controls include a firm fixed effect (α_j), a 3-digit industry-by-time fixed effect ($\alpha_{k(j),t}$), and time-varying firm level controls ($X_{j,t}$), which primarily consist of a firm's log employment in 2007 (pre-lottery) interacted with time fixed effects.

Our key variables of interest are the event study variables $\widehat{wr}_j \times \mathbb{1}(\tau = t)$ which interact the combined 07-08 win rate for firm j, as defined in Equation (1), with time dummies. The event study dummies measure the differential response in outcome $y_{j,t+1}$ across firms with different values of the estimated win rate, in year t with respect to the omitted period 2006.

The event study setup helps us address concerns about the win rate proxy. First, we can make sure firms with different win rates were following similar trends before the lottery. Second, the rich set of fixed effects in the regression allow 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.

We limit the sample to firms that apply for at least one lottery-subject LCA in 2007 or 2008 and are present in our LEHD sample. In the end, we are left with 14,500 firms and 116,000 firm-year observations which we use throughout the analysis.[11]

5.2 H-1B Hiring

We begin the analysis by testing whether our measure of the lottery is consistent with the expected firm-level response in terms of I-129 records for new employment and employment renewal. A firm that is successful in the lottery, should have an increase in the observed I-129 records for new employment during the lottery years of 2007 and 2008. Similarly, successful firms are expected to have higher I-129 records for renewal 3 years later, in 2010 and 2011, assuming the H-1B workers remain employed at the firm.

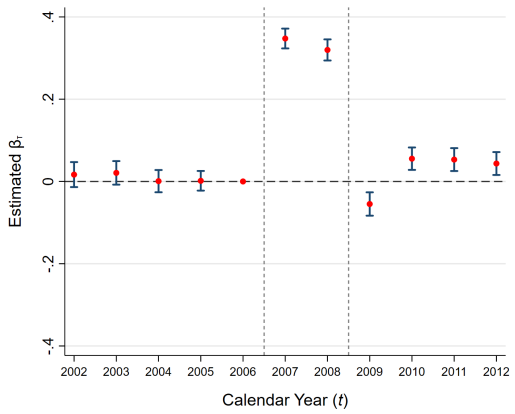
As shown in Figure 3a, there are no visible pre-trends regarding I-129 records between winning and losing firms. However, there is a clear spike in the observed I-129 records for new employment during the lottery years of 2007 and 2008. Interestingly, we see a negative and significant coefficient for year 2009, likely due to lottery-losing firms being more likely to apply for I-129s for new employment in 2009, when no lottery took place. Similarly, as shown in Figure 3b, winning firms are more likely to have I-129 records for employment renewal in 2010 and 2011,

[11] This restricts our attention to H-1B lottery firms that were operating in the 25 states for which we have LEHD information. This was done to limit disclosure avoidance review. In future drafts, we will also provide results using the full sample of lottery firms that we can match to the LBD, which adds thousands of firms to our analysis but can only be used to analyze employment and revenues, not firm composition.

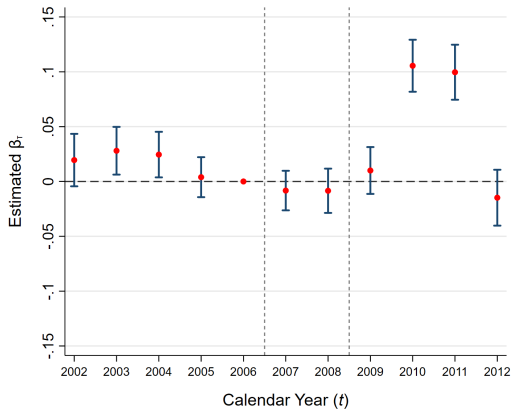
3 years after the lotteries of 2007 and 2008. Further, we also do not find an effect on renewals during the lottery years, which can be thought of as placebo outcomes.

Figure 3: Effect of Win Rate on Granted H-1B Petitions for Fiscal Year $t+1$

(a) New Employment



(b) Renewals



Note. We plot the estimated coefficients β_τ and their 95% confidence intervals. Vertical dashed lines bracket lottery years. Omitted period is 2006. Number of observations is 116,000. Number of firms is 14,500. Standard errors clustered at the firm level. The dependent variable in subfigure 3a is the log number of I-129 petitions for new employment at time t. The dependent variable in subfigure 3b is the log number of I-129 petitions for renewal of employment at time t. These data are from the USCIS. In both cases we add 1 to the outcome before taking logs to include observations with 0 value. Regressions include firm fixed effects, industry-time fixed effects, and log employment in 2006 interacted with time fixed effects.

We proceed by testing the ability of our LEHD data to pick up this "first stage." We should observe that winning firms become more likely to hire immigrants with characteristics that make them likely to H-1B visa holders than losing firms. This basic test is a key contribution of our project relative to prior literature that has used the H-1B lotteries (Dimmock et al., 2021; Doran et al., forth.; Glennon, 2020). Due to the lack of firm-level data on immigrant and native employment composition, prior literature was not able to confirm whether the H-1B lotteries indeed increased immigrant employment at the firm level. Hence, it was hard to establish that increased immigration was the predominant mechanism through which lottery success affected firm-level outcomes.

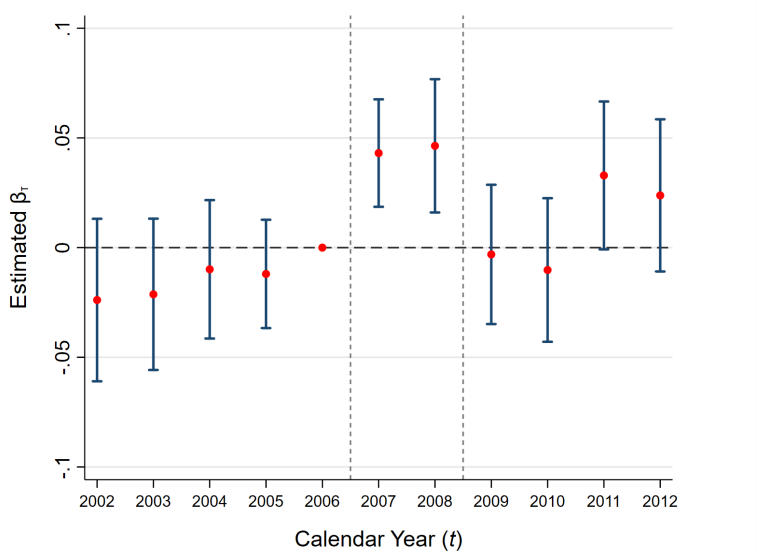
We define likely H-1B immigrants as those workers in the LEHD who were born outside of the United States and Puerto Rico, who have a college degree, who are between the ages of 25 and 40, and have less that 3 years of tenure at the firm. While the definition is likely to include some non-H-1B workers, it will almost surely include the firm's lottery-subject H-1B immigrant hires.[12]

As shown in Figure 4, there is a clear and significant increase in H-1B like hiring for winning firms relative to losing firms. The coefficient β_t measures the differential employment of H-1B

[12]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 not be subject to the 07-08 lotteries.

like immigrants in year $t+1$ across firms with different win rates. As an example, the 2007 coefficient implies that when comparing firms that win all of their applications (win rate of 1) with firms that lose all of their applications (win rate of 0), winning firms increase their H-1B-like immigrant employment by 12% in 2008. To put these numbers into context, each additional win adds, on average, 0.2 likely H-1B immigrants to the winning firm.

Figure 4: Effect of Win Rate on Employment of H-1B-Like Immigrants in Year $t+1$



Note. We plot the estimated coefficients β_τ and their 95% confidence intervals. Vertical dashed lines bracket lottery years. Omitted period is 2006. Number of observations is 116,000. Number of firms is 14,500. Standard errors clustered at the firm level. The dependent variable is the log number of H-1B-like immigrants employed at the firm at time $t+1$. This outcome is measured from the LEHD. We add 1 to the outcome before taking logs to include observations with 0 value. H-1B-like immigrants are defined as employees with less than 3 years of tenure at the firm, with a college degree, between the ages of 25 and 40, and born outside of the United States. Regressions include firm fixed effects, industry-time fixed effects, and log employment in 2007 interacted with time fixed effects.

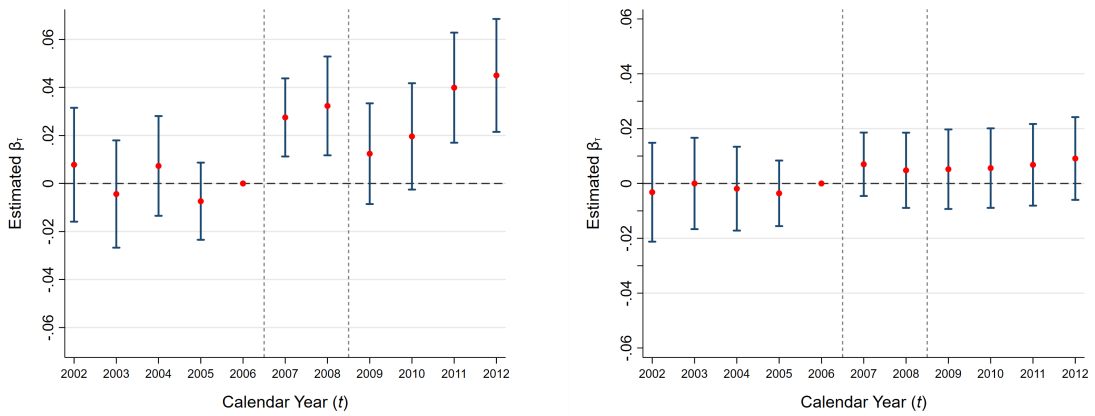
Interestingly, the effect becomes close to zero in 2010 and 2011 but rebounds the last two years of the sample. Part of this explanation could be networks, as the new immigrants the firm hires in 2008 and 2009 might be more likely to promote future immigrant hiring through better information and screening.

To further explore whether our lottery measure is likely capturing H-1B hiring, we split likely H-1B immigrants by nationality. As shown in Figure 5, we separately run our event study for Indian H-1B-like workers and Canadian/Mexican H-1B-like workers. 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 US under a TN employment status, which is a specific visa category established as part of NAFTA. Therefore, we should expect the hiring of Indians to primarily respond to the H-1B lottery and a lower response for Canadian and Mexican immigrants—indeed, this employment of this group can be thought of as a placebo outcome for our lottery measure. As shown in

Figures 5a and 5b, there is a clear and significant increase on H-1B like Indians during the lottery years. However, there is no significant increase for Mexicans and Canadian H-1B like workers.

Figure 5: Effect of Win Rate on Employment of H-1B-Like Immigrants in Year $t+1$, by Country of Birth

(a) Indians (b) Canadians/Mexicans



Note. We plot the estimated coefficients β_τ and their 95% confidence intervals. Vertical dashed lines bracket lottery years. Omitted period is 2006. Number of observations is 116,000. Number of firms is 14,500. Standard errors clustered at the firm level. The dependent variable is the log number of H-1B like immigrants employed at the firm at time $t+1$. We add 1 to the outcome before taking logs to include observations with 0 value. H-1B like immigrants are defined as employees with less than 3 years of tenure at the firm, with a college degree, between the ages of 25 and 40, and born outside of the United States. Sub-figure 5a only includes H-1B like workers born in India, while sub-figure 5b only includes H-1B like workers born in Canada or Mexico. Regressions include firm fixed effects, industry-time fixed effects, and log employment in 2007 interacted with time fixed effects.

5.3 Firm Performance

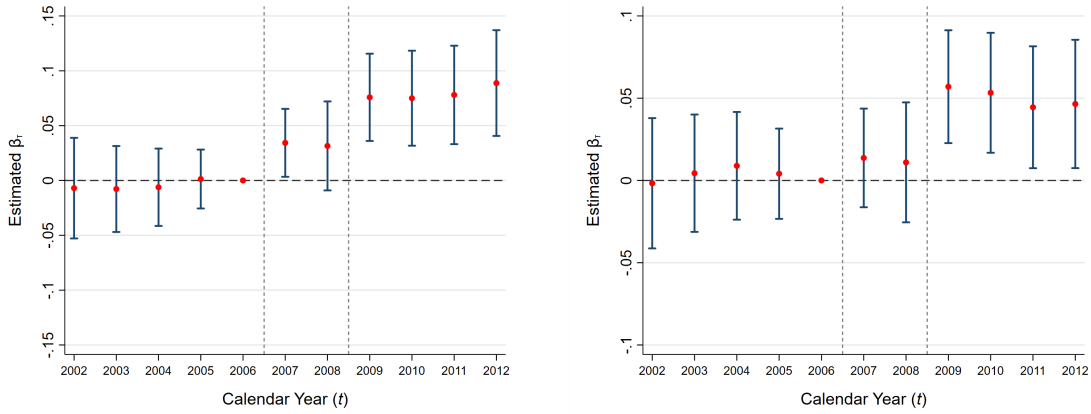
Having validated that lottery wins generate a "first stage"—an effect on both true and LEHD-measured, likely H-1B hiring—we next turn to the effect of lottery wins on firm revenues and labor productivity, measured as revenues per worker. If H-1B workers are uniquely important to the firms that apply for their services, we would expect to see improvements in firm performance—or, perhaps as accurately, deterioration in the performance of lottery losers. On the other hand, if H-1B workers are simply serving as (cheaper), near-perfect substitutes for otherwise-available native workers, access to H-1B workers may not impact revenues or labor productivity (as found in Doran et al., forth.).

Figure 6 displays the results from estimating Equation (2) using these outcomes. Figure 6a shows that lottery winners experience immediate increases in revenues upon employing H-1B workers. More strikingly, even after lottery winners' advantage in H-1B hiring initially subsides (recall Figure 4), lottery winners' revenue advantage persists. Four years after the lotteries, going from a win rate of 0 to 1 leads to a nearly 10% rise in revenues ($\hat{\beta}_{2012}$). The persistent, longer-run advantage is perhaps most clearly illustrated in Figure 6b, which shows

that a long-run labor productivity advantage due to lottery wins emerges *after* lottery losers have initially caught up with winners in terms of their H-1B workforce. Four years after the lotteries, going from a win rate of 0 to 1 leads to a nearly 5% rise in labor productivity, as measured by revenues per worker. These dynamics are consistent with previous literature that finds a relationship between H-1B workers and innovative activity, which often requires several years before producing revenue generating activity (see, e.g., Kerr and Lincoln, 2010; Peri et al., 2015b). This is a hypothesis we can test directly using United States Patenting and Trademark Office (USPTO) data in future drafts.

Figure 6: Effect of Win Rate on Firm Performance in Year $t+1$

(a) Revenues (b) Revenues per Worker



Note. We plot the estimated coefficients β_τ and their 95% confidence intervals. Vertical dashed lines bracket lottery years. Omitted period is 2006. Number of observations is 116,000. Number of firms is 14,500. Standard errors clustered at the firm level. The dependent variable in subfigure 6a is the log revenues at time $t+1$. The dependent variable in subfigure 6b is the log revenues per worker at time $t+1$. Regressions include firm fixed effects, industry-time fixed effects, and log employment in 2007 interacted with time fixed effects.

5.4 Workforce Composition

We next ask how the hiring of H-1B workers alters the worker composition of a firm, fully utilizing the unique data contained in the LEHD. Of first order importance is the canonical question of how much and whether 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 of 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 7. Figure 7a demonstrates that the temporary relative increase in H-1B workforce at lottery winners relative to losers also generates an increase in employment, While imprecisely estimated in later years, our results suggest that this increase in employment lasts beyond the lottery window.

Figure 7b does not find statistically significant evidence for substantial crowd out of "H-1B-like

natives"—who we define as workers who were born in the United States or Puerto Rico, who have a college degree, who are between the ages of 25 and 40, and have less that 3 years of tenure at the firm. However, all point estimates are negative after the lottery takes place, providing some suggestive evidence of crowd-out of native workers with the most potential substitutability with H-1B workers. As described in Section 6, we believe future drafts may deliver increased precision on these estimates.

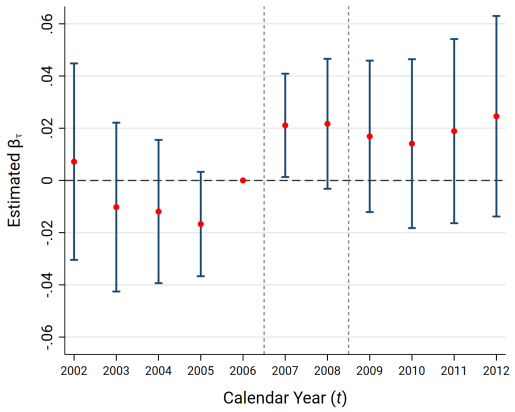
Figure 7c studies native workers with a college degree who are either over 40 years old, have more than 3 years of tenure at the firm, or both. These are workers who could be substitutable with H-1B workers based on their educational profiles, but may also complement H-1B workers because of their different experience profiles. Though our estimates are once again imprecise, we find stronger evidence for the latter notion.

Finally, we do not find strong evidence of either an increase or decrease in hiring of workers without a college degree. It appears that this group of workers are relatively insulated from the dynamics spawned by H-1B lottery wins and losses. We view this result as interesting on two fronts: 1) H-1B workers and non-college workers are not necessarily complements in hiring, and 2) a lack of effect on non-college workers cuts against the notion that the economically significant impacts on firm performance found in Section 5.3 are the result of omitted demand shocks that correlate with lottery wins.

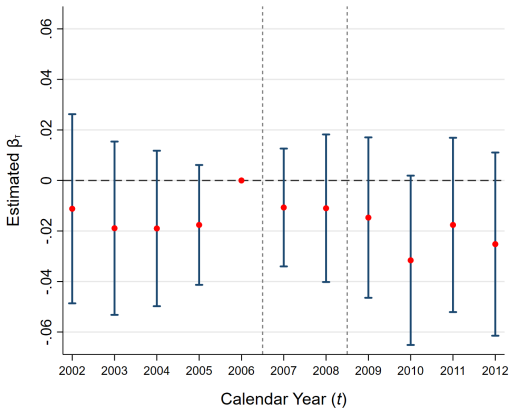
To put these numbers into context, each lottery win is expected to increase total employment at the firm by 1.38 employees in 2009, suggesting that immigrants have a net positive impact on firm employment. While imprecisely estimated, each lottery win is expected to decrease the number of H-1B-like natives by 0.06 workers, increase the number of non-H-1B like native college graduates by 0.24 workers, and increase the number of non-college graduates by 0.43 workers.

Figure 7: Effect of Win Rate on Employment in Year $t+1$, by Selected Subgroups

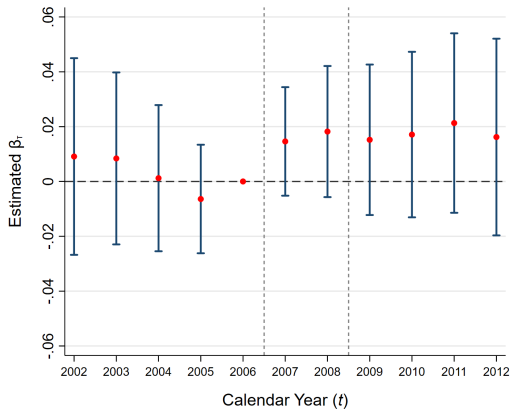
(a) Total Employment



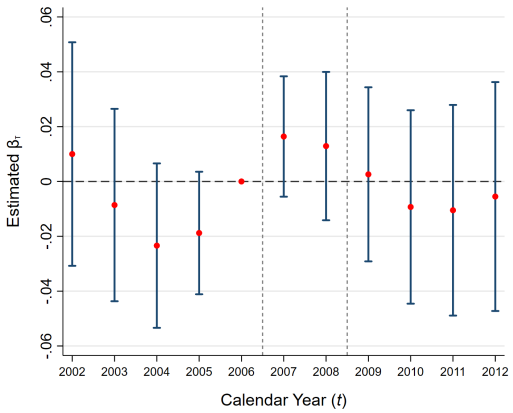
(b) H-1B-Like Natives



(c) Non-H-1B-Like, College Natives



(d) Non-College Workers



Note. We plot the estimated coefficients β_τ and their 95% confidence intervals. Vertical dashed lines bracket lottery years. Omitted period is 2006. Number of observations is 116,000. Number of firms is 14,500. Standard errors clustered at the firm level. The dependent variable in subfigure 7a is the log employment at time $t+1$. The dependent variable in subfigure 7b is the log of employment of H-1B like natives at time $t+1$. H-1B like natives are defined as workers born in the US, with less than 3 years of tenure at the firm, between the ages of 25-40, and with a college degree. The dependent variable in subfigure 7c is the log employment of non-H-1B like college natives at time $t+1$. These workers are employees with a college degree that do not classify as H-1B like workers. The dependent variable in subfigure 7d is the log of employment of H-1B like natives at time $t+1$. We add a 1 before taking logs to keep observations with 0 employment. Regressions include firm fixed effects, industry-time fixed effects, and log employment in 2007 interacted with time fixed effects.

6 Next Steps

All told, this paper provides several new, well-identified estimates to the empirical literature on skilled immigration to the US. Its 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 several ways.

First, recent advances in our data collection will generate several benefits in future drafts. We obtained a new I-129 data set that allows us to identify those I-129 records likely subject to the ADE cap because we now have infomation on each applicant's educational status when applying to the H-1B. This will allow to measure the lottery more precisely in 2007 by excluding ADE applicants. Similarly, as mentioned above, the new data set includes EINs, which are also present in the LBD. Thus, in many cases, we will be able to avoid using a fuzzy match and will instead be able to directly link our H-1B data to the LBD and LEHD. This should reduce measurement error in $\hat{wr}_j$.

Second, we plan to investigate whether our results on firm performance and composition differ by observable characteristics of the firm. Key dimensions of heterogeneity include firm size, skill composition, immigrant intensity, and average pay of the workers at the firm. In addition, we will extend our analysis to look into earnings responses at the firm level. This will help us evaluate the extent to which firms value H-1B workers as sources of lower labor costs.

Finally, in light of our firm-level results, 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 3:

$$y_{i,t+1} = \sum_{\tau \neq 2006} \beta_\tau \left[\widehat{wr}_{j(i)} \times \mathbb{1}(\tau = t) \right] + \Gamma X_{i,t} + \alpha_i + \varepsilon_{i,t+1} \tag{3}$$

where $j(i)$ indexes the firm that employed individual i in 2006.

These models provide the first direct evidence of how incumbent individuals' career paths are altered by exposure to H-1B workers. We estimate Equation (3)—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 workers presence 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., forth.; Kerr and Lincoln, 2010; Peri et al., 2015b).

References

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

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*, forth.

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.

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.

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 2 shows for each fiscal year, how long the H-1B cap took to distribute. The vertical axis shows days-in-filing, calculated as the number of days between the start of the application filing period (usually April 1st or 2nd), 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 be doing so around April, with start dates on or after October 1st, 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 noticably 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 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 1st, which usually corresponds to week 13 or 14 of the calendar year). The weeks go up until week 40, the week of October 1st and the start of the fiscal year.

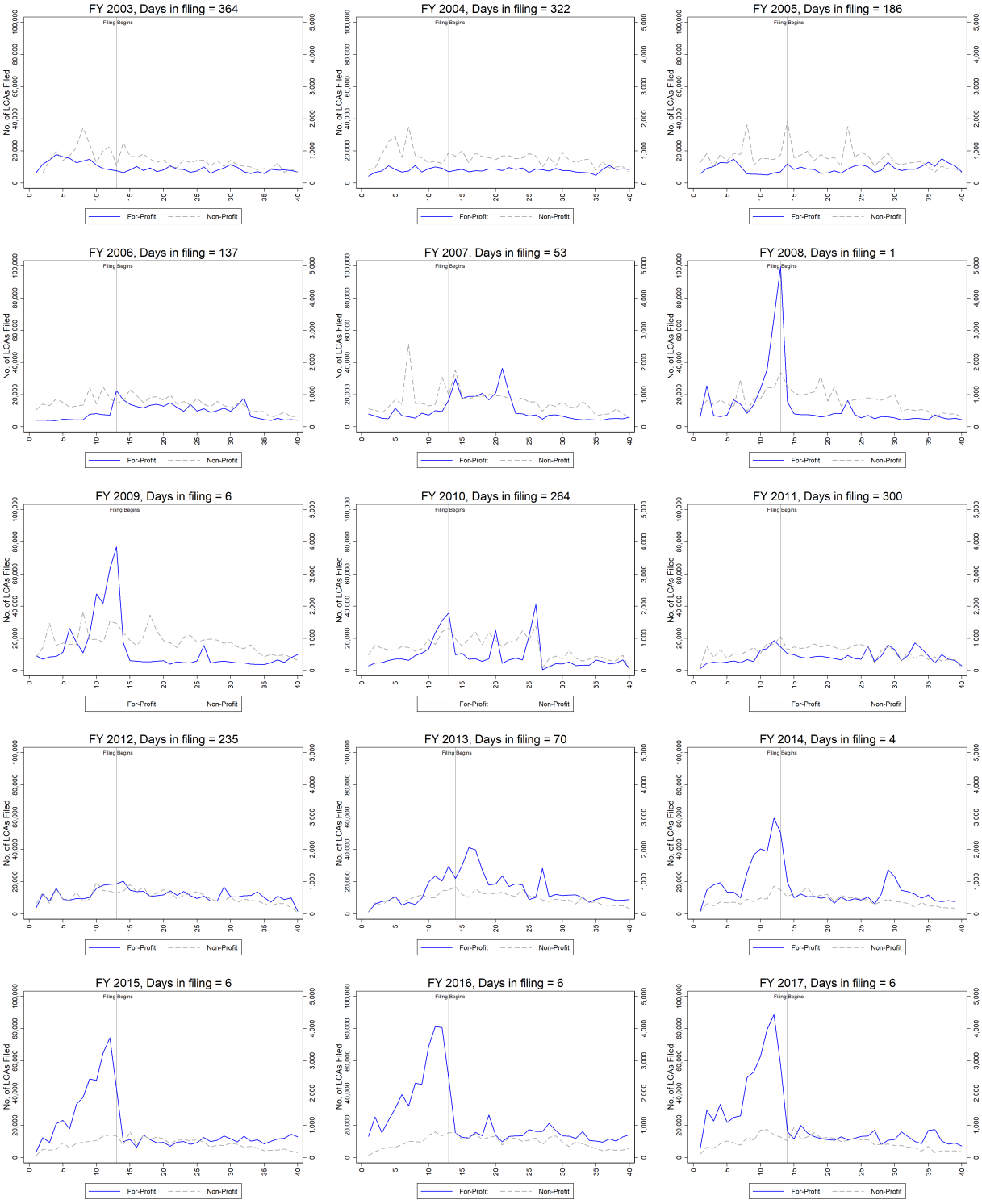
In times of low demand, when the H-1B cap takes a long time to distribute, we se roughly uniform LCA filing throughout the calendar year. As days-in-filing starts to decrease, we start to see a mass of LCA applications grow starting right around April 1st. FY 2008 and 2009 are clear that a huge mass of LCA applications are filed and approved not only on April 1st, but 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 1st. Since FY 2014, it is clear that predating is now the norm, most LCA applications appear to be filed before April 1st.

Such predating behavior presents at least two issues. First, this implies particular selection of firms into the lottery. In its simplest form, this means that firms who could anticipate the high demand and that a lottery might take place, ended up predating LCA applications to ensure they could submit a complete H-1B application by April 1st. 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 non-lottery 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 9. 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 1st.

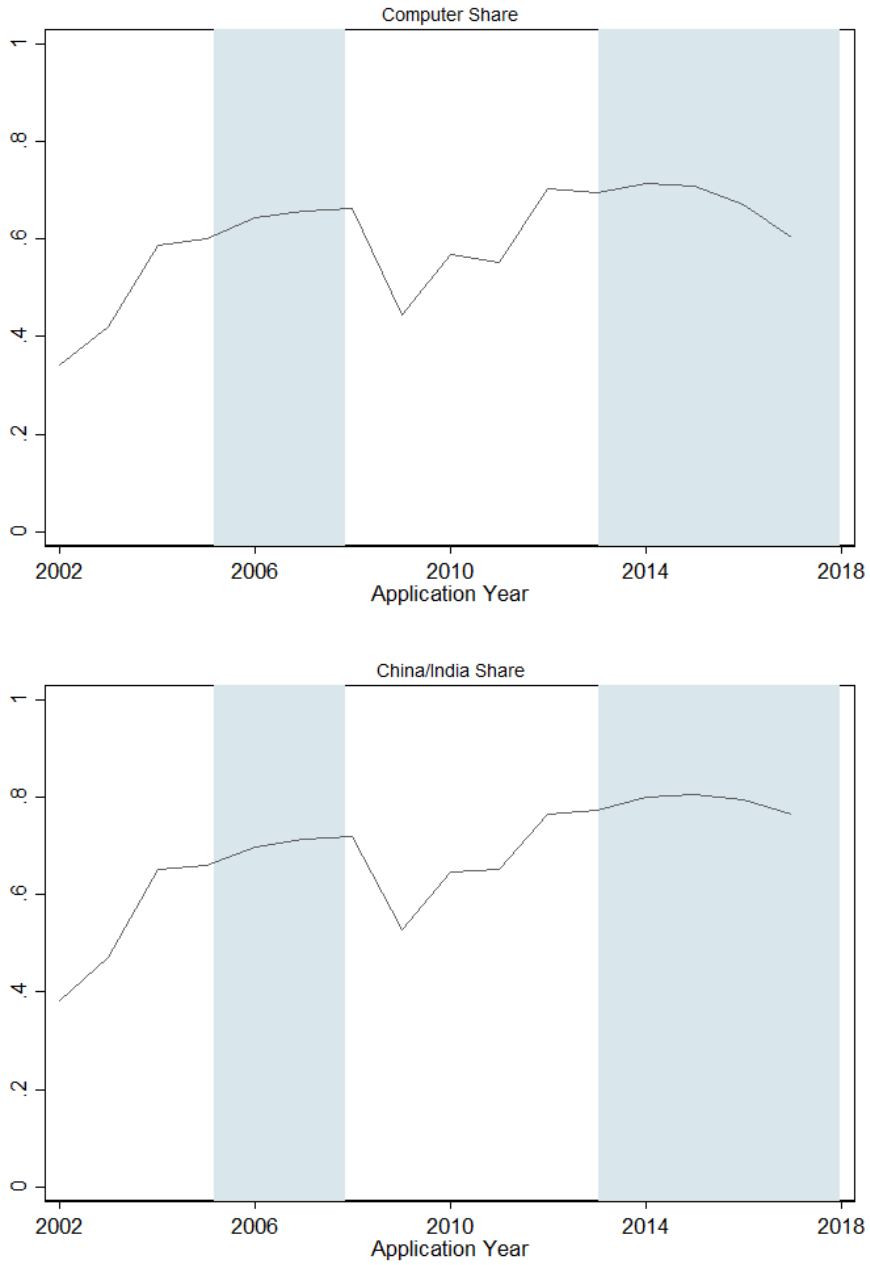
A second issue pertains to the internal validity of this natural experiment. Firms anticipating the lottery may be able to partially mitigate its impact. For example, a firm that anticipates a lottery and believes the odds to be equivalent to a coin flip can mitigate the lottery rationing by applying 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 8: LCA Application Filing by Week, FY 2003-2017



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

Figure 9: 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.