Looking for the "Best and Brightest":

Labor shortages and high-skilled foreign-workers

Morgan Raux [*]

December 2018

Abstract

The H-1B visa policy is at the core of the public debate on high-skilled migrations in the US. The controversy focuses especially on firms' motives to hire foreign workers. While firms claim that their vacancies would stay unfilled without recruiting foreign workers, the lack of data limits the investigation of this problem. In this paper, I address this gap by scraping job posting data from Indeed.com to measure labor shortages. I study the online duration of job ads to approximate firms' difficulties to fill their vacancies. Completing this analysis with administrative data, I observe for each job advertised on Indeed.com whether it leads to a demand for an H-1B visa. Exploiting variations at both extensive and intensive margins allows me to show that labor shortages explain a part of the demand for high-skilled foreign workers. By taking advantage of a discontinuity in H-1B visas application rules, I ensure the causal identification of this relationship.

KEYWORDS: H-1B Work Permit, Labor Shortages, Web Scraping.

JEL CLASSIFICATION: J61, J2, F22.

---

[*]Aix-Marseille University, CNRS, EHESS, Centrale Marseille, AMSE.

Mail: morgan.raux@univ-amu.fr

1   Introduction

Contrary to most developed countries, the public debate on migrations in the US focuses also on high-skilled foreign workers. More especially, most critics are concentrated towards the H-1B visas program as it represents the main gate through which high-skilled foreign workers enter the US on a temporary basis. As the number of visas is subject to a quota, the most debated question is whether it should be raised or decreased. To answer this question, all eyes are on the impact of this policy on the labor market outcomes of natives. Because this program relies on firm sponsorship, obtaining a visa necessitates first to obtain a job in a US company. The firm then proceed with the administrative process and pay a fee to make the hiring possible. As sponsorizing a foreign worker is costly[1], the public debate speculates on compensations obtained by firms. More especially, critics argue that big companies take advantage from these recruitements to cut wages. On the opposite, companies claim that their vacancies would stay unfilled without recruiting among the foreign labor force. In this paper, I investigate whether firms' demand for high-skilled foreign workers responds to labor shortages.

Deciding whether H-1B quotas should be raised or decreased depends on the impact of the program on natives' welfare. As summarized by Bound et al. (2018), this global impact takes into account both effects on innovation, profits and labor market outcomes. As the latter is the most debatted concern, the literature mainly tries to evaluate the impact of the program on labor market outcomes of natives.

To assess this effect, there is a large literature testing the degree of complementarity (respectively substitution) between native and foreign workers at different level of analysis. Exploiting variations across US cities, Kerr and Lincoln (2010) and Peri et al. (2015) reject the substitution hypothesis between native and foreign workers. By focusing on city level, both papers report average effects of the program. They capture hereby both direct and

---

[1]According to U.S. General Accounting Office (2011), hiring one H-1B worker costs between \$2,300 and \$7,500 dollars.

indirect effects on natives' employment outcomes. This average estimate first depends from the increase in labor competition for natives who are close substitute to H-1B workers. But it also relies on indirect employment effects resulting from aggregate changes in production and consumption due to the arrival of high-skilled foreign workers.

Another strand of the literature focuses on variations across firms to disentangle both channels. By assessing firms' employment responses to arrival of new H-1B workers, these studies get rid of the latter mechanism. Exploiting reforms in the quota of visas as natural experiments, a first set of papers report pieces of evidence supporting the complementarity hypothesis (Kerr et al., 2015, Mayda et al., 2016, Mayda et al., 2018). In addition, by taking advantage from very precise data on audit companies, Aobdia et al. (2018) provide us with a detailed description of hiring practices in line with this result. On the opposite, Doran et al. (2016) reject the complementarity hypothesis by exploiting a setting similar to a randomized experiment. In this paper, the authors focus on the marginal effect of hiring H-1B workers on employment growth. They take advantage from the lottery distributing visas once the cap has been reached. Exploiting variations across firms in their degree of luck, they show that winning firms do not experience a larger employment growth, contradicting hereby the complementarity scenario.

Exploiting similar random variations through the H-1B lottery, Peri et al. (2015) and Doran et al. (2016) report opposite conclusions. While the former present pieces of evidence consistent with complementarity between native and foreign workers, the latter reject this hypothesis in their analysis. Taking into account the presence of labor shortages offers a way to reconciliate both results. Considering the case where firms' expansion is constrained by the lack of workers can explain why winning firms in the lottery did not experienced a larger size growth. Simultaneously, this does not exclude the existence of complementarity at city level as it can take place between occupations and industries.

Therefore, focusing on the elasticity of substitution is not enough to evaluate the impact of the program. More especially, this elasticity should be considered within a broader reflexion

on firms' motives to hire foreign workers. Hiring cheap labor and filling labor shortages are usually opposed in the public debate to explain the demand of U.S. companies for H-1B visas. Disentangling both explanations, most of the literature focuses on the earnings difference between H-1B workers and their different counterparts. Testing for this wage difference, previous studies do not report any evidence about a wage penalty against new H-1B workers (Mithas and Lucas, 2010, Lofstrom and Hayes, 2011, Aobdia and Srivastava, 2018). Moreover, these results are consistent across industries and comparison groups. On the opposite, there is no direct nor causal evidence about the role of labor shortages in the demand of firms for H-1B foreign workers. Only Sparber (2015) reports indirect evidence about the existence of labor shortages. His identification strategy relies on the fact that for-profit firms cannot longer hire new H-1B foreign workers for the fiscal year once the 85,000 cap has been reached. As this limitation does not apply for non-profit firms, he focuses on a difference-in-difference estimator comparing the growth in wage offers between both groups. Reporting a larger wage growth in for-proft firms after the cap is reached, Sparber interprets its results according to market forces. More especially, he argues that such differences should not appear if natives were readily available on labor markets to fill these positions. Hereby, he reports indirect evidence of the role of labor shortages in the demand for high-skilled foreign workers.

Surprisingly, there is no direct evidence about the relationship between labor shortages and inflows of high-skilled foreign workers. This represents an importang gap in the literature for two reasons. While the existence of labor shortages can theoretically reconciliate different results of the literature, economists do not have a clear vision of this phenomenon. Therefore, reporting evidence on this relationship matters from a research perspective. This is also important from a political point of view. Since the role of labor shortages in demand for high-skilled foreign workers is constantly highlighted by tech companies that lobby the administration to expand visas quota, assessing this link would inform policy makers. The main explanation to this gap in the literature relies on scarcity of the data. According to

Barnow et al. (2013), usual survey data capture information that are too aggregated to assess labor shortages at the skill level. Its most striking example is the Job Openings and Labor Turnover Survey (JOLTS) that only delivers information at the industry level.

In this paper, I address this challenge by collecting the complete set of U.S. job postings advertised on Indeed.com during February and March 2018. I focus on the advertisement duration of these job postings to proxy firms' difficulties to fill their vacancies. I rely hereby on a measure of the labor shortage experienced by the company itself. To assess the link with the demand for high-skilled foreign workers, I match this information with administrative data on demand for H-1B visas. More especially, I first look whether each job advertised on Indeed.com leads to a demand for a H-1B visa. I exploit hereby variations in ads duration to explain the likelihood of demanding a H-1B visa. By comparing similar jobs within a same commuting zone, my results are twofold. At the extensive margin, I find that jobs that are still advertised by the beginning of the visa application period are more likely to lead to a H-1B demand. At the intensive margin, longer job posting durations are associated with a larger probability to fill a visa application. Finally, I provide additional evidence that job postings duration truly reflect firms' difficulties to fill their vacancies.

By scraping over time the complete set of job postings published on Indeed.com, I propose a new method to assess the role of labor shortages in the demand for high-skilled foreign workers. First, these data allow me to address the aggregation issue highlighted by Barnow et al. (2013). By construction, this study takes jobs as unit of analysis, where jobs are identified by their title, company and location. On the other hand, administrative data also report H-1B demands at the job level. Because these visas are worker and job specific, companies first have to fill a Labor Condition Application (LCA) ensuring that the job is eligible. Focusing on the universe of jobs advertised on Indeed.com, I identify whether they are also included in the list of H-1B LCAs.

Moreover, matching both sources of data necessitates to take into account the specific timing of the program. For the last few years, the demand for visas largely exceeded the

85,000 quota imposed to for-profit firms. As visas are delivered on a first-come first-served basis, companies are now aware of the need to apply by the first week of April to get a chance of obtaining one of them. As a consequence, I focus especially on job postings advertised online until the first week of April 2018. Hereby, I target jobs that are, a priori, still vacant at the beginning of the visa application period.

To investigate the causal role of labor shortages on firms' demand for foreign workers, I assess whether online duration of ads is a good predictor of the demand for H-1B visas. More especially, I explore this relationship at both extensive and intensive margins. I first exploit variations between job postings that last or disappeared from Indeed.com by the beginning of April 2018. I show that jobs that are still advertised by this date are more likely to give rise to H-1B demand. I interpret this result as a first piece of evidence to say that firms' difficulties to fill their vacancies play a significant role in the demand for high-skilled foreign workers. In addition, I explore further the relationship between the actual duration of ads and H-1B demand. Taking advantage from the longitudinal feature of my data collected online, I show that jobs that are advertised during a longer period of time are more likely to lead to H-1B applications. I interpret this result as an additional piece of evidence in line with firms' recruiting difficulties. By gathering evidence at both extensive and intensive margins, I show that labor shortages, measured by firms' difficulties to recruit, play a significant role in the demand of H-1B foreign workers.

Finally, I test whether online duration of ads truly reflect firms' difficulties to fill their vacancies. This test takes advantage of a discontinuity in the H-1B visa program. To prevent firms from using H-1B visas to cut wages, the administration requires from companies to prove their commitment to the law. More especially, they have to prove that they first failed in recruiting a native worker before to apply for a visa. Usually employers provide the administration with the bill of their job posting to meet this requirement. Hereby, this feature threatens my identification strategy. More especially, the positive correlation between ads duration and demand for foreign workers could result from an alternative scenario where

6

employers keep voluntarily their job posting online to get the right to apply for H-1B visas. To address this issue and ensure the validity of my identification strategy, I take advantage of the fact that jobs whose annual wage exceed 60,000 dollars are not subject to this requirement. Similarly to a regression discontinuity design, I compare job posting durations on both sides of the threshold. By rejecting the presence of significant difference between both groups, I ensure that ads duration is a relevant proxy for labor shortages. Consequently, it supports the causal effect of labor shortages on the demand for H-1B foreign workers.

While most of the literature focuses on the supply side of foreign workers to evaluate the labor market impact of the H-1B program, I take the opposite view. More especially, I build on the specificity of the H-1B program: firm sponsorship. I try to understand the reasons of firms for recruiting within the foreign labor force. I show hereby that rejecting the complementarity between native and foreign workers is not sufficient to conclude about a displacement effect against natives.

The next section describes the H-1B visas program and the public debate on high-skilled migrations. Section 3 presents the job posting data collected online alongside with administrative data on H-1B demands. I pay special attention to both matching and timing dimensions. Section 4 details the identification strategy. I present the resuts in section 5 and details robustness checks in section 6. Section 7 concludes.

2 The H-1B program in the public debate

H-1B visas allow employers to employ high-skilled foreign workers on a temporary basis. Targeting high-skilled workers, this program is restricted to individuals graduated at least from a bachelor degree. They are considered as "non-immigrants" as the duration of the visa lasts for 3 years and can be renewed only once. In this section, I describe the specificities and the timing of H-1B visas. I conclude by detailling the arguments used in the public debate on this question.

Because H-1B visas represent the main channel through which high-skilled foreign workers enter the United States, the relevance and the scale of the progam is largely debatted. For the last few years, approximately 180,000 entered the country through this program. Recipients of H-1B visas must be distinguished according to their employer. While there is no restrictions for individuals coming to work for a non-profit company, the number of visas delivered to for-profit firms workers is limited to 85,000 among which 20,000 are restricted to foreign born individuals graduated from U.S. universities.

The specificity of the H-1B program relies on firm sponsorship. This imply that visas are job and worker specific. To obtain one of them, foreign workers first have to find a job in a U.S. company. The employer then have to cope with the administrative procedure. This process is made of two steps. They first have to submit a Labor Condition Application (LCA) to the U.S. Department of Labor. This form intends to make sure that the job meets the legal requirements, namely that H-1B workers will benefit from equivalent wage and working conditions than other workers similarly employed. Then, only approved applications can actually compete to obtain a H-1B visa. This second step takes place by filling a petition for a non-immigrant worker, namely the form I-129.

Since 2014, the number of petitions sent during the first week of the application period has exceeded the H-1B quota. In 2015, the U.S Citizenship and Immigration Services (USCIS) received 172,500 cap based petitions while the quota was 85,000. To address this issue, Immigration services changed their way to allocate visas. More especially, they replaced their rule based on a first-come first-served basis by a lottery taking into account all applications received during this first week of April. Delivered for each fiscal year, H-1B visas represent working authorizations from October onwards. The application period opens 6 month before, on April 1st. Since 2014, companies are now aware that they have to apply during the first week of April to get enrolled in the lottery that will give them a chance to obtain a visa. The lottery then takes place by mid-April and employers are notified whether their petition is accepted or denied.

Because the demand for H-1B visas largely exceeds the supply allowed by this cap, the quota takes an important place in public discussions on this topic. On the one hand, big companies invest in lobbying activities to convince policy makers to increase the cap. Their main argument relies on labor shortages. Facing difficulties to fill some positions with native workers, they present H-1B visas as a solution to address this issue. Recruiting worlwide allow them to compensate skill mismatches that take place on the U.S. labor market. But on the opposite, these firms are accused to take advantage of foreign workers to reduce their labor cost. More especially, critics argue that their recruitement policy create displacement effects hurting native workers. To inform this debate, I document the role of labor shortages in firms' demand for high-skilled foreign workers through H-1B visas.
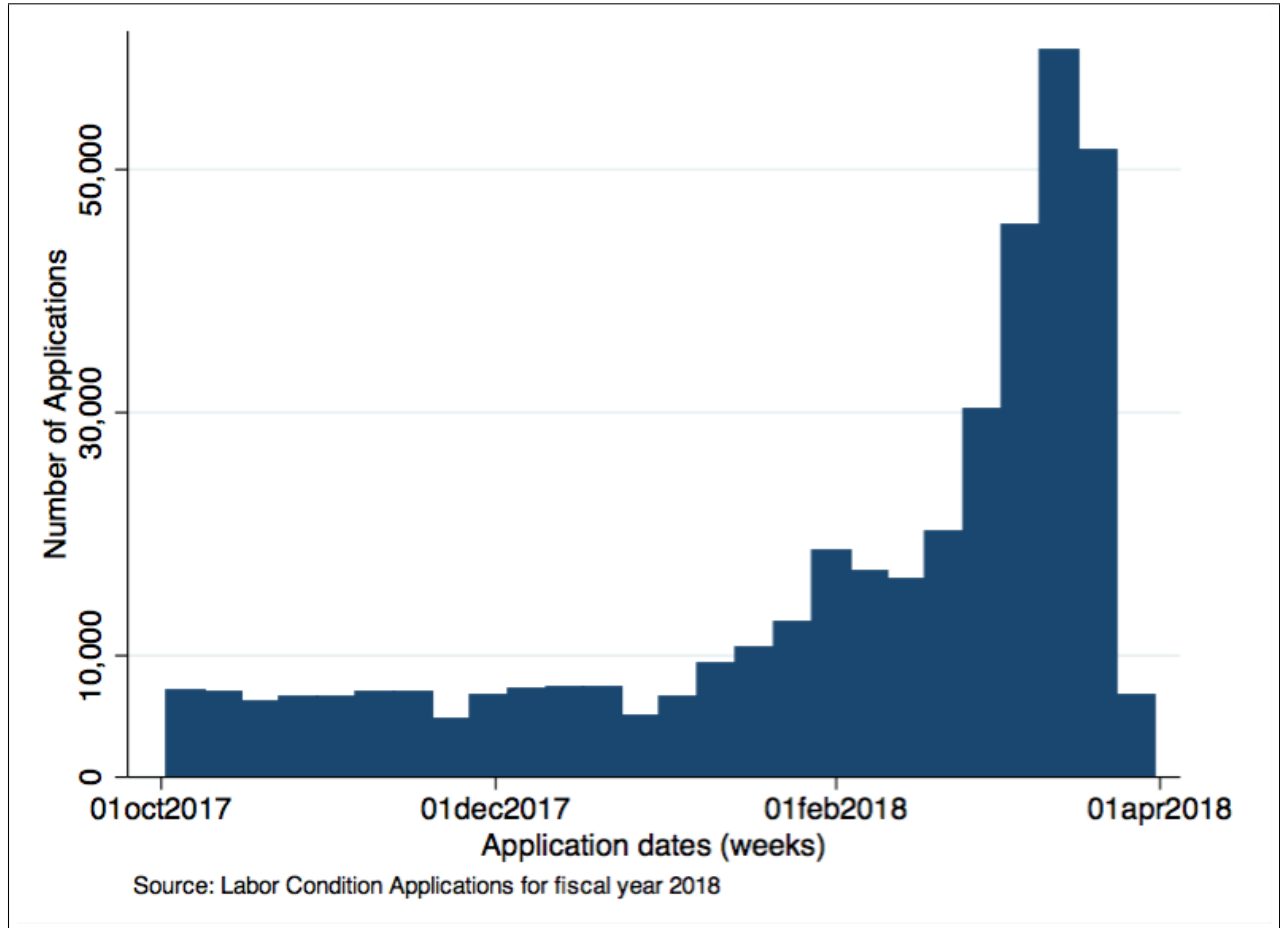
3   Data

In this section, I present both administrative and web data on which the analysis focuses. The paper builds on the administrative data released by the immigration Administration on firms' demands for H-1B visas. The section starts by presenting this first source of information. Moreover, the contribution of this paper mostly relies on a new source of data that allows me to exploit variations in labor shortages at the job level. I pay special attention to detail job postings information collected online. Eventually, this section concludes on the description of the merging procedure delivering the final data used in the analysis.

3.1   Observing H-1B demands via Labor Condition Applications

This paper first relies on demands made by firms for H-1B visas. I observe this demand through the exhaustive list of Labor Condition Applications sent by U.S. companies to the Department of Labor. Here, I first detail the information embodied in these forms. Finally, I describe both firms and jobs represented among H-1B demands.
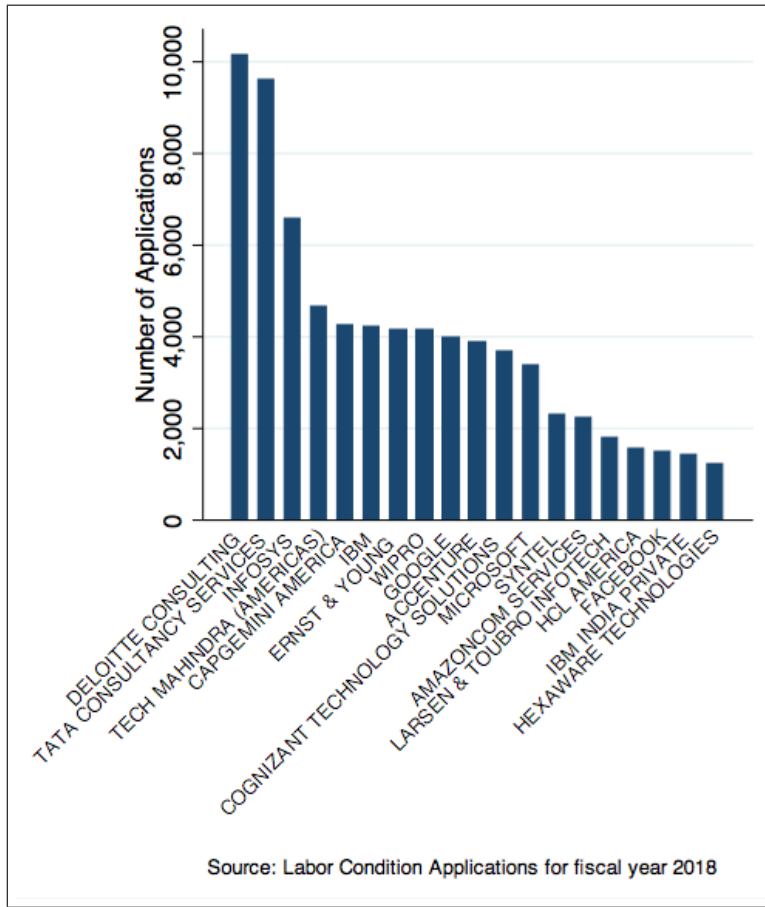
Observing Labor Condition Applications received by the U.S. Department of Labor gives a sense of the demand for high-skilled foreign workers. On the one hand, it is important to note that filling this form does not correspond to the official demand for a H-1B visa. All

Figure 1: Distribution of Labor Condition Applications' submission date

employers submitting a LCA do not necessarily fill the final petition. But on the other hand, it represents a first necessary step in this process. Therefore, focusing on Labor Condition Applications rather than on actual petitions (observed through the form I-129) allows me to observe the expected demand for H-1B visas until March 2018. Because employers subject to the quota need to submit their I-129 form by the beginning of April, their LCA must be sent by the mid/end of March. Even if most of them are waiting for the last moment to meet this requirement, some employers do send their application over the year. By focusing on these data, I take these observations into account regardless of whether they send the I-129 form in April.

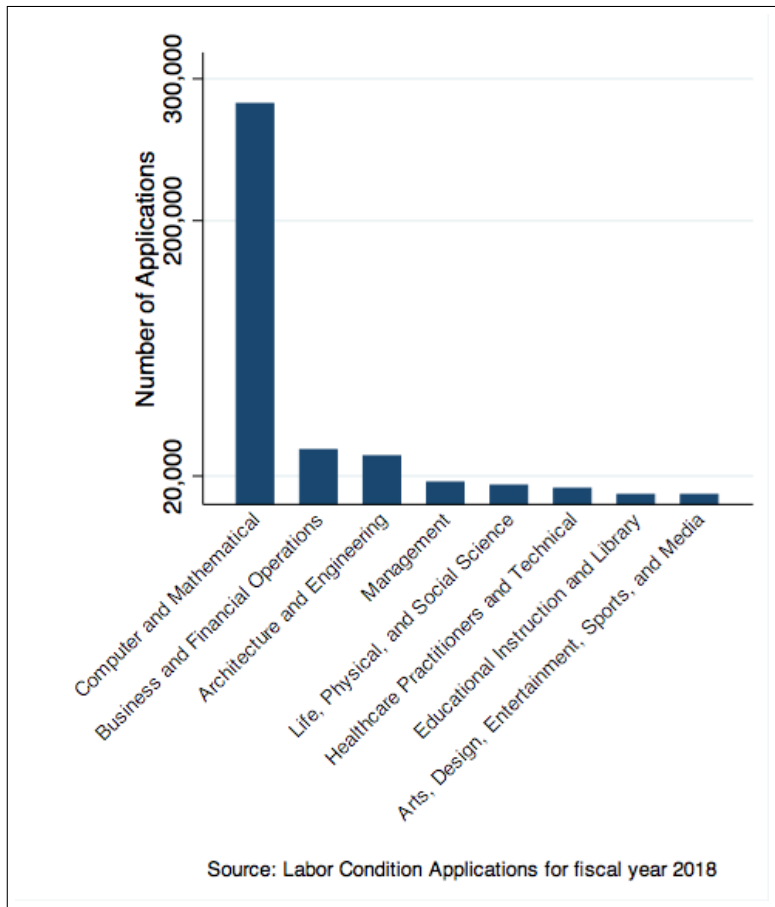Figure 2: Distribution of employers in Labor Condition Applications

Source: Labor Condition Applications for fiscal year 2018

Notes: This graph presents the distribution of largest employers gathering 20% of LCAs. (Reading) Deloitte Consulting sent more than 10,000 LCAs to the Department of Labor from October 2017 to March 2018.

Studying LCA submission dates provides me with new insights about the timing of firms recruitments. Employers who face the H-1B quota can submit their LCA from October to the end of March. Figure 1 describes submission dates over the fiscal year 2018. The distribution is largely skewed towards the end of the period. While the number of applications never exceeds 10,000 per weeks until the end of January, it rises to 60,000 during the second week of March. It suggests hereby that most of employers are waiting for the last moment before starting the administrative process to hire foreign workers.

LCA data enable me to identify visa demands at the job level. Labor Condition Applications are used to ensure that jobs meet the official requirements. Therefore, it gathers

enough information to identify precisely the position for which the company needs a visa. More especially, vacancies are identified by their job title, company name and precise location. Since LCA are required by the Administration to make sure that H-1B workers are paid similarly to their native counterparts, these data include information on the prevailling wage associated with the job. However it worth noting that they do no gather any information on the worker himself. Therefore, the analysis really focuses on the demand side, namely the firms. The main advantage of LCA relies on the precision of information concerning the latter. More especially, employers are identified by their name and are not anonymised. This features delivers a precise idea about the main petitioners for H-1B visas.
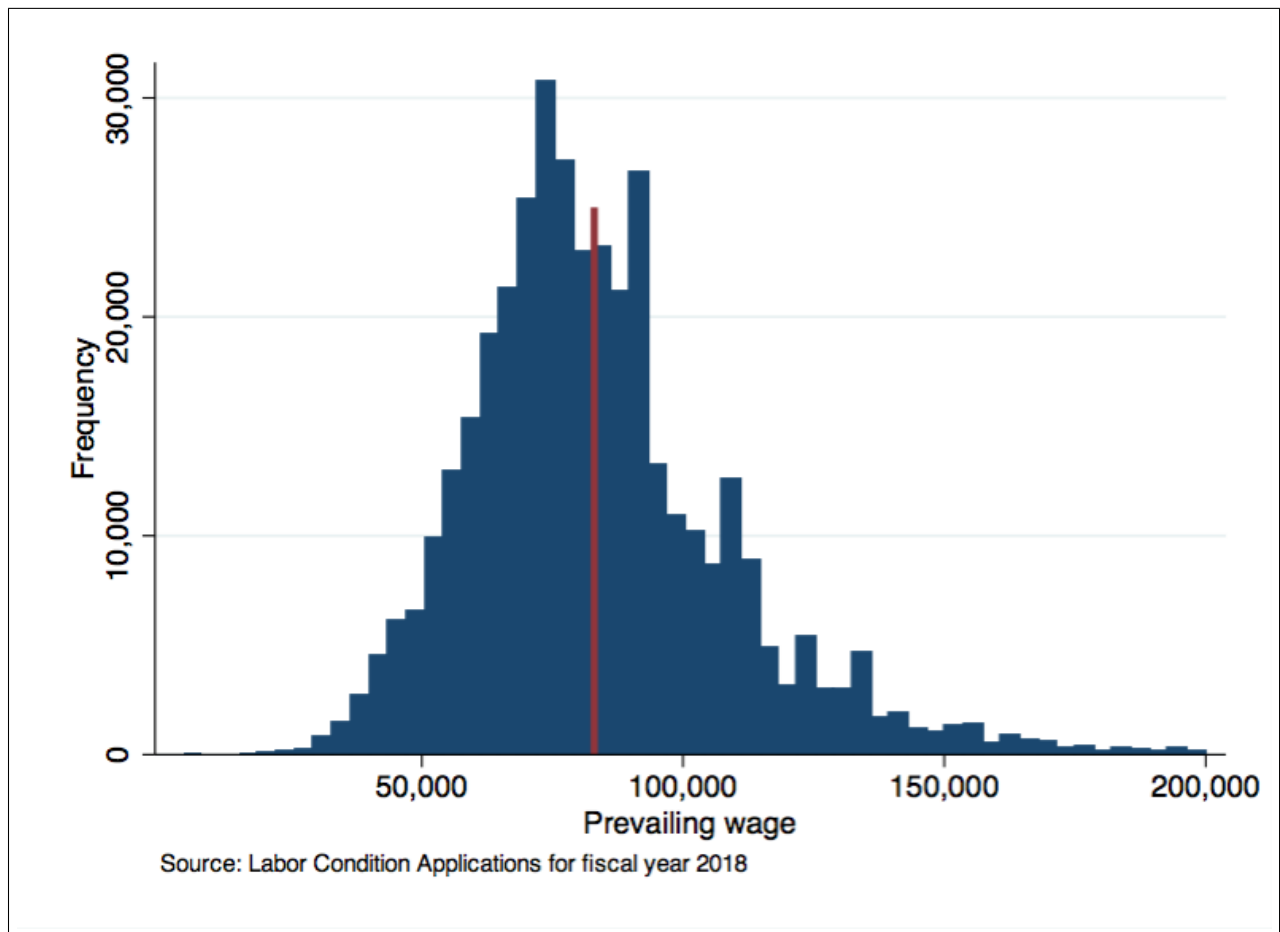
Figure 3: Distribution of SOC codes in Labor Condition Applications

Source: Labor Condition Applications for fiscal year 2018

Notes: This graph gathers 99% of LCAs submitted between October 2017 and March 2018. (Reading) Almost 300,000 LCAs sent during this period are related to a computer and mathematical job (i.e. STEM).

Big companies who lobby the administration to increase the H-1B quota are also the main petitioners to Labor Condition Applications. Figure 2 depicts the distribution of employers by number of LCAs. It worth noting that 20% of applications are gathered by 19 companies. Two sectors of activities are represented in this list: new technologies and the audit industry. Among these top 19 employers, the number of applications is largely skewed towards three big companies, namely Deloitte Consulting, Tata Consultancy Services and Infosys. Again, the conversion of LCA into H-1B petition varies from one employer to another. Nevertheless, it gives a broad idea about relative use and needs of H-1B visas.

Figure 4: Distribution of prevailing wage in Labor Condition Applications

Source: Labor Condition Applications for fiscal year 2018

Notes: The vertical red line represents the average of annual prevailing wages.

While most applications are splited between two industries, most of them are concentrated among STEM jobs. More especially, 70% of Labor Condition Applications are submitted for Computer and Mathematical occupations. Figure 3 represents this distribution by SOC code. This graph highlights the large bias in H-1B demand towards STEM jobs. It is therefore consistent with the focus of the public debate on H-1B visas as both proponents and opponents focus principally on the situation of STEM workers in tech companies.

Finally, prevailing wages in LCAs provides me with a proxy for the quality of workers that firms are looking for. Figure 4 presents the yearly wage distribution. Half of the applications refer to jobs paid from 40,000 to 80,000 dollars per year. On the right part of the distribution, wages are much more dispersed. This latter observation supports the idea that employers are looking for top skills in the foreign labor force. However, focusing on the wage dimension may rise doubt about the level of skills really needed for jobs located in the lower part of the distribution.

3.2 Measuring hiring difficulties on Indeed.com

Internet data on job postings provide me with an alternative to proxy labor shortages. In this section I present the data collected on Indeed.com. I first describe the data collection procedure and pay special attention to its timing. More especially, I detail the link made between job postings observed online and the H-1B application process. After describing job postings duration, I explain how it allows me to study the relationship between labor shortages and firms' demand for high-skilled foreign workers. Finally, I list the advantages of focusing on Indeed.com for the study of H-1B visas.
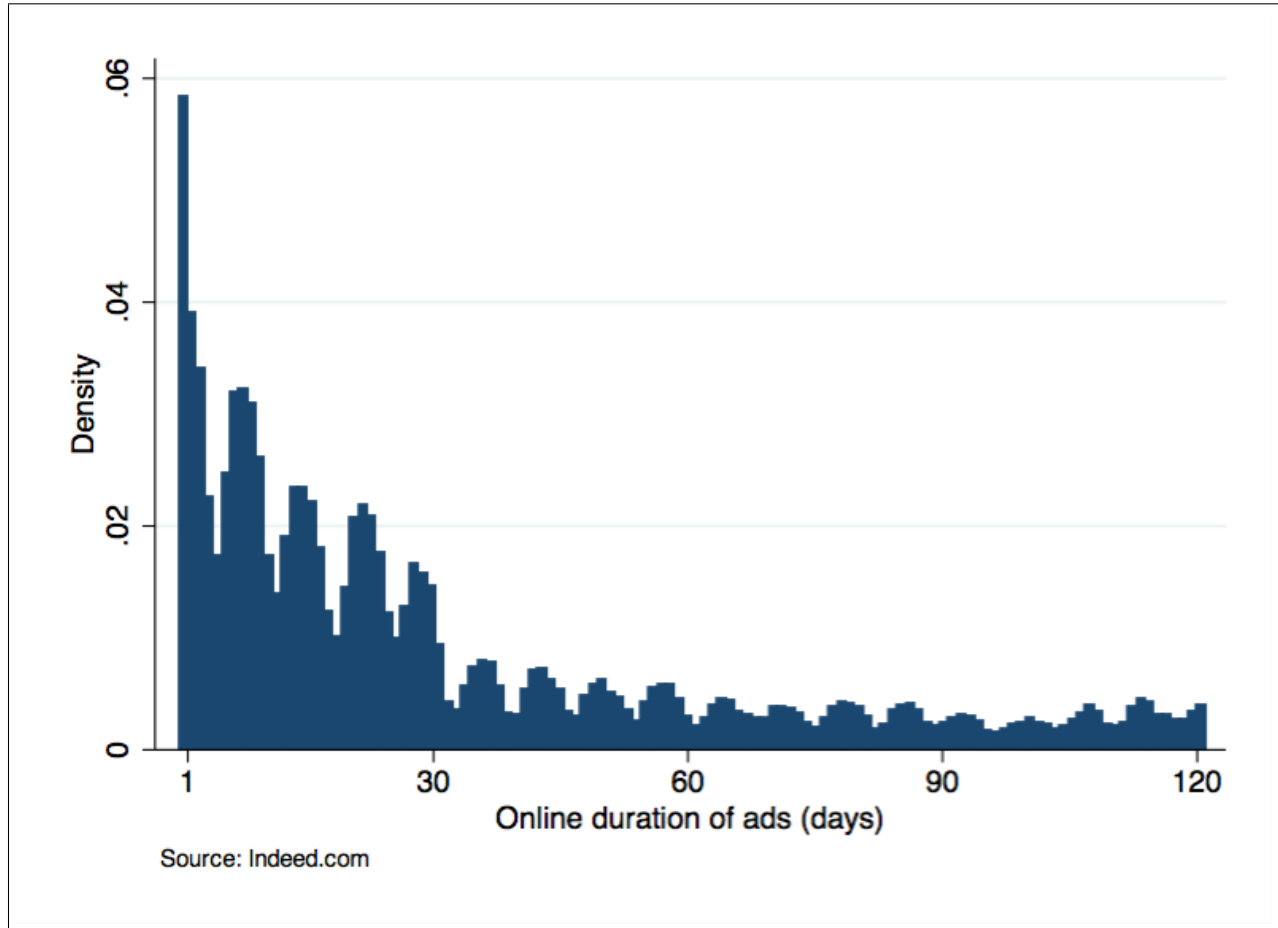
To understand how labor shortages affect the demand for H-1B visas, I focus on the exhaustive set of vacancies advertised on Indeed.com before April 2018. Focusing on this panel of ads provides me with a precise description of the demand on the U.S. labor market. Since actual applications for H-1B visas take place in the beginning of April, it necessitates to measure labor market forces at the same period to understand the relationship between both phenomenons. I have collected four times the entire portfolio of vacancies published on

Indeed.com. By scraping data every two weeks from February 15 to March 31, I have created a sample of ads lasting from November 2017 to March 2018. This allows me to observe the evolution of the demand for high-skilled jobs in the U.S. Eventually, I investigate how it translates into a specific demand for foreign workers.

By taking advantage from new data collection techniques, I address the aggregation problem raised by Barnow et al. (2013). Compared to usual surveys, internet data allows me to observe more precisely labor markets. I exploit variations in the duration of job ads to approach labor shortages. More especially, the time that job postings stay online partly reflects firms' difficulties to fill their vacancies. Publishing vacancies online is one of the main channels through which employers are searching for workers. Longer durations reflect hereby a longer period of search. Therefore, it can be interpreted as a proxy of the excess of demand for this particular type of skills. Respectively, between-jobs variations reflect the relative exposition to labor shortages.

Studying duration of ads on Indeed.com provides me with insights on variations on which relies the following analysis. Figure 5 describes the distribution of ads duration in the sample. First, it worth noting that the distribution is right truncated. Because Indeed systematically removes ads after 120 days spent online, I do not directly observe between ads variations exceeding this cap. This affects directly the composition of my sample. Since the data collection procedure started on February 15, earliest ads gathered in the data have been posted on November 2017. Among these ads, the distribution lies in two distinct parts. During the first month of publication, most of the job postings disappear from the website. On the other hand, older job ads mostly last for four months before being removed by Indeed. Measuring firms' recruitment difficulties mostly relies on these variations. Therefore, the credibility of the analysis relies on the accuracy of this duration information. Focusing on Indeed.com supports this identifying assumption. Because this specific job board collects 90% of its ads by scraping employers carreer pages, it maximizes this accuracy. More especially, it takes advantage of the fact that employers are more likely to remove directly their ad

Figure 5: Distribution of ads duration on Indeed.com

from their own website once their vacancy is filled. Therefore, observing duration of ads on Indeed.com provides me with the most accurate information on firms' difficulties to fill their vacancies.

Indeed.com presents an accurate description of the high-skilled labor market in the US. First, it gathers more than three millions job postings at any point in time and represents hereby the largest job board in the country. In addition, it mostly focuses on big companies that are also the main petitioners for H-1B visas. This represents a first argument to use Indeed data while studying high-skilled migrations. This overrepresentation of large firms is due to the platform's strategy. To gather the largest spectrum of job postings, Indeed.com does not only rely on ads directly posted by employers. In addition, they scrape career

webpages of big companies to extend their portfolio of vacancies. Because high-skilled jobs are overrepresented in big companies, they compose also the largest part of the job board.

3.3 Merging labor shortages to H-1B demand

Finally, I investigate whether each job advertised on Indeed ended up by giving rise to a demand for a H-1B visa. By focusing on the job level, I can identify those that are included in both data sources. More especially, their job title, company and location allows me to identify these pairs. In the following analysis, I explore the determinants explaining the presence of jobs among H-1B petitioners. In this section, I first present the final sample obtained from the merge of both Indeed data and Labor Condition Applications.

Merging both data relies on the three main dimensions identifying a job. It first necessitates to identify its location. Focusing on city level, this first step is relatively straightforward as city names are uniformely reported in both datasets. On the opposite, identifying similar companies across sources is more complicated. Because the name of a same company can slightly vary between an online job posting and an admistrative form sent to the Department of Labor, the matching becomes more difficult. To adress this issue, I adapt the algorithm to take into account small variations in company names and identify similarly "Facebook", "Facebook inc." and "Facebook Incorporation". Finally, the biggest challenge relies in the job title dimension. Merging administrative and internet data together is quite challenging as jobs can be called very differently across both sources. For instance, an employer looking for a "Java Developer" on Indeed.com, could indicate "IT Engineer" on its corresponding LCA. This feature leads to limit the number of matches identified across data sources.

Applying restrictive criteria ensures that job postings matched to LCA really identify a same job but limits drastically the number of matches across samples. More especially, I focus on a sample of 3,500 jobs that are both advertised on Indeed.com and associated with a Labor Condition Application. To test the labor shortage hypothesis, the following analysis exploits between jobs variations within a same local labor market. The latter is defined by the association between a commuting zone and a type of job following the Standard Occu-

17

pation Classification System at the two digits level. Focusing on similar jobs on Indeed.com that are not associated with a LCA, my comparison group gathers 200,280 observations. It worth noting that this group might include jobs associated with a LCA but not identified by the matching algorithm. Assuming that the probability of missing a match is uniformly distributed along the duration distribution ensures the validity of the identification strategy

To assess the role of labor shortages in the demand for high-skilled foreign workers, I take advantage of two key pieces of information in the data. First, the analysis exploits the correlation between ads removal and LCA submission. Hereby, it focuses on the extensive margin of the relationship. More especially, it compares the proportion of jobs that disappeared from Indeed.com by LCA status. Table 1 summarizes the average difference. Gross statistics highlight that jobs associated with a LCA are more likely to get a job posting still online by the end of March. Secondly, the analysis focuses on the correlation between online duration of job postings and their probability to submit a LCA. It explores hereby the intensive margin of this link. The second row of table 1 compares this duration between both LCA groups. It shows that jobs associated with a LCA last in average three more days than their comparison group. While table 1 only describes gross statistics, I continue this analysis in section 5.

Table 1: Final sample: descriptive statistics

	All jobs on Indeed.com	Jobs looking for an H-1B visa	Jobs not looking for an H-1B visa
Observations	203,789	3,510	200,280
Mean duration (days)	31.76	34.82	31.73
Percentage of jobs Still Advertized (SA)	33.09%	58.58%	32.80%

Notes: This table describes the final sample gathering Indeed and LCA data. It compares average statistics between jobs associated to a LCA and the others.

4 Identification strategy

This section presents the identification strategy to assess the role of labor shortages on the demand for high-skilled foreign workers. It first focuses on this relationship at the extensive margin by exploiting variations in job postings removal. It then explores the intensive margin of this phenomemon by taking advantage from variations in job posting durations.

4.1 Focusing on the extensive margin

To assess the role of labor shortages in the demand for foreign workers, I first test whether employers are more likely to submit a LCA when their ad is still advertised on Indeed.com. The underlying assumption considers that employers who no longer advertise their ad do not face a shortage for this position. Here, I focus on the effect of shortage at the extensive margin. By exploiting the full information contained in the data, this section distinguishes between the role of actual shortages from the role of shortages perceived by employers.

Comparing the advertising status of jobs by the end of March provides me with a relative measure of firms' recruitment difficulties, at a given point in time. Assessing the role of labor shortages consists here in comparing the percentage of jobs associated with a LCA between those still advertised on March 31 and those that have been removed before. Because applications are submitted before the end of March, variations captured here could be interpreted as the effect of actual shortages on the demand for foreign workers.

As the analysis focuses on firms' demand for high-skilled foreign workers, the second part studies the role of labor shortages perceived by employers. More especially, it accounts for variations in LCA submission dates. As many LCAs are submitted way before the end of March, I modify the measure of shortages to take into account labor market forces at the moment of submission. The identification still relies on comparing percentages of jobs associated with a LCA between two groups. The group perceiving shortage gathers jobs that were still advertised online while submitting their LCA. It also take into account jobs that were still on Indeed.com by the end of March but did not sent a LCA. On the opposite, the

group not affected by shortage is incudes jobs whose job posting was removed before the LCA submitting date. It also includes jobs not associated with a LCA and removed from Indeed before the end of March. By taking into account these slight variations in the data, this second estimate captures the effect of employers' perception of labor shortages.

To identify the role of labor shortages, the statistical analysis focuses on local labor markets. More especially, it exploits between-jobs variations within each market defined by a commuting zone and a job type. Focusing on this level of analysis results from the fact that labor shortages apply uniformly to all employers looking for the same skills in a same location. Equation (1) takes this specificity into account. Estimating the determinants of LCA submission for a job j, it first controls for commuting zone (CBSA) and job-type (SOC) fixed effects. Within these cells, it then studies the average effect of getting a job still advertised, denoted SA, on the probability of submitting a LCA:

$$\Pr[\text{LCA} = 1\,]_j \;\;=\;\; \alpha_0 + \text{CBSA}_{c(j)} + \text{SOC}_{s(j)} + \alpha_1 \text{SA}_j + u_j. \tag{1}$$

Studying the probability to submit a LCA rather than petitioning for a H-1B visa allows me to focus on the unconstrained demand of firms for high-skilled foreign workers. The estimator α_1 captures therefore how employers adapt their expected recruitment strategy while facing difficulties to fill their position.

Interpreting this correlation in causal terms necessitates to identify and observe the counterfactual of interest, namely: would the job be associated with a LCA if its ad was no longer advertised online? Hereby, the identifying condition assumes that job postings are only removed from Indeed.com once the position is filled. Because 90% of ads on this specific job board are direclty scraped from employers career webpages, it supports this key assumption. More especially, employers have no incentive to withdraw their ad before filling the vacancy. This argument holds even more when they publish their ad on their own career webpage as the cost is almost null. Hereby, I identify the role of labor shortages on H-1B demand.

4.2 Focusing on the intensive margin

Exploiting variations in the duration of ads allows me to explore a second dimension of the relationship between labor shortages and demand for foreign workers. By considering the marginal effect of advertising duration on the probability to submit a LCA for job j, this section explores the intensive margin of labor shortages. More especially, it compares how ads duration explain differences in LCA status between two jobs similar in all other aspects. Equation (2) estimates this effect:

$$\Pr[\text{LCA} = 1\,]_j \;\;=\;\; \beta_0 + \text{CBSA}_{c(j)} + \text{SOC}_{s(j)} + \beta_1 \text{Duration}_j + \epsilon_j. \tag{2}$$

Similarly to equation 1, assessing the effect of labor shortages through ads duration necessitates to exploits between-jobs variations within each local labor markets. It controls for these variations through commuting zone (CBSA) and job type (SOC) fixed effects. Therefore, the coefficient β_1 estimates how ads duration affects the probability to submit a LCA for this job.

Interpreting this correlation in causal terms necessitates to observe the counterfactual situation, namely what would be the LCA status of jobs if their online duration was shorter. Therefore, the identification relies on two conditions. First, duration of ads must reflect the true duration of vacancies and must not be manipulated. Second, labor shortages must be the cause of variations in ads duration.

Since Indeed.com minimizes the presence of obsolete information on its platform, focusing on its data ensures the accuracy of the information. In other words, it makes it more likely that online duration of ads reflects the true vacancy duration. More importantly, section 6 ensures the second requirement. By implementing a McCrary test, it rejects the scenario under which job postings duration would be manipulated. Therefore the coefficient β_1 captures the causal effect of ads duration on the submission of LCA. Finally, this relationship can be interpreted as the causal effect of labor shortages on the expected demand for high-skilled foreign workers.

5 Results

This section presents estimations for both equations (1) and (2). Exploring successively both extensive and intensive margins, I show that labor shortages partly explain firms' demand for high-skilled foreign workers. First of all, table 2 highlights the role of perceived labor shortages. Moreover, table 3 precises this effect by exploring its intensive margin. It shows that jobs advertised for a longer period of time are more likely to lead to a Labor Condition Application supporting hereby the causal effect of labor shortages on H-1B demand.

5.1 Focusing on the extensive margin

This section starts by exploring how the demand for foreign workers is affected by labor shortages at the extensive margin. To explore this effect, I distinguish the role of actual from perceived labor shortages in the estimation of equation (1). According to the different measures presented in section 4, I find a significant and important effect of perceived labor shortage. On the contrary, my measure of actual shortages does not exhibit any significant difference on the demand for foreign workers.

After implementing a probit estimation of equation 1, I report the results in table 2. Successive estimations using respectively measures of actual and perceived labor shortages are reported in each column. Both specifications account for commuting zone (CBSA) and job type (SOC) fixed effects. Therefore, estimations rely on variations across jobs competing within a same labor market. Negative constant estimates reflect that most jobs included on Indeed.com are not associated with a LCA. Both measures of labor shortages do not return similar results. While both of them focus on the extensive margin of this phenomenon, table 2 indicates that employers' perceptions are more important to determine the demand for foreign workers.

On the one hand, I do not find a significant difference in LCA submission between jobs still advertised on March 31 and others. The coefficient is reported in the first column of the table. It seems that variations in labor market forces measured on March 31 do not
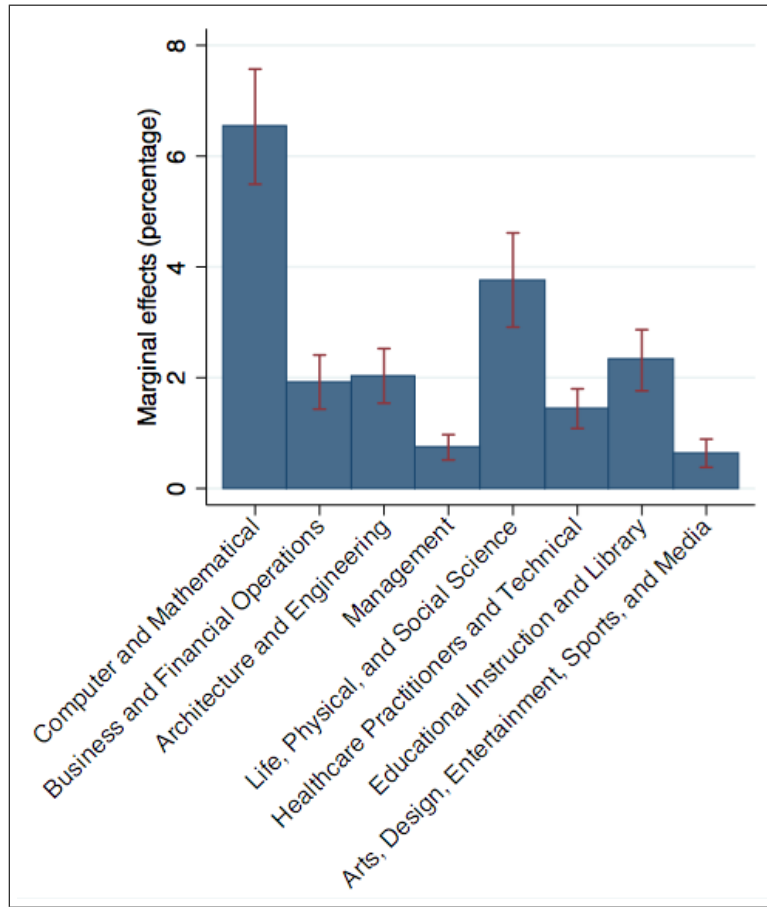
Table 2: Ads duration and demand for H-1B visas: Extensive margin

Dependent variable:	Demand(H-1B) (1)	Demand(H-1B) (2)
Still Advertized on March 31 (Actual shortage)	-0.0009 (0.0174)	
Still Advertized (Perceived shortage)		0.4897*** (0.0169)
Constant	-3.1103*** (0.1494)	-2.8942*** (0.1529)
Observations	203,789	203,789
CBSA Fixed effect	Yes	Yes
SOC Fixed effect	Yes	Yes

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

determine variations in LCA submissions. It worth noting that this result could be due to measurement errors. More especially, it could be driven by the truncated distribution of ads duration. Because job postings are automatically removed after 120 days spent online, older job ads can be measured as removed on March 31 while they are still advertised online with a new identifier. This phenomenon biases hereby the estimate towards zero. Therefore, rejecting the difference between both groups is not enough to conclude that labor shortages do not affect firms' demand for foreign workers.

On the other hand, I do find a significant difference in LCA submission when taking into account the labor shortage situation at the date of submission. Contrary to column (1), this indicator focuses on employers' perception of labor shortage by taking into account the time dimension. This specification shows that firms do respond to hiring difficulties by preparing the recruitment of a foreign worker. The second column of table 2 reports this positive and significant result. It means hereby that employers are more likely to look for foreign workers when they face difficulties to fill their position.

Figure 6: Marginal effects by SOC type: Extensive margin

Notes: This graph presents marginal effects of perceived labor shortages on the probability to send a LCA. This effect is distinguished by SOC type. Blue bars depicts the average estimate. Red whiskers represent the 95% confidence interval.

Because the magnitude of the effect is not consistent with its significance, I explore to which extent this result can be interpreted as a lower bound estimate. The coefficient reported in column (2) corresponds to a marginal effect at means of 1%. In average, jobs removed from Indeed.com are 1% less likely to submit a Labor Condition Application than their counterparts that are still advertised. Even if this effect is statistically significant, it seems very small in economical terms. Two reasons can explain the size of the average estimate: heterogeneity and measurement errors.

First of all, the small size of the effect results from its heterogeneity. As labor shortages are more or less striking across job types and commuting zones, the average estimate is not

the most meaningfull statistic to document the current debate that mostly focuses on STEM shortages. Figure 6 details this heterogeneity across main SOC categories. I find that the effect of labor shortages is more important for STEM jobs, consistently with the focus point of the H-1B public debate. Focusing on this category, the marginal effect increases from 1% to 7%. Figure 6 highlights the separation between STEM and other job categories. This difference is particularly striking with Business and Financial occupations. Whereas financial companies are represented among the largest petitioners of Labor Condition Applications, the effect of labor shortages on their demand for H-1B visas seems quite limited.

Last but not least, measurement errors lead to underestimate the true effect. In this case, errors are twofold. First, missing matches between Indeed and LCA data sources biases the average estimate towards zero as it mechanically decreases the prediction power of labor shortages. Second, measurement errors in the duration distribution of job postings increase the number of false negative reinforcing hereby the underestimation.

5.2   Focusing on the intensive margin

This section exploits variations in the number of days spent online to assess the role of labor shortages at the intensive margin. While previous section took advantage of ads removal, this part assesses whether ads posted earlier are more likely to be associated with a Labor Condition Application. It strengthens previous results by showing that labor shortages at the intensive margin cause also an increase in demand for foreign workers. After presenting the probit estimates, it reports marginal effects for one additional month spent online. I conclude this section by discussing the magnitude of the effect.

Table 3 reports probit estimates for two different specifications. Column (1) only focuses on the relationship between ad duration and the probability to submit a LCA. In addition, column (2) augments this estimation by taking into account whether jobs were still advertised by the date of submission. As previous estimations, both specifications control for commuting zone (CBSA) and job type (SOC) fixed effects. Hereby, they take into account the specificities of labor shortages and exploit variations within local labor markets. Similarly to table 2,

Table 3: Ads duration and demand for H-1B visas: Intensive margin

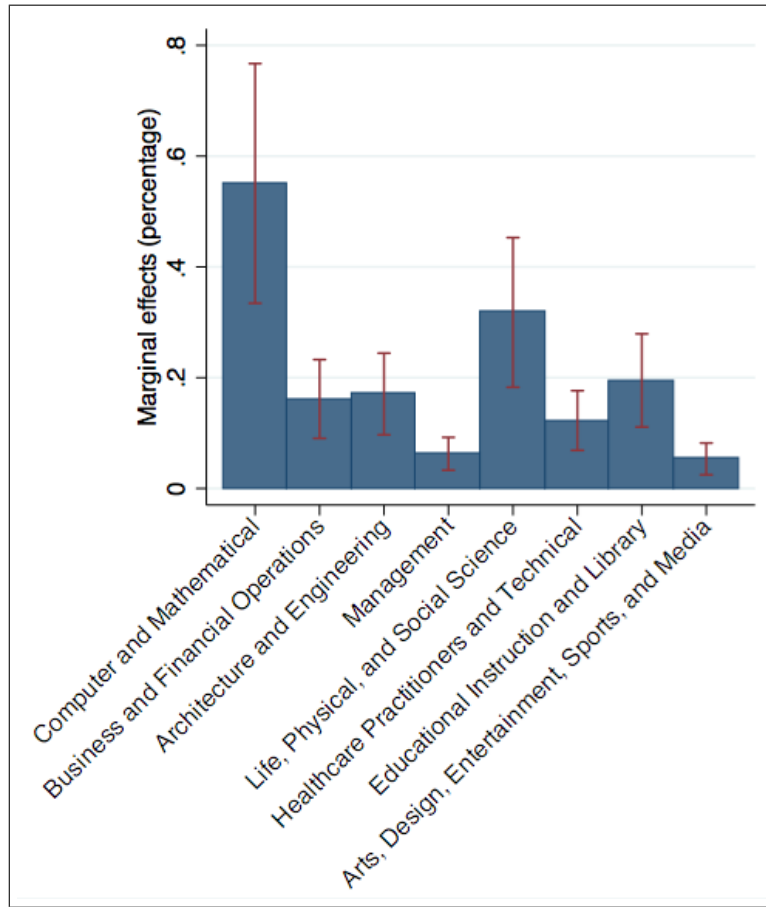| Dependent variable:	Demand(H-1B) (1)	Demand(H-1B) (2)
Job posting duration	0.0006*** (0.0002)	0.0014*** (0.0003)
Still Advertized		0.4986*** (0.0170)
Constant	-3.1317*** (0.1494)	-2.9370*** (0.1533)
Observations	203,789	203,789
CBSA Fixed effect	Yes	Yes
SOC Fixed effect	Yes	Yes

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

negative constants highlight that most jobs observed on Indeed are not associated with a LCA.

Across both specifications, there is a positive and very significant effect of job posting duration on the probability to submit a Labor Condition Application. In other words, employers who advertise longer their job on Indeed.com are more likely to start the administrative process to obtain a H-1B visa. This empirical evidence supports the idea that firms' demand for H-1B visas is a response to their difficulties to fill their vacancies. Column (2) controls also for ads that are no longer advertised by the date of submission. The effect of labor shortage measured at the extensive margin is similar to table 2. In this specification, the point estimate associated with job posting duration is twice as large as in the first column. By controling for a part of the heterogeneity, this specification consistently reports a larger effect.

Figure 7 presents marginal effects of one additional month spent online by SOC category. The relative importance of labor shortages on H-1B demand across SOC types is consistent with previous results. Similarly to figure 6, the marginal effect is way larger for STEM jobs

Figure 7: Marginal effects by SOC type: Intensive margin

Notes: This graph presents marginal effects of one additional month spent online on the probability to send a LCA. The effect is distinguished by SOC type. Blue bars depicts the average estimate. Red whiskers represent the 95% confidence interval.

compared to other categories. Moreover, demand for foreign workers in business and financial occupations is still much less impacted by labor shortages. However, the big difference with figure 6 lies on the magnitude of marginal effects. Considering job postings duration at the intensive margin returns much smaller estimates. More especially, I find for STEM jobs that one additional month spent online increases the probability to send a LCA by approximately 0.6%. This effect decreases to 0.016% when focusing on business and financial occupations.

Similarly to previous results, these figures can be interpreted as lower bound estimates. While magnitudes of marginal effects do not really make sense in economical terms, the estimates are significant at the one percent level. This supports the idea that these results

are underestimated. As measurement errors are particularly striking when considering the intensive measure of job posting durations, it may explain small magnitudes associated with these coefficients.
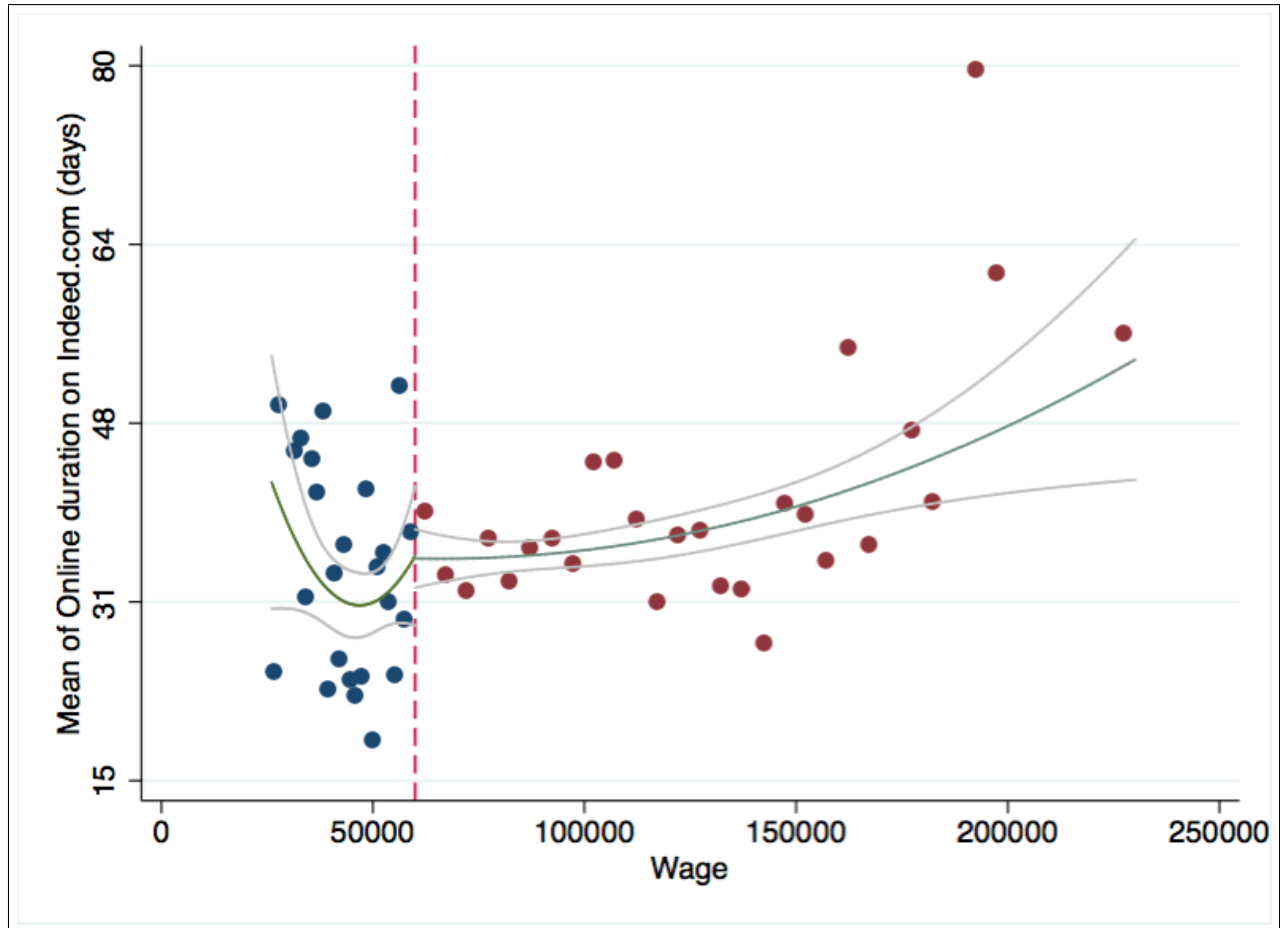
6 Robustness

The causal identification of this paper relies on the validity of labor shortage proxies. In other words, it depends to which extent job posting durations truly reflect firms' difficulties to fill their vacancies. In this section, I present the only alternative scenario that might affect the causal interpretation. Implementing a McCrary test allows me to reject this identification threat and ensures the causal effect of labor shortages on the demand for foreign workers.

Equation (2) exploits variations in job posting durations to predict LCA submission. Because the measure of ad duration is anterior to employer's decision to search for foreign workers, the identification strategy can be interpreted as Granger causality. Consequently, identification holds under the assumption that employers do not modify the duration of their job posting because they anticipate to demand a H-1B visa. For jobs paid under 60,000 dollars per year, this hypothesis is not straightforward. To obtain a H-1B visa, employers proposing this kind of jobs have to prove that they are facing labor shortages. Usually, they provide the administration with their job posting bill. By showing that they have advertised their vacancy for a long period of time without finding an American worker to fill the job, employers obtain the right to compete for a H-1B visa. Imagining a scenario where employers would like to hire foreign workers for another reason than labor shortages, this administrative feature would still provide them with an incentive to keep their ad online. Without rejecting this scenario, the causal interpretation presented in section 5 no longer holds.

To test whether this scenario represents a credible threat to my identification strategy, I exploit this 60,000 wage threshold similarly to a regression discontinuity design. Comparing the average duration of job postings around the threshold provides me with a test for manipulation. By implementing a McCrary test, I study whether jobs under the threshold

are advertised for a longer period of time. Because the adminstrative incentive for manipulation vanishes above this threshold, this alternative scenario should give rise to a significant difference between both groups.

Figure 8: McCrary test: duration discontinuity

Notes: This graph tests whether jobs paid under 60,000 $/year atificially increase their ad duration to obtain H-1B visas. Reporting no significant difference, it rejects this hypothesis.

The result of this test is graphically represented on figure 8. More formally, this procedure fits two quadratic polynomials to the data, on both sides of the threshold. It then measures the discontinuity at the cutoff and assesses its significance. As highlighted on figure 8, there is no significant difference when comparing the average advertising duration between jobs paid below and above the threshold. This test rejects hereby the scenario under which low wage employers manipulate their job posting duration to obtain H-1B visas.

Finally, increasing the prevailing wage to escape this administrative requirement does not change the result of the McCrary test. Employers paying jobs just under the threshold could slightly increase this wage to get rid of this rule. Under this scenario, they would face no incentive to voluntarily increase their job posting duration. If really not concerned by labor shortages, they would participate to decrease the average duration above the threshold, increasing hereby the observed difference. By rejecting the presence of significant difference, I ensure that demand for high-skilled foreign workers is partly caused by labor shortages.

7   Conclusion

In order to document the public debate on the H-1B program in the US, I investigate the main argument raised by the proponents of this policy. In order to increase the annual cap of visas, employers lobby the administration by arguing that their vacancies would stay unfilled without recruiting foreign workers. Due to a lack of data, this question has been overlooked in the previous literature. I address this gap by scraping job posting data from Indeed.com and provides hereby a new measure of labor shortages. I study the online duration of job ads to approximate firms' difficulties to fill their vacancies. Completing this analysis with administrative data, I observe for each job advertised on Indeed.com whether it is associated with a Labor Condition Application. Submitting this form to the Department of Labor represents the first step in the process to recruit foreign workers under H-1B visas. By exploiting variations at both extensive and intensive margins, I show that labor shortages explain a part of the demand for high-skilled foreign workers. While the magnitude of these effects is quite small, I list the reasons leading to an underestimation of the effect. I formally test the identification condition of Granger causality by implementing a McCrary test. Ensuring that job posting duration truly reflects hiring difficulties, I show that labor shortages cause firms' demand for high-skilled foreign workers. These new pieces of evidence support hereby the argument raised by employers in the debate on H-1B visas.

References

Aobdia, D. and A. Srivastava (2018). Do u.s. corporations hire u.s. educated skilled immigrants to lower their labor costs? evidence from the audit industry.

Aobdia, D., A. Srivastava, and E. Wang (2018). Are immigrants complements or substitutes? evidence from the audit industry. *Management Science 64*(5), 1997–2012.

Barnow, B. S., J. Trutko, and J. S. Piatak (2013). *Occupational Labor Shortages: Concepts, Causes, Consequences, and Cures*. Upjohn Institute for Employment Research.

Bound, J., N. Morales, and G. Khanna (2018, May). Understanding the economic impact of the h-1b program on the u.s. In G. Hanson, W. R. Kerr, and S. Turner (Eds.), *High-Skilled Migration to the United States and its Economic Consequences*, pp. 109–175. NBER.

Doran, K., A. Gelber, and A. Isen (2016). The effects of high-skilled immigration policy on firms: Evidence from visa lotteries.

Kerr, S. P., W. R. Kerr, and W. F. Lincoln (2015, June). Skilled immigration and the employment structures of us firms. *Journal of Labor Economics 33*(3), 147–186.

Kerr, W. R. and W. F. Lincoln (2010, July). The supply side of innovation: H-1b visa reforms and u.s. ethnic invention. *Journal of Labor Economics 28*(3), 473–508.

Lofstrom, M. and J. Hayes (2011, December). H-1bs: How do they stack up to us born workers? *IZA Discussion Paper 6259*.

Mayda, A. M., A. Ghosh, and F. Ortega (2016). The impact of skilled foreign workers on firms: an investigation of publicly traded u.s. firms. *IZA Discussion Paper 8684*.

Mayda, A. M., F. Ortega, G. Peri, K. Shih, and C. Sparber (2018, September). The effect of the h-1b quota on the employment and selection of foreign-born labor. *European Economic Review 108*, 105–128.

Mithas, S. and H. C. Lucas (2010, May). Are foreign it workers cheaper? u.s. visa policies and comensation of information technology professionals. *Management Science 56*(5), 745–765.

Peri, G., C. Sparber, and K. Shih (2015, May). Foreign and native skilled workers: What can we learn from h-1b lotteries? *NBER Working papers*.

Sparber, C. (2015, August). The value of h-1b status in times of scarcity. *CReAM Discussion Paper*.

U.S. General Accounting Office (2011). H-1b visa program: Reforms are needed to minimize the risks and costs of current program. Technical Report GAO/HEHS 00-157, U.S. General Accounting Office, Washington, D.C.