Vacancy Duration and Demand for H-1B Labor in the United States

Guillermo Carranza Jordan*

April 2018

Abstract

The H-1B visa program allows employers to hire skilled foreign labor in the United States. While there is a broad literature analyzing the effects of the program, there is marked lack of research analyzing its drivers. This paper analyzes the impact of job market tightness as measured by online job posting duration on H-1B labor demand. I conduct a panel analysis on national units of occupation from 2008 to 2016, after merging data from aggregated job postings and public H-1B disclosure data. I find a significant effect of yearly vacancy duration on H-1B demand, with a 10-day increase in vacancy duration associated with a 2% to 5% increase in demand of H-1B visas. These results indicate a positive relation between H-1B growth and the difficulty of filling job openings in the domestic market, evidence consistent with the intended purpose of the program.

*Senior Thesis submitted to the Department of Economics of Columbia University in partial fulfillment of the requirements for Bachelor of Arts' degree with Honors.

Acknowledgments

First of all, I would like to thank Professor Wojciech Kopczuk for his support throughout my thesis project. I was blessed with his willingness to share his knowledge and expertise during this past year. His insights have definitely helped me become a better researcher during this exercise.

I would also like to thank Professor Michael Best, the leader of our senior thesis seminar, for organizing the mechanics of our seminar and for the extra time he graciously offered me to discuss my project. For the same reason, I thank all the many members of the Columbia Economics' department faculty that were open to listen to my thoughts regarding my project and gave me insightful feedback.

I am very grateful to the folks at Linkup who were willing to freely share their data for use in this project. This present paper most definitely would not have been possible without the support from your company.

I could not have finished this project with the help of the amazing friends I have made at Columbia, who help me perfect my work and motivated me throughout this process.

Finally, this paper is also fruit of the support and motivation I received from my parents from the moment I started writing my thesis. I cannot not properly describe how grateful I am for your investment and support in the projects that I embark on. My educational journey is a reflection of your love. Los quiero mucho.

1 Introduction

The H-1B visa program has been at the center of a political debate about its perceived effects on employment and domestic wages since its inception in 1990. The intention of the H-1B program is to allow employers to hire skilled foreign labor when equivalent domestic labor is not available. Supporters of the program, including large tech companies who hire a very significant proportion of H-1B workers, argue that the program is necessary to improve domestic productivity and for their industries to remain globally competitive. Critics of the program, on the other hand, counter that the H-1B program is a means for companies to decrease their wage costs by hiring less expensive foreign labor (Torres, 2017). While there has been an increased research interest in understanding the effects of the visa program, much less attention has been given to analyzing its drivers and causal factors, like how domestic labor market conditions impact the demand for this program. This paper elucidates the effect of job market tightness, as measured by online job vacancy posting duration, on the number of H-1B applications submitted in a given year by occupation. The goal is to empirically answer to what extent the demand of H-1B visas is driven by domestic employers' difficulties in finding appropriate workers. If there is no relationship found between job market tightness and H-1B demand, the argument in support of the program based on a lack of appropriate labor supply from domestic applicants becomes less tenable.

To answer this question, I construct a new dataset at the occupational level that incorporates aggregated vacancy data and public H-1B disclosure data. The vacancy dataset used for this project is a representative sample of online job postings from 2008 to 2016. After classifying the job postings into occupations, I am able to reaggregate measures of vacancies over finer units of occupation and different time frames. This finer detail is an improvement on comparable datasets that aggregate at higher levels of industry and groups of occupations. I exploit the large number of occupation units to conduct a panel regression of the number of H-1B applications on average vacancy duration and number of new jobs created per occupation. The second variable is included to control for overall expansions or contractions in the demand for workers of a given occupation. Furthermore, I also weight the occupation units by their prevalence in the H-1B data, in order to represent more accurately the dynamics of the H-1B distribution of occupations.

I find a statistically significant result for vacancy duration on H-1B demand, at the 1% significance level. A 10-day increase in vacancy duration is associated with a 1.8% to 4.9% increase in the number of H-1B visas demanded, and the effect is significant regardless of whether a weight was used or not. This result indicates that there exists a positive relationship between labor market tightness and H-1B demand, which corroborates the arguments made in support of the program. Employers do in fact appear to respond to increased difficulty of filling positions by hiring skilled H-1B foreign workers. On the other hand, the coefficient estimate of vacancy count throughout these regressions is surprisingly negative, indicating a negative relationship between domestic job creation and H-1B demand. However, this unexpected result in this control variable is likely caused by bias from firm and occupation heterogeneity.

After establishing a significance vacancy effect, I am also interested in analyzing differences in this effect for STEM and non-STEM occupations. The most prevalent occupations in the H-1B dataset are in STEM roles, and they are usually the focus of the policy considerations towards the program. I also introduce variables that control for education requirements and wage differences between H-1B applications and domestic workers. I find that the effect of vacancy duration is even more significant for STEM occupations. A 10-day increase in vacancy duration is associated with a 7.7 to 10% increase in H-1B demand, using weighted regressions and the control variables. On the other hand, the effect of vacancy duration is not significant for non-STEM occupations. This last result suggests that the specific skill composition of STEM occupations generates a larger effect of vacancy duration, indicating that employers have higher difficulty finding workers with STEM skills. As a secondary question, I also analyze the temporal dimension of the effect. This question is in dialogue with vacancy literature analyzing the perception and response of firms to hiring difficulty. In this sense, I find that the effect found above for STEM occupations is strongly placed in the second quarter of a given application cycle's previous year. There does not appear to be a continuous prolonged effect of vacancy duration throughout the year. The location of the effect at Q2, not in the closest quarter to H-1B application submission, suggest that the perception of the difficulty of filling positions and the decision to hire these workers occurs significantly earlier in the year.

The results of this paper are robust to the high degree of demand from outsourcing firms observed in the H-1B dataset. The number of H-1B applications filled by outsourcing increased rapidly in the regression time frame, with a peak in 2014. The large proportion of outsourcing firms is one of the main arguments against the necessity of the program, since they are associated with wage decreasing strategies in firms and not hiring difficulties (Torres, 2017). Even after removing the demand from these firms, the results stay statistically significant, with a slightly greater vacancy duration effect for STEM occupations. Therefore, the growth of this sourcing practice does not negate the previous conclusion of the positive effect of hiring constraints on H-1B demand. Finally, our results are also robust to other internal validity assessments, including the relaxation of the panel assumption of common trends across occupations.

The rest of this paper is structured as follows. Section 2 gives an overview of the H-1B program and discusses previous research on the employment visa and market vacancy tightness. Section 3 describes and summarizes the main trends in the two datasets, online job vacancies and H-1B applications, as well as the sources of data for the control variables: wage and education. Section 4 summarizes the panel identification strategy. Section 5 describes the results and later discusses their significance, identifies possible threats and further assesses robustness. Finally, Section 6 concludes and explores further avenues of research with the present datasets.

2 Background Information

2.1 The H-1B Program

According to the Department of Labor, the H-1B program is an employment visa that allows American employers to "hire nonimmigrant aliens as workers in specialty occupations or as fashion models of distinguished merit and ability". A "specialty" occupation is defined as one requiring the use of a highly specialized body of knowledge, usually related to the completion of at least the equivalent of a US bachelor degree. The application for a H-1B visa is a process that greatly depends on employer's actions. It usually consists of 3 steps. First, the employer needs to submit a Labor Condition Application (LCA) to the Department of Labor, in order to ensure that a set of labor regulations established to protect domestic employment are followed. Most important is the wage protection rule. According to the US Citizen and Immigration Services (USCIS, 2018):

> The employer will pay the beneficiary a wage which is no less than the wage paid to similarly qualified workers or, if greater, the prevailing wage for your position in the geographic area in which you (the applicant) will be working

In the same spirit, the employer must also ensure that the applicant will not be used for union-busting purposes. After the LCA has been approved, the employer submits the "Petition For a Non-Immigrant Worker", Form I-129, to the USCIS. If the employer's petition is approved, as a final step, the applicant can then apply for their visa documentation in a US embassy abroad. In 2005, a cap of 65,000 visas was established for employee applicants with bachelor's degrees only, and an extra 20,000 visas for applicants with masters or more advanced degrees. If the number of applications exceeds the established cap, the applications are selected using the H-1B lottery system. Applications for occupations in institutions of higher education, non-profits or government research agencies are not subject to the cap (USCIS, 2018). The USCIS starts receiving applications for the next fiscal year on the first week of April, and releases employment authorizations in October. Every year starting from 2014, the number of applications has reached the H-1B cap within the first five days of the start of application season, which causes the USCIS to deny any further applications for that given year. The fast filling behavior of the H-1B quota has been given as evidence of the large domestic demand for skilled foreign labor through the program (Kight, 2018).

In this paper, I use the number of Labor Condition Applications as a measure of H-1B demand. By regulations of the Immigration and Nationality Act, the Department of Labor discloses basic information from each LCA received every year. However, information about the I-129 applications and the final authorizations are not made public, and the present project does not have access to this approval data. Given my research question, it is important to note that, in order to assess demand, information about approvals is not necessary. The fact that an employer submits a LCA already indicates a willingness to hire a H-1B worker. It is therefore possible to understand demand looking solely at the LCA figures. This data consideration is in line with previous H-1B research measuring demand (Clifford, 2014; Kerr et al., 2013; Wilson et al., 2012).

2.2 Previous Research

There has been an increased interest in economic research concerning the H-1B program in the last few years. Most of the research has been focused on the effect of H-1B workers on domestic labor markets, in terms of employment, wages and firm innovation. The state of the present literature in terms of effects does not give conclusive evidence on an overall positive or negative effect. Mayda et al. (2017), for example, find that, after the H-1B cap was instituted in 2004, the employment of H-1B workers fell sharply, while the employment of domestic workers in the same firms did not increase. Peri et al. (2015) find negative native employment effects for negative H-1B supply shocks, estimated using the random lottery assignments of 2007 and 2008. These results seem to indicate a complementarity between native workers and foreign workers in H-1B occupations. In terms of productivity and innovation, Kerr and Lincoln (2010) show increases in invention and patenting in cities with higher proportion of foreign H-1B employees. Ghosh et al. (2014) find similar results. On the other side of the H-1B argument, Doran et al. (2014) exploit the introduction of the lottery in 2006 and 2007 and show negative crowding out in terms of employment in firms using H-1B employees, a decrease in domestic wages and an increase in firm profits. Focusing on computer scientists, Bound et al. (2017) build a model of IT skill demand and find that wages would have been 5% higher in 2001 without the introduction of the program.

Furthermore, critics of the program often argue against the actual necessity of the program, regarding whether or not H-1B positions can be filled with domestic labor supply. This is a concern that until recently has not received the same attention as the question of effect. Related to this issue, there is debate, for example, on whether H-1B workers are over or underpaid compared to their domestic counterparts (Mukhopadhyay and Oxborrow, 2012; Rothwell and Ruiz, 2013). The fact that a large proportion of the H-1B applications and approved visas go to outsourcing firms paying only the market's prevailing wage is a further source of concern (Hira, 2014). The present project contributes to the analysis of the H-1B program in terms of this second question, the necessity of these foreign workers. Assuming job market tightness indicates a higher difficulty of filling a position, this project examines directly H-1B trends in the face of domestic occupational needs, empirically testing the argument that H-1B demand is disjointed from employer needs. The research most directly related to this project analyzes the demand of the H-1B program in terms of STEM skill shortages specifically. Rothwell and Ruiz (2013) conclude that a higher proportion of jobs usually filled with H-1B visas go unfulfilled after 30 days compared to an average position, using Help Wanted Online Data from 2011. This number is calculated using inflows of new positions and employment figures, and not vacancy duration directly. Wilson et al. (2012) show that H-1B capped visas follow cyclical macroeconomic trends, but that uncapped visas are relatively stable across time. Clifford (2014), from the Federal Reserve of Boston, finds that there is a very high proportion of H-1B visas in IT-related occupations in metropolitan markets in the North East, although, worryingly, a high proportion of the employment comes from outsourcing firms. The present project contributes directly to this literature by analyzing the impact of hiring difficulty in H-1B demand at a national level and with a longer time frame (2008-2016). The project also does not focus solely on IT-related occupations, and systematically includes all occupations represented in the H-1B dataset, while controlling for relative frequency and importance.

Another question of interest for this project is the effect of vacancy duration on firms' hiring decisions. Under the Diamond-Mortensen-Pissarides model of employment, firms face a cost from not filling a vacancy quickly, either due to their search efforts or lost productivity. An increased level of job market tightness decreases the firm's probability of filling a position and therefore increases search costs. Given how cumbersome and costly it is to fill a position with a H-1B employee, the search cost would need to be higher than the cost of the H-1B hiring for the firm to decide to sponsor the employee. In this sense, this present project is in conversation with previous literature that examines firm's responses to vacancy constraints.

An interesting line of research examines the effect of job market tightness on discrimination. Baert et al. (2015) finds that foreign sounding names are more likely to receive callbacks in the Belgian youth labor market when applying for positions that are more difficult to fill. Moss and Tilly (2000), using employer's surveys, show that employers are more willing to hire African American workers in the United States during the 90s when local job market tightens, even though their stated preferences towards black workers do not change. This research indicates that tightness can force employers to alter their hiring preferences for less desired workers. Furthermore, there is also some research on the effect of job tightness on firm's hiring scheduling. Gorter et al. (2003) find that increased labor tightness decreases the probability of hiring all their employees at once and is linked with an increase in gradual hiring. The literature is scarce on further analysis of tightness effects on firm hiring behavior. This project analyzes the possible response of hiring foreign labor in the face of hiring constraints. Moreover, the nature of my constructed dataset makes it possible to temporally place when the vacancy effect is more significant on H-1B demand in relation to the calendar year. This approach will inform whether the effect is spread throughout the year, concentrated in time, or if vacancy constraints are only significant closer to the filling deadline.

3 Data

3.1 Classification of Job Titles

As a first step, I will discuss the classification structure used for job titles in the vacancies and the H-1B applications dataset. The identification strategy of this paper requires the assortment of observations in occupation units, in order to build a panel dataset of vacancies and H-1B demand. To this end, I will use the 2017 version of the O*NET-SOC codes from the Occupation Information Network (O*NET). Maintained by the U.S Department of Labor, the O*NET program collects statistical information for different occupations and industries in the United States, and also maintains a classification structure for occupations, based on 950 codes. Built upon the previous Standard Occupational Classification System, the O*NET-SOC codes are a series of 8 numbers, each consecutive number indicating a smaller level of aggregation. The O*NET program also hosts an online classification tool called O*NET Code Connector, which helps employers and the public identify the correct classification code based on a job title or description.

I will now briefly describe how the classification was conducted for the job titles in the datasets. The main focus of the classification effort pertained to the vacancy dataset, as the observations were classified using a proprietary structure that was too aggregated for comparison with the H-1B applications dataset. After cleaning the strings for non-alphanumerical characters, the titles were ranked according to their number of occurrences in the dataset. A list of about 135,000 most frequent strings were then each passed through the O*NET Code Connector Application, and an occupational code was retrieved for each. For the other non-frequent strings in the dataset, I searched through each job title to find matches with the code-assigned most frequent list. For example, this allowed for the classification of "Critical Nurse" and "Critical Nurse Miami full-time" with the same "29-1141.03" code (Registered Nurses). At the end, this simple classification algorithm allowed for 92% of the vacancy titles to be code-assigned. It is important to note, however, that the Code Connector matching tool is not perfectly accurate, which introduces a degree of measurement error in our classification process. Our empirical strategy will specifically try to deal with this source of bias by clustering within larger occupational groups. This empirical approach will be further discussed in the Method Section. Finally, on the side of the H-1B dataset, the same algorithm was used to classify applications before 2009, given that employers were not required to use the SOC codes with the LCA's fax submission system in use before this year.

At many points of this paper, including my empiric specification, I rely on the notion of occupational clusters, a group of similar occupations working on the same field. For the definition of occupational clusters, I used the "Career Cluster" and "Career Pathway" structure from the O*NET code taxonomy. According to its documentation, "Career Clusters contain occupations in the same field of work that require similar skills" [1]. Appendix Table B.1 shows a sample of occupations with their corresponding career cluster and career pathway. Some clusters contain a higher degree of homogeneity compared to others, but most of them still contain a large number of within group occupation code variation. The present analysis will use the duo of career cluster and career pathway as a "Occupational Cluster". The advantage of using this type of grouping instead of industry or other type classification is that these clusters associate occupations that are very related to one another while still maintaining a degree of heterogeneity. One single profession can be found across different industries and as such, it does not allow for a clean separation in groups. At the same time, the cluster heterogeneity is important for our panel identification strategy. Furthermore, the internal diversity of the cluster organization will help reduce the degree of measurement error as related occupations that could be wrongly assigned are associated within the same group.

3.2 Vacancy Data

The information on vacancies comes from the online job aggregator LinkUp. The company hosts a job search engine, populated with job ads compiled directly from companies' websites. By indexing straight from employer's websites, Linkup ensures that their data set contains fewer duplicate postings, expired jobs and "unauthentic" vacancies (scams, fraud, etc). The removal of job board "pollution" gives the LinkUp an advantage in terms of measurement of labor demand and job market prediction, compared to alternative datasets like the Conference Board's Help Wanted Online series (Dayton, 2016). The dataset contains observations from approximately 45,000 employers, starting from late 2007 until August of 2017. Each observation of the dataset corresponds to one job posting. Each data point contains information on the advertised position job title, employer information

[1] Data and information on the clusters was obtained from the O*NET website. Available at: https://www.onetonline.org/find/career?c=0&g=Go

including geographical location and, most importantly for our analysis, the date the posting was created and the last date the posting was online. The difference between these two values is our measure for online job duration for a given observation. Job duration was then aggregated for a particular occupational code by estimating the mean duration value for vacancies that were posted within a given time frame (e.g "Mean duration of job ads for Registered Nurses between January 1st and January 31st 2010"). I use this aggregation measure instead of the median to take advantage of the large sample properties of the mean, including smaller variance in the series. For a given level of aggregation, I additionally constructed a count variable that aggregates the number of observations for an occupation at a given time period.

One important source of concern for the construction of vacancy duration variable based on online data concerns its relationship with true job market tightness. The online nature introduces possible sources of bias. There could be a disconnect between the end of the employer's search and the end of the job posting. For example, it is possible that some employers leave the position posted after the position has been filled. The existence of this behavior could generate the presence of large outlier values, and shift the values of the variable upwards from real vacancy duration. For this paper, online positions with outlier values larger than 200 days were capped at this limit value. In the Robustness subsection of Results, I carry the panel analysis using a different measure of vacancy duration. Instead of measuring average online post duration, I calculate the proportion of postings with duration longer than 60 days, similar to the analysis of Northman and Ruiz. The 60 days' threshold was chosen to reflect a limit point above which vacancies are likely to remain unfilled. This share of "filled" positions measure is less susceptible to outlier values, as the magnitude of the value itself does not enter the aggregation measure.

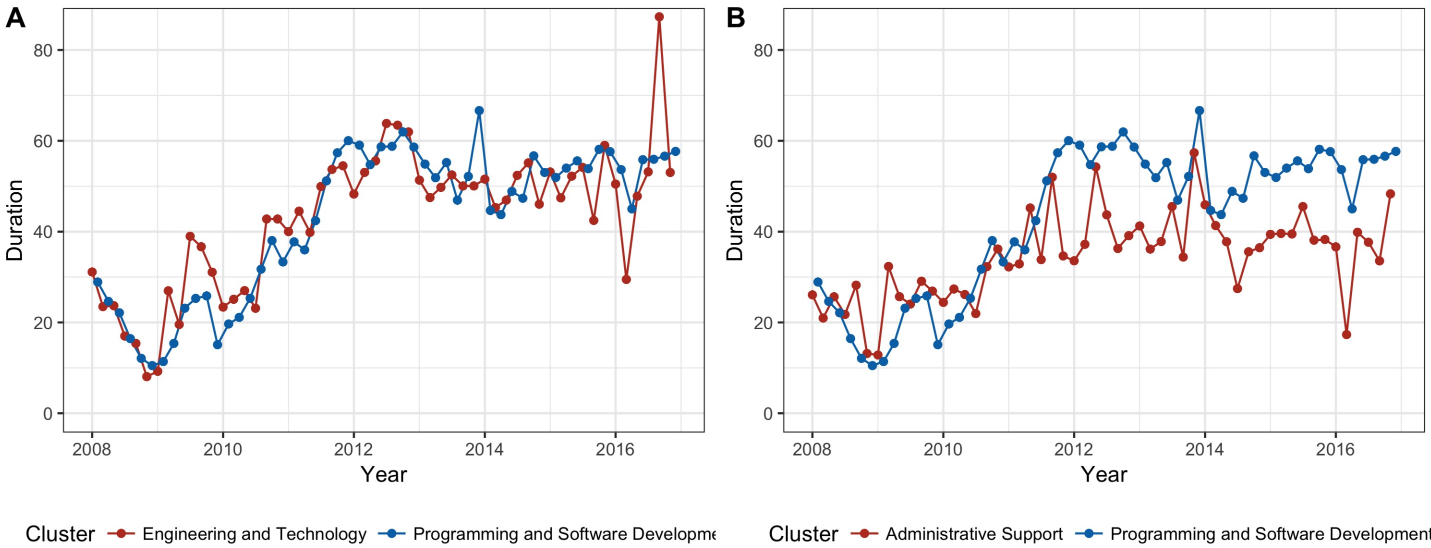
A natural question that follows from this proxy formulation of vacancy is if our present variable, mean online job duration, matches up with other indicators of vacancy tightness. Vacancy duration is usually defined as the time it takes an employer to fill a position, from the moment the position is posted until an offer is accepted. In the analysis of vacancy duration, employer surveys have been the predominant way this job market tightness measure has been estimated (Rothwell, 2014). Another approach has been to model vacancy duration using inflows and outflows of vacancies and employment figures. Davis et al. (2010) models a matching function with vacancies, unemployment and recruitment effort as inputs with data from JOLTS to estimate vacancy duration. This model is the basis for the DHI-DHF Vacancy Duration Measure. [2] In Figure A.1, I compare mean online job duration with the DHI-DHF Index. We can see that the mean online job duration is much more volatile in variance compared to the DHI-DHF Index. Its range of values is also greater, probably due to the presence of outliers discussed before. At the same time, the constructed series and the DHI-DHF capture similar trends in the domestic labor market. There is a marked valley in both series during the Great Recession, with a slow increase in the recovery years and a more recent leveling of the growth rate. This pattern is consistent with our understanding of vacancy duration during recessions. As employers are able to choose quickly from a larger pool of applicants, vacancy duration drops quickly. In the recovery process, job applicants have greater leverage in recruitment and vacancy duration increases. From this comparison, it does not appear that using mean online job duration to stand for vacancy duration is an inappropriate measure to capture macroeconomic trends within occupations.

After the observations are classified by occupation using the classification algorithm, we can briefly visualize how the vacancy trends differ, using occupational clusters. As expected, occupations that are similar in skills or roles have similar trends of vacancy duration. For example, the "Programming and Software Development" cluster has a time series that is very close to that of "Engineering and Technology", as shown in Figure 1B. Similarly, if we compare the "Programming" cluster with the "Administrative Services" cluster, which includes occupations like bookkeepers, office clerks and receptionists, we see a divergence starting from the post-Recession period. In Figure 1B, the later cluster shows similar values to the "Programming" cluster during the recession valley, but displays lower values afterwards. Since the administrative occupations are not high-skilled, it is not surprising that employers have less difficulty finding employees for this occupation compared to computer programmers. At the same time, it is important to remark that the relationship of skill with vacancy time is not so clear cut. For example, heavy truck drivers have even higher vacancy duration numbers than programmers, which is explained by the shortage of applicants and a rising shipping demand in the past few years (Raphelson, 2018). As outlined above, there is significant variation between occupations in terms of vacancy duration trends that we will exploit with our identification strategy.

Since our identification will also account for the effect of the vacancy counts of occupations,

[2] DHI-DHF Time Series was obtained from FRED https://fred.stlouisfed.org/series/DHIDFHMVDM

Figure 1: Comparison of Mean Vacancy Duration between Clusters



Note: Data on online job postings obtained from LinkUp from 2008 to 2016. The observations were classified into occupations using a classification algorithm based on the O*NET Code Connector Tool. The occupations were then grouped to larger clusters based on the "Career Pathway" classification of O*NET. Each data point is the mean value of the duration of job postings from a cluster, that were started in a given month.

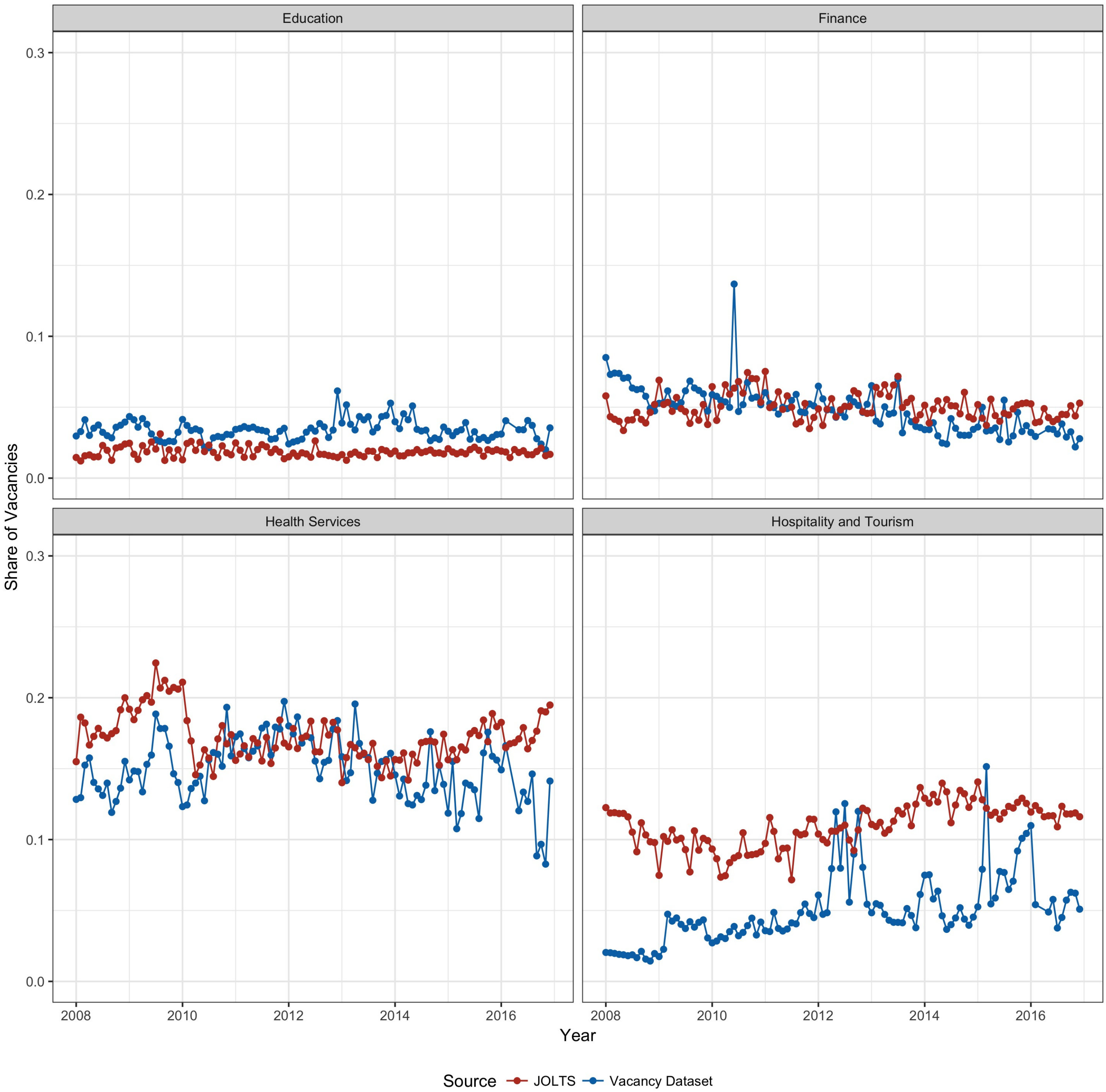
it is also important to explore the representativeness of this variable in relation to true domestic dynamics. An important characteristic of this vacancy dataset to consider is that the efficiency of the compilation of jobs increased throughout the years, due to improvements in the compilation algorithm. Furthermore, there may also be concerns about representativeness of the jobs compiled. If the scrapping algorithm focused first on jobs coming from certain industries, like IT or healthcare, the evolution of vacancy count would reflect a trend towards a more representative sample, and not the dynamics of the true domestic vacancy numbers. On the other hand, if the distribution by occupation of new postings in the vacancy dataset reflects the same domestic distribution across years, this concern can be addressed by the panel empirical strategy. After converting vacancy count figures to log form, the ratio difference between the true domestic vacancy number and the sample vacancy count would be picked up by the year fixed effects, if this ratio is constant across occupations. Evidence of a similar distribution can also help establish the representativeness of the sample of vacancies from LinkUp. Therefore, an important assumption to test is whether the share of vacancies by occupation in our sample reflects other estimates of the domestic share of new jobs created.

An established source of estimates on job openings comes from the Job Openings and Labor Turnover Survey (JOLTS)[3], compiled by the Bureau of Labor Statistics. Unfortunately, the JOLTS's industry estimates are too aggregated for direct comparison with our occupation or cluster numbers. In order to compare the distributions between sources, I aggregate certain clusters to approximate the industry estimates available from JOLTS. The groups I aggregate are "Finance", "Health Services", "Education" and "Hospitality and Tourism". The first three are selected because they are very prevalent in the H-1B dataset, and because they are less broad compared to other industry groups in JOLTS. Unfortunately, there is no easy comparison for IT occupations, because their JOLTS's "group" (Professional and Business Services) is too broad for a clear comparison. After establishing similar groups in our vacancy dataset, I calculate the share of total new vacancies for each group using both sources. The results for the "Health Services" and "Education" groups can be found in Figure 2. Overall, the shares from both data sources appear to have similar magnitudes, with "Education" appearing to have a slightly higher proportion in our vacancy dataset. More importantly, the similarity in magnitudes appears to remain constant throughout the years. "Health Services" appear to have a smaller share in the vacancy dataset only slightly after 2015. This similarity in shares gives evidence to the assumption of a constant distribution of new vacancies between the occupations described above.

For the "Finance" group, the share from the vacancies dataset appears to match well with the JOLTS estimates, except again for a slight decreasing share in the later years. The graph for "Hospitality and Tourism" shows a similar positive trend between the two groups, but the JOLTS share appears to have higher magnitude values overall. This gap could be due to a difference in the composition of the groups between the two sources of data. In any case, the increase in proportion in our dataset over time can be explained by a similar increase in the domestic share as evidenced by JOLTS, and not completely by changes in the compilation algorithm. This aggregated comparison supports our assumption of representative sampling in the vacancy dataset. Further assuming similar composition in the smaller occupation units, we can use the vacancy count per occupation in the identification strategy to examine H-1B dynamics, as will be outlined in the Method section.

[3]Data series obtained from the FRED website at https://fred.stlouisfed.org/categories/32241

Figure 2: Comparison of Share of Vacancies from Occupational Cluster by Source



Note: Each data point corresponds to the share of vacancies from a certain group of occupations with the total number of vacancies in a given month. The red points were calculated from the Job Openings and Labor Turnover Survey. The blue points represent monthly share from the LinkUp's vacancy dataset. The vacancy observations were classified into occupations and then aggregated for comparison with the JOLTS's industries. The months of March and April of 2016 were removed because they are outlier values due to unrepresentative collection of certain occupations.

3.3 H-1B Disclosure Data

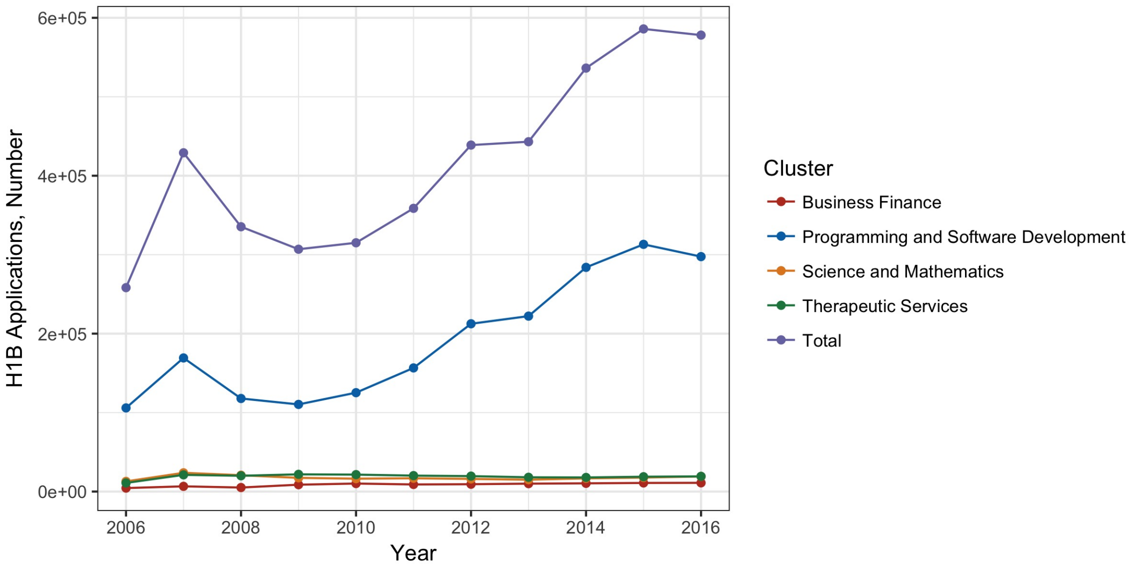
As a first step to apply for H-1B visas, employers need to submit Labor Condition Applications (LCAs) to the Department of Labor's Office of Foreign Labor Certification (OFLC). The LCA database is available publicly on the OFLC website[4], starting with applications from 2001. For the present analysis, I take the subset of the available data starting from 2007 to match the time span of the vacancy dataset. Furthermore, applications that were withdrawn by the employer were also removed. The dataset contains the job title of the position, employer information including location and the number of workers requested for each position. Each LCA application also contains a prevalent market wage, usually taken from the OFLC office, and a proposed wage for the position, which should be on par with the prevalent wage. Finally, in terms of classification, each observation processed after 2009, with the "iCert System", comes with a SOC code. SOC codes are very easy to match with the O*NET classification, as they are very similar and several merging cross-walks are available. Observations before 2008 only include a more general 3-digit Dictionary of Occupational Title (DOT) code. For these observations, observations were classified applying the job title algorithm used with the vacancy dataset.

On the other hand, the H-1B count series shows a strong seasonal pattern throughout the years, as it can be seen in Figure A.2. Since applications for next year's available visas are received at the start of April, there is a huge spike in the number of LCA applications submitted for certification in March. Moreover, in Figure A.2, it can be observed that the number of applications during the rest of the year is minimal compared to the huge spike that occurs before April. Therefore, it is not possible to obtain monthly-level assessment of H-1B visa demand from the LCA application count. It is clear that most employers wait to submit their applications right before the next application cycle. With this motivation, I aggregated the H-1B count at a yearly level between April of a given year and April of the next year, and deal with a yearly level of H-1B demand for all of the following analysis.

In terms of macroeconomic trends, the number of H-1B applications also shows a marked decrease during the recession period, similar to the vacancy duration series. The number of applications did not recover completely to pre-Recession numbers until the 2013 application cycle.

[4] Available at: https://www.foreignlaborcert.doleta.gov/performancedata.cfm

Figure 3: Number of H-1B Applications of Selected Clusters and Total



Note: Each data point is the number of Labor Condition Applications between April 1st of a given year and March 31st of the next year. Labor Condition Applications are publicly available on the Office of Foreign Labor Certification website. Using each observation's SOC code, the LCAs were classified into clusters using the "Career Pathway" classification of O*NET. The graph shows data for 4 of the most representative clusters and the total count.

Exploring a bit more the composition of the LCA applications by occupation, we can see a marked prevalence of IT-related occupations. The H-1B numbers is dominated by one particular cluster: "Programming and Software Development". In Figure 3, we can see that the overall H-1B count and the "Programming" cluster count follow very similar shapes, while the rest of the most popular clusters show minimal numbers in comparison. In Appendix Figure A.3, I zoom in the graph to observe clusters other than "Programming", and we can see how they show heterogeneity in trends and not a clear pattern compared to the overall H-1B figures. Overall, the H-1B count is driven to a large extent by the dynamics of the occupations inside the "Programming" cluster. These patterns raise the issue of whether to control for the relative importance of occupations for the present empirical analysis. The effect of vacancy time for the occupations in the "Programming" cluster could drive the regression results if given an inappropriate weight. Alternatively, not weighting occupations appropriately would fail to reflect the overall dynamics in the program. An inaccurate weighting would be equivalent to over representing smaller occupations. In this project, I will show and compare results using mean H-1B count between 2008-2016 as weights as a way to address the relative prevalence of occupations.

3.4 Wage and Education

One of the most common charges leveled against the H-1B program is one of wage-undercutting, that is firms using the H-1B program not out of a need for skilled labor but to decrease labor costs (Kirkegaard, 2005). A higher degree of wage undercutting, if it exists, would impact vacancy duration, since employers' recruitment effort and behavior would be influenced by the desire to lower wages. The estimation of the effect of vacancy duration will then most likely be biased, since its coefficient will also capture the effect of this wage behavior. Unfortunately, the real wages payed to H-1B workers by industry are not easily available. It is also difficult to measure the degree of the depression of wages due to the lack of wage data after the firm has hired the H-1B worker. Most of the present research on wage comparison relies on ex-post surveys or USCIS disclosure requests. As such, there is no direct available variable we can use to control for the extent of this behavior.

For this project, I will try to estimate and control for this effect by using the proposed wage information in the H-1B dataset as a proxy. Due to H-1B regulation, each LCA application sent for certification contains a "proposed wage" for the position, which is required to be at least 100% of a "prevailing wage" as determined by the OFLC. The Department of Labor defines a prevailing wage as "the average wage paid to similarly employed workers in a specific occupation in the area of intended employment". Employers can search for prevailing wages for their applicants at the OFLC website. The website lists four tiers of wages, from Level I for entry level positions to Level 4. Therefore, the values of proposed wages are constrained by the lower bound given by the OFLC.

From the H-1B dataset, I calculate the proposed wage quartiles (25%,50% and 75%) for a given occupation code in a year. As a first glance, we can compare the quartiles of the proposed wages against their domestic counterparts. Estimates of domestic wage quartiles for every occupational code can be obtained from the Bureau of Labor Statistics, which publishes Occupational Employment Statistics yearly[5]. Figure 4 shows the evolution of the quartiles from the H-1B dataset and the OES estimates for "Computer Programmers" (O*NET Code=15-1131.00). We can see that the range of proposed wages comes from the lower tail of the domestic wage distribution for computer programmers. Moreover, the H-1B wage distribution also has a much smaller variance, which could be explained by the H-1B regulations. The wages for "Computer Programmers" are therefore very

[5] Available at: https://www.bls.gov/oes/tables.htm

concentrated around a value smaller than the domestic average. The relative position of this proposed wage distribution is not evidence in itself of wage reduction behavior, but it does tell us that the H-1B population is most likely composed of cheaper entry-level Level I positions. It is important to note that there is a high degree of heterogeneity in relation to relative wage distribution position between occupations. Appendix Figure A.4 shows the wage distributions for "Accountants" and "Economists". While "Accountants" follow a similar trend to "Computer Programmers", opening a bit in the later years, the trend for "Economists" is starkly different. For "Economists", the distribution of H-1B wage is very similar to its domestic counterpart for this occupation.

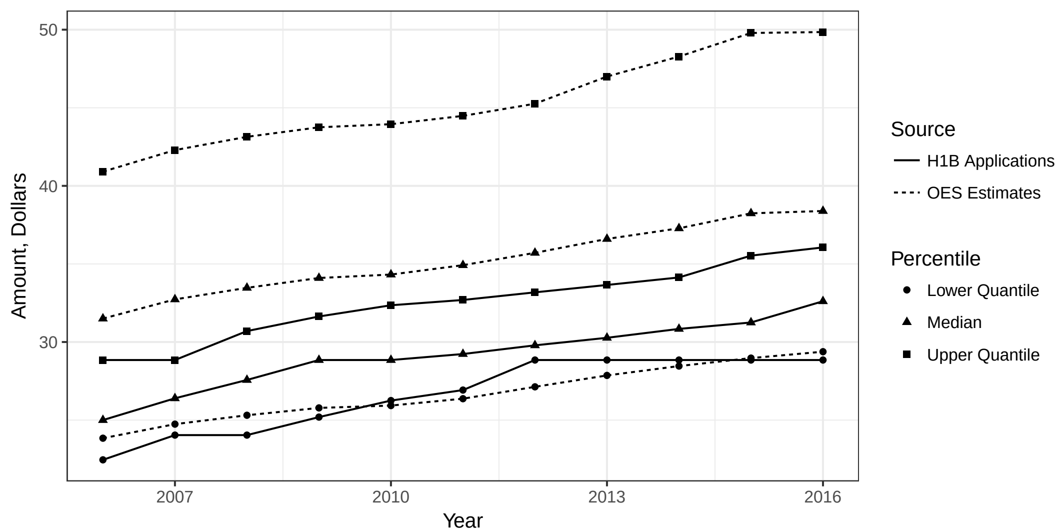
I create a set of wage variables to analyze the effect of this variation in the H-1B dynamics. The variables are simply the ratio between a given H-1B proposed wage quartile divided by the domestic quartile. For the lower 25% quartile, for example, this amounts to:

$$r_{25\%} = \frac{wageforeign_{25\%}}{wagedomestic_{25\%}}$$

The evolution of these ratios captures the relative change of the H-1B distribution compared to the domestic wage distribution. For an occupation where the H-1B wage growth is highly outpaced by the domestic wage growth, we can assume there are some wage incentives on the part of the employer that are causing this disparity. It is important to note that if employers are willing to pay a wage that is lower than the average market wage, they have a higher incentive to propose the lowest "proposed" wage possible at the moment of application. This dynamic would cause these quartile ratios to decrease over time. Furthermore, the magnitude of these ratios themselves also allows for inter-occupation comparison of the relative position of their H-1B wage distributions. To reiterate, the H-1B proposed wages are not the actual wages payed to the employees once the hiring process is completed. Still, due to lack of availability of reliable H-1B wage information, I will try to control for wage incentives looking at the relative growth rates of wages in these two populations, given that the absolute difference is not available.

Education and skill level is another important control to include in this analysis. The vacancy duration variable is a measure of overall job market tightness, which is affected by the composition and skill level of the labor supply. It might also be important to control and examine the effect of this skill level effect separately. For my identification strategy, the inclusion of this variable

Figure 4: Distribution of H-1B Proposed Wages and Domestic Wages for Computer Programmers



Note: Each Labor Condition Application specifies a "Proposed" wage. The full lines represent the yearly "Proposed" wage quartiles for computer programmers in the LCA dataset, from 2006 to 2016. The dashed lines are estimates for domestic wage quartiles for computer programmers. The estimates are obtained from the Office of Employment Statistics.

would require data for skills and educational requirements for each occupation code and each year of analysis. If there is no significant change in the educational skillset within our time frame of analysis, the fixed educational effects will be picked up by fixed occupation effects in the panel regression. As part of the O*NET project, the organization also compiles information for mean educational requirements per occupation (O*NET, n.d.). Unfortunately, this information is not updated yearly. For our period of analysis, each occupation is updated at most twice. Moreover, average educational requirement is not the only component of skill level, since work experience and training must also be important components of skill. Our present education data is not the best dynamic assessment of skill requirement for these professions, but it does help us control to an extent with the available data. Furthermore, the O*NET mean educational requirement also helps to compare the effect of education on the vacancy values across occupations. Finally, as part of our identification strategy, I try to estimate the difference in effects if any between STEM and non-STEM applications. Again, for this separation, I rely on the classification of the O*NET database, which classifies certain occupational codes as STEM based on the nature of skills used in the occupation and the field of work usually associated with a given occupation (O*NET, n.d.). For the panel dataset, 373 occupations are classified as non-STEM and 175 are classified as STEM.

4 Method

This paper hypothesizes that vacancy duration is associated with an increase in H-1B demand, because it is representative of an employer's difficulty in finding employees with the appropriate characteristics for the position. As explored in the "Previous Research" section, this effect can also be understood as a relative decrease in the cost of filling a position with an H-1B worker relative to the cost of continuing searching, when a position goes unfulfilled for a long time. Furthermore, from a search and match theory, we can also assume that a higher number of vacancies creates longer periods of vacancy duration, as employers have to compete for the same pool of candidates and their probabilities of finding appropriate matches decrease. In this sense, vacancy duration and vacancy count are positively correlated. For this reason, assuming a relationship between the number of domestic jobs and the number of H-1B visas demanded, the exclusion of vacancy count could create a biased estimate in vacancy duration. This possible bias is why I include both measures in all of the empirical specifications outlined below.

In terms of the empirical strategy, I estimate the effect of vacancy duration via a panel regression using occupation codes as units of analysis. The panel specification allows for the controlling of occupation and year fixed effects, unobservable variables that could bias the estimation of the effect of vacancy duration. Occupation fixed effects includes occupation heterogeneity that is constant across years, like occupation specific hiring practices and occupation specific H-1B rules (university professors). Year fixed effects would capture time-changing variables that are constant across occupations, including macroeconomic domestic trends. The nature of panel data generates autocorrelation between unit observations, that is addressed using clustered errors. In this specification, the errors are clustered at the occupational cluster level, and not at the code level. This approach is taken in order to overcome the possibility of measurement error in the occupational code's classification. If some observations are misclassified between codes from the same cluster, then the two occupational codes should not be treated as having independent error term variance. Furthermore, beyond measurement error, the occupational codes are also most likely not independent from one another in that they share common industry trends, like hiring practices, industry shocks, etc. Finally, as a last data cleaning step, I only select occupations that have LCA applications at least five years between 2008 and 2016. This selection is done to remove the very rare occupations in the

H-1B dataset. At the end, there are 556 occupation units in the panel dataset, with observations from 2008 to 2016.

I conduct the following panel regression:

$$y_{it} = c + \alpha_i + \beta_t + \delta_1 x_{it-1} + \delta_2 z_{it-1} + \epsilon_{it}$$

where y_{it} is H-1B log application count for an occupation in a given year, x_{it-1} is vacancy duration of the application cycle's previous year, z_{it-1} is the log number of domestic vacancies of the application cycle's previous year , α_i is a vector of occupation fixed effects and β_t is a vector of year fixed effects. For comparison, I also run an OLS regression and panel regression without year fixed effects. Given the discussion of the composition of the H-1B dataset, with the strong preponderance of certain occupations, I also estimate the previous regressions using weights, to control for prevalence disparity. I take as weights the averages of the H-1B counts per occupation during the 9 years of the panel observations. The weights also take into account the difference in accuracy of the mean values of the vacancy variables, since occupations with smaller number of observations will have mean and interval estimates with a higher level of imprecision.

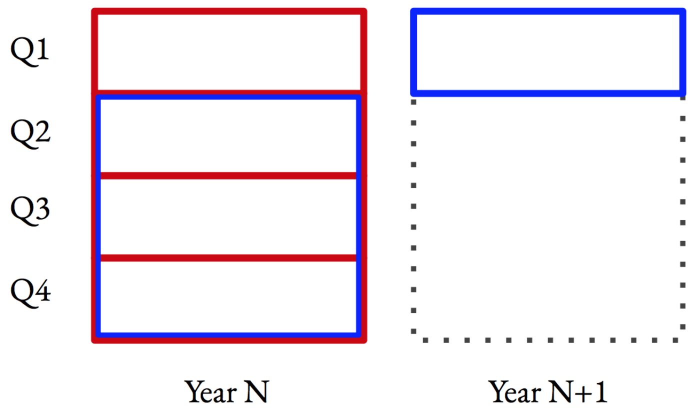
Instead of yearly aggregate estimations, I then run the previous regressions after disaggregating the vacancy variables into quarters. This specification allows me to examine the precise temporal effects of vacancy in relation to its position in the application cycle. The above regression then becomes:

$$y_{it} = c + \alpha_i + \beta_t + \delta_1 X_{it-1} + \delta_2 Z_{it-1} + \epsilon_{it}$$

, where now X_{it-1} is a vector of vacancy duration from Q1 to Q4 of the application cycle's previous year. Similarly, Z_{it-1} is a vector of vacancy count in Q1 to Q4 of the previous year. To reiterate, due to the dynamics of the H-1B cycle, I take the yearly H-1B count from Q2 of a given year up until and including Q1 of the next year, which is when most of the applications are submitted. This specification of time frames can be visualized in Figure 5, where the red cells correspond to the vacancy quarters and the blue squares correspond to the span of the yearly H-1B count aggregation.

Furthermore, there is additional data on education and wage, described in the previous Method

Figure 5: "Diagram of the Time Frame of Vacancy and H-1B count variables"



section, that we use to control for possible bias in estimation of the vacancy effect. The model with controls is:

$$y_{it} = c + \alpha_i + \beta_t + \delta_1 X_{it-1} + \delta_2 Z_{it-1} + \delta_3 C_{it-1} + \epsilon_{it}$$

where C_{it-1} is a vector of control variables during the previous year which includes average required years of education and the wage comparison variables: the ratio of the H-1B proposed wage percentile divided by the domestic percentile estimates for the 25%, 50% and 75% percentiles. Year and occupation fixed effects are also included throughout. At this stage, I also separate the occupations into STEM and non-STEM groups, as defined by the O*NET taxonomy to examine the difference in effects.

Finally, in the Robustness section of Results, I relax the assumption of common trends across occupations, and include cluster-specific time trends and cluster-specific year fixed effects. With cluster-specific year trends, the model is:

$$y_{it} = c + \alpha_i + \beta_t + \delta_1 X_{it-1} + \delta_2 Z_{it-1} + \delta_3 C_{it-1} + \delta_4 \theta_i * t + \epsilon_{it}$$

where θ_i is the linear time trend for a specific cluster. The previous model assumes a linear trend in the growth of H-1B variables for a given cluster, which in a post-recession period might be inappropriate. As an alternative, I also run a model with cluster-specific year fixed effect, which modifies the overall year fixed effect with cluster-specific differences. The model becomes:

$$y_{it} = c + \alpha_i + \beta_t + \delta_1 X_{it-1} + \delta_2 Z_{it-1} + \delta_3 C_{it-1} + \delta_4 \theta_i * \beta_t + \epsilon_{it}$$

where $\theta_i * \beta_t$ is the cluster specific extra effect on the year fixed effect.

5 Results

5.1 Initial Results

Table 1: Descriptive Statistics by Occupational Cluster

	Occupational Cluster	Vacancy Duration	H-1B Count	H-1B Percentage
1	Programming and Software Development	44.06	2204842	46.33
2	Information Support and Services	39.21	333834	7.02
3	Engineering and Technology	42.92	285017	5.99
4	Therapeutic Services	37.36	215153	4.52
5	Science and Mathematics	39.93	197904	4.16
6	Operations Management	40.40	177110	3.72
7	Teaching/Training	43.10	167760	3.53
8	Network Systems	38.69	162593	3.42
9	Business Finance	54.39	98312	2.07
10	Accounting	45.58	97634	2.05

Table 1 contains some descriptive statistics for the occupational clusters with the largest prevalence in the H-1B dataset. The table shows mean vacancy duration, the total number of H-1B applications and their relative frequency. Similar to what was found in Figure 3, the “Programming and Software Development” cluster is almost half of the total number of H-1B applications between 2008 and 2016. With a sharp decrease in relevant frequency, the next most prevalent occupational clusters are also STEM related, including “Information Support and Services” and “Engineering and Technology”. The end of the list is composed of teaching, medical and business positions. In terms of vacancy duration, most of the average duration values for these clusters range between 35 and 45, with the exception of the “Business Finance” cluster with a value of 54.4.

I further examine the relationship between the nature of occupations and vacancy duration by analyzing the impact of education and skill on this measure. Table 2 summarizes the results of regressions of vacancy duration on education and skills. Regression 1 is an OLS regression with average educational requirement in years as the independent variable. As outlined in Section 3.3, the O*NET database contains estimates of educational requirements by occupation. Regression 1

Table 2: Effect of Skills on Vacancy Duration

	(1) OLS	(2) Panel	(3) OLS	(4) Panel
Education Years	0.55***	0.45+		
	(0.149)	(0.233)		
STEM			2.85***	2.91***
			(0.681)	(0.599)
Constant	40.4***	21.4***	41.9***	22.7***
	(0.818)	(1.255)	(0.365)	(0.553)
Observations	4569	4569	4956	4956
Adjusted R^2	0.00429	0.217	0.00351	0.214

Standard errors in parentheses

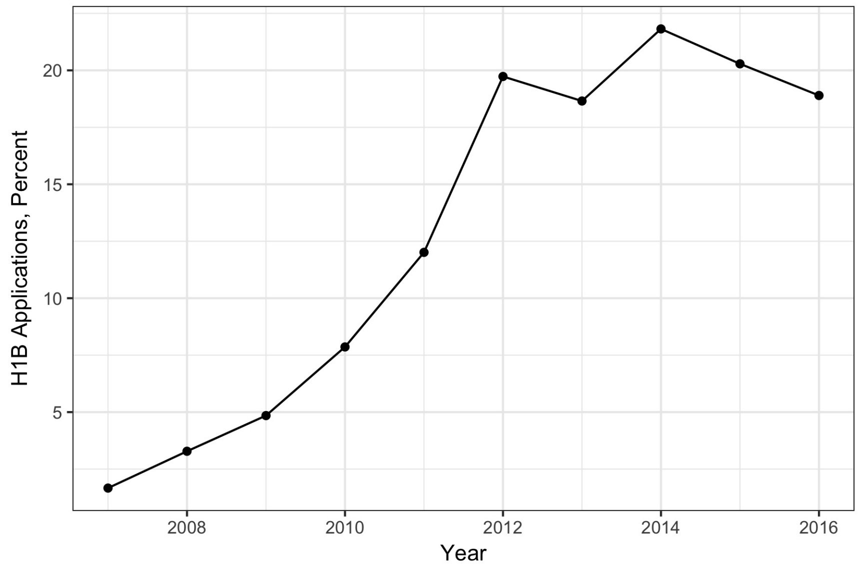
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Top 10 Employers in H-1B Dataset

	Employer	H-1B Count
1	Infosys	150810
2	Tata Consultancy Services	81079
3	Wipro	51475
4	Deloitte Consulting	47232
5	Accenture	39970
6	IBM India	37474
7	Microsoft	35657
8	HCL America	27816
9	Capgemini America	24938
10	Ernst & Young	24308

indicates that 1 additional year of education is associated with 0.55 more days of vacancy duration. Regression 2 includes year fixed effects and clusters errors for each occupational unit. The effect of vacancy duration is now 0.45 and it is statistically significant at the 10% level. The time fixed effects remove common variables affecting all occupations each year, which increases the robustness of the education coefficient. For the last two regressions, I compare the difference in vacancy duration for STEM and non-STEM occupations. The list of STEM occupations again comes directly from the O*NET database. STEM occupations have longer vacancy durations by approximately 3 days. The difference is significant at the 0.1% level after accounting for year fixed effects. These results seem to corroborate Northwell and Ruiz's results which find that STEM occupations take longer to fill compared to non-STEM ones. Furthermore, the previous result also implies that there is a more general effect of education on vacancy duration beyond the STEM nature of the occupation.

Figure 6: LCA Applications from Outsourcing Firms by Year



Note: The percent of H-1B LCA applications relative to the yearly total count coming from a list of outsourcing firms. Each LCA application contains the name of the employer demanding the position. The outsourcing firms were selected from the firms that appeared at least once in the Global 100 Outsourcing List between 2008 and 2016.

Analyzing now the employer composition of the H-1B dataset, we can see a marked prevalence of IT consulting and outsourcing firms. Since the LCA applications dataset contains employer information, I use this information to analyze the extent of H-1B demand from these types of firms. Table 3 shows the top 10 employers in terms of LCA applications submitted between 2008 and 2016. Except for Microsoft and Deloitte, all of the remaining firms are involved in business sourcing practices. Five of them are large IT firms from India: Infosys, Tata Consulting, Wipro, IBM India and HCL (Park, 2015). The overall make-up of the H-1B firm composition corroborates the 2014 finding from Ron Hira on the prevalence of applications from outsourcing firms. Furthermore, I also analyze the temporal evolution of this outsourcing prevalence. I compile a list of the outsourcing firms that appear in the Global Outsourcing 100 ranking from 2008 to 2016. The members of the annual list are selected by the International Association of Outsourcing Professional, and gives recognition to the firms with best practices in the industry (IAOP, n.d.). While most major and important players in the H-1B dataset appear once in the list in this period, I should note that the compiled list does not completely contain all firms that supply this service. Still, this small list of around 200 firms composes around 20% of all LCA applications between 2012 and 2014, as can be observed in Figure 6. The percent of applications from outsourcing firms has grown from a very small level before 2007 to a peak of around 22% in 2014, the year that Hira analyzed. At the same time, it can be seen how in 2015 and 2016, the number of visas from these firms slightly decreased from its 2014 peak, which points to a possible present downward trend. In the coming Robustness section, I address the prevalence of this demand in the regression estimates, factoring out its impact in the analysis of vacancy duration effects.

5.2 Vacancy Effect

Table 4: Effect of Vacancy Duration at Code Aggregation

	(1) OLS	(2) Panel	(3) Panel	(4) W. OLS	(5) W. Panel	(6) W. Panel
Vacancy Duration	-0.0046	-0.0035^{+}	0.0018**	0.011*	0.018***	0.0049***
	(0.00347)	(0.00176)	(0.000644)	(0.00526)	(0.00342)	(0.000931)
Vacancy Count	0.28**	-0.17**	-0.050	0.85***	0.16	-0.28***
	(0.0890)	(0.0533)	(0.0726)	(0.0869)	(0.155)	(0.0652)
Constant	1.47^{+}	5.05***	4.28***	0.17	6.82***	11.4***
	(0.752)	(0.480)	(0.646)	(0.835)	(1.683)	(0.693)
Observations	4956	4956	4956	4943	4943	4943
Adjusted R^2	0.0473	0.0315	0.0846	0.307	0.238	0.347
Year FE		No	Yes		No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4 summarizes the results of the effect of the vacancy variables on H-1B demand at the occupational code level, with errors clustered within an occupational cluster. To reiterate, occupational codes constitute what is usually understood as a well-defined occupation. Table 4 shows the results using yearly aggregated vacancy variables. The unweighted panel regression shows a statistically significant coefficient with magnitude of 0.018. The coefficient increases in magnitude when the regression is weighted by H-1B representativeness. In Regression 6, the coefficient on vacancy duration increases to 0.0049. These results indicate that a 10-day increase in vacancy duration is associated with an increase in H-1B applications from 1.8 to 4.9%, depending on whether it is weighted or not. On the other hand, the effect of vacancy count is negative and statistically significant at the 0.1% significance level using the weighted panel model in Regression 6. Since both the H-1B demand variable and vacancy count are in log form, the coefficient estimate indicates that

Table 5: Effect of Vacancy Duration at Code Aggregation by STEM status

	(1) Panel	(2) Panel	(3) W. Panel	(4) W. Panel
Vacancy Duration	0.00097	0.0026	0.0025	0.0034^{*}
	(0.00131)	(0.00158)	(0.00332)	(0.00148)
Vacancy Count	-0.0019	-0.032	-0.34^{**}	-0.065
	(0.0509)	(0.117)	(0.0978)	(0.178)
Constant	3.53^{***}	4.98^{***}	9.99^{***}	9.65^{***}
	(0.475)	(0.975)	(0.879)	(1.849)
Observations	3375	1581	3362	1581
Adjusted R^2	0.162	0.0162	0.0763	0.461
STEM	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a 1% increase in domestic vacancies is related to approximately a 0.28% decrease in the number of H-1B applications. The point estimate for the unweighted regression is also negative, but not significant. This result is surprising. If the number of domestic positions available increases, we would expect that the H-1Bs used to fill a share of the positions would increase as well. The estimate found in Regression 6 suggest the opposite.

In Table 5, I estimate again the effect of vacancy duration, while separating the occupations by STEM status. Firstly, for STEM occupations, the effect of vacancy duration is 0.0034 and is significant at the 5% level. The point estimate of vacancy count is again negative but not statistically significant. This estimate indicates that a 10-day increase in vacancy duration increases the number of STEM H-1B applications by about 6.8%. On the other hand, for non-STEM occupations, while its point estimate is positive and in line with the previous results, the effect of vacancy duration cannot be distinguished from the null alternative. It is important to note that no significance can be found in the unweighted Regressions 1 or 2, which could be a function of the higher degree of measurement imprecision in the less common occupations. Still, the point estimate of vacancy duration for STEM occupation is positive and its magnitude is within the range of previous results.

From these results, we can conclude that vacancy duration seems to have a significant effect on H-1B demand using yearly aggregated vacancy measures. The point estimates range from 0.0018 to 0.0049. The estimate found in Table 4, Regression 3, is even significant without weighting.

Moreover, using the whole sample, a negative effect of vacancy count was also found, with a magnitude of -0.28. These results do not appear to be the same for STEM and non-STEM occupations. Vacancy effect only appears significant for STEM occupation and the negative vacancy count effect only present for non-STEM occupation. This unexpected result will be further discussed at the end of next section. Using yearly vacancy variables, we find that job market tightness is a significant factor in STEM H-1B hiring, which is the predominant group of occupations that use this visa.

5.3 Temporal Effect Disaggregation

Table 6 now shows the effects of vacancy count and duration disaggregated at a quarter level as discussed in the Method section. Regressions 1 and 3 are the same as Regressions 3 and 6 in Table 4. In Regression 5, the effect of vacancy duration during the second quarter is significant at the 1% level, with a coefficient of 0.0055. Vacancy duration at Q1 also has a coefficient of magnitude 0.0031, with a p-value of 0.06. In the unweighted Regression 2, vacancy duration during Q1 is significant at the 1% level, with a coefficient of 0.0022. Both regressions indicate that the effect of vacancy duration is concentrated at the earlier quarters. It does not appear like the effect is distributed throughout the year, since the coefficients of vacancy duration at Q3 and Q4 are essentially zero. Furthermore, the earlier sharp effect at Q2 is unexpected, since I had hypothesized that the marked effect should appear closer to the filling deadline, if at all. The results earlier in the year could be explained if employers make the decision at the later quarters, but are affected by information gathered in earlier quarters. Employers could be affected by vacancy duration at Q2 if their understanding of the job market, and its level of hiring difficulty, was done during this period. Considering how troublesome applying for a H-1B visa is, the existence of a delay between information assessment and the hiring decision is probable. Furthermore, there appears to be again a negative effect of vacancy count at Q4, with a coefficient of -0.27 significant at the 1% level. No significance is found in the unweighted regression.

Table 6 also contains the tentative control variables discussed in the Data section. The inclusion of the control variables does not seem to have a major effect in the estimation of the coefficients of interest. It is important to note that the sample size of the regression with the full array of control variables is smaller compared to ones without it. As it will be seen in the Robustness section, the sub-selection of the sample size does not seem to have a major effect in the results. Going

Table 6: Effect of Vacancy Duration at Code Aggregation with Quarter Vacancy, Controls

	(1) Panel	(2) Panel	(3) Panel	(4) W. Panel	(5) W. Panel	(6) W. Panel
Vacancy Duration	0.0018**			0.0049***		
	(0.000644)			(0.000931)		
Vacancy Count	-0.050			-0.28***		
	(0.0726)			(0.0652)		
Duration Q1		0.0022***	0.0018**		0.0031+	0.0035*
		(0.000580)	(0.000628)		(0.00160)	(0.00178)
Duration Q2		-0.00016	0.00019		0.0055**	0.0049**
		(0.000652)	(0.000879)		(0.00193)	(0.00144)
Duration Q3		-0.00018	0.00092		-0.00078	0.00062
		(0.00109)	(0.000762)		(0.00142)	(0.00102)
Duration Q4		-0.00023	0.000018		-0.00026	0.0014
		(0.000525)	(0.000596)		(0.00176)	(0.00188)
Vacancy Count Q1		0.0023	-0.0043		-0.036	-0.023
		(0.0215)	(0.0269)		(0.0401)	(0.0509)
Vacancy Count Q2		0.015	0.045		-0.024	-0.025
		(0.0475)	(0.0665)		(0.0849)	(0.0496)
Vacancy Count Q3		-0.017	-0.060		0.071	-0.020
		(0.0654)	(0.0383)		(0.0583)	(0.0555)
Vacancy Count Q4		-0.027	-0.026		-0.27**	-0.26***
		(0.0340)	(0.0312)		(0.0832)	(0.0653)
Education Years			-0.026			-0.11+
			(0.0640)			(0.0605)
Lower Quartile			0.14***			0.11
			(0.0320)			(0.733)
Median			0.037			-0.85
			(0.212)			(0.712)
Upper Quartile			-0.87***			0.50
			(0.150)			(0.607)
Constant	4.28***	4.16***	5.53***	11.4***	10.8***	12.3***
	(0.646)	(0.617)	(0.313)	(0.693)	(0.722)	(2.020)
Observations	4956	4883	3823	4943	4881	3823
Adjusted R^2	0.0846	0.0972	0.186	0.347	0.356	0.459
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

back to the table, in weighted Regression 6, the magnitude of Q2 vacancy duration diminishes slightly from 0.0055 to 0.0049 after including the controls. A small decrease in the coefficient also occurs in unweighted Regression 2. Interpreting the effects of the controls briefly, required years of education has a negative but insignificant effect with and without weighting. Furthermore, the wage variables do not seem to have any explanatory power for the weighted regression (the p-value of their joint significance test is 0.71). Interestingly, however, these variables do become significant in the unweighted specification. The lower quartile ratio has a coefficient of 0.14 and the upper quartile ratio a coefficient of -0.88, both statistically significant at the 0.1% level. The wage variables capture the relative growth of H-1B proposed wage relative to its domestic counterpart. The fact that these variables do not have a significant effect in the weighted specification indicate their effect is more representative for smaller occupations in the dataset.

In Table 7, I separate occupations by STEM status and estimate again the regressions of Table 6. All the regressions in Table 7 are weighted. For STEM occupations, we find again the marked effect at Q2 found in the previous table. In Regression 3, the coefficient is 0.01, double of what is estimated using the whole sample. This effect is also statistically significant at the 0.1%. After including the control variables, the magnitude of the coefficient decreases to around 0.0077. The decrease in sample size is not a worrying concern after including the control variables because most STEM occupations have the full array of variables. These results indicate that a 10-day increase in vacancy duration for STEM professions causes a 7.7 to 10% increase in the number of H-1B applications. Furthermore, vacancy count does not appear to have a significant effect at any quarter. The control variables also do not appear to have explanatory power for this sub-sample. Regressions 1 and 2 hold the same results for non-STEM applications. Vacancy duration does not have a significant effect at any quarter. On the other hand, vacancy count seems to have a significant and negative effect at Q1, with a coefficient of magnitude -0.24. There also appears to be a significant effect at Q4, an effect which probably caused the similar whole-sample significance at Q4 found in Table 6. Interestingly, two of the wage variables also have a significant effect in the non-STEM sample. The median ratio has a coefficient of -1.58 and the upper quartile ratio, a coefficient of -1.08. An increase in these wage ratio causes a decrease in the number of H-1B applications, which indicates an inverse relationship between proposed wage increases and H-1B demand. This is tentative evidence of wage pressure incentives for non-STEM occupations. Such

Table 7: Effect of Vacancy Duration with Quarter Vacancy and Controls, by STEM status

	(1) W. Panel	(2) W. Panel	(3) W. Panel	(4) W. Panel
Duration Q1	0.0043^{+}	0.0016	0.0011	0.0038
	(0.00238)	(0.00135)	(0.00218)	(0.00270)
Duration Q2	-0.00016	-0.00088	0.010***	0.0077***
	(0.00212)	(0.00127)	(0.00248)	(0.00195)
Duration Q3	-0.0016	-0.00027	-0.00063	0.0014
	(0.00184)	(0.00158)	(0.00147)	(0.000863)
Duration Q4	0.00061	0.00092	-0.0017	0.00032
	(0.000967)	(0.00139)	(0.00261)	(0.00231)
Vacancy Count Q1	-0.24***	-0.24***	0.096	0.15^{+}
	(0.0608)	(0.0605)	(0.0694)	(0.0822)
Vacancy Count Q2	0.045	0.13^{+}	0.056	-0.0044
	(0.0725)	(0.0687)	(0.166)	(0.0809)
Vacancy Count Q3	0.035	-0.0044	0.050	0.0070
	(0.0860)	(0.0730)	(0.0698)	(0.0756)
Vacancy Count Q4	-0.19*	-0.16**	-0.14	-0.099
	(0.0753)	(0.0529)	(0.0932)	(0.0904)
Education Years		-0.030		-0.099
		(0.0828)		(0.0798)
Lower Quartile		0.17		-0.027
		(0.248)		(1.505)
Median		-1.58**		-0.61
		(0.505)		(1.260)
Upper Quartile		-1.08**		1.01
		(0.328)		(0.771)
Constant	9.60***	12.4***	8.33***	8.27*
	(0.867)	(0.987)	(1.934)	(3.202)
Observations	3325	2508	1556	1315
Adjusted R^2	0.0828	0.340	0.473	0.591
STEM	No	No	Yes	Yes

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a relationship is not found for the STEM sample.

The past two tables have shown a significant effect of vacancy duration at Q2. Similar to what was found in Table 5, in Table 7 we have shown very disparate dynamics between STEM and non-STEM occupations. As indicated above, a 10-day increase is associated with a 7.9 to 10% increase in H-1B demand for STEM occupations. The effect is only significant at Q2, which indicates a small time frame under which STEM employers are affected by tight job market conditions. Given that the average vacancy duration for occupations in the "Programming Cluster" is 44 days, a 10-day increase would represent a 22.7% increase in duration at the mean, which is a very significant effect in magnitude. Furthermore, thE high significance of this vacancy duration effect is in line with result found in Section 5.1, that STEM workers are more difficult to hire than average. It appears this increased difficulty also translates into a higher demand of H-1B visas.

Vacancy count appears to have a negative effect for non-STEM occupations at Q1 and Q4. The effect is not significant for STEM occupations. As stated before this result is unexpected, because we would assume a positive effect of higher number of domestic vacancies in H-1B demand. I will outline two considerations to better understand this stated negative effect. First of all, the inclusion of both vacancy duration and vacancy count in the regression require the coefficient of the later variable to capture effects uncorrelated with the effect of vacancy duration. As stated in the Method section, a larger number of vacancies causes a tightening of the labor market. For STEM occupations, it appears vacancy duration captures all the explanatory power contained in vacancy count, i.e. there is no apparent effect independent from labor market tightening. The lack of significance could therefore be explained by the relationship between vacancy count and vacancy duration. Secondly, the vacancy count measure is also likely to be correlated with firm and occupation heterogeneity changing across time. An increase in vacancy count indicate a period of expansion and economic vitality for industries. During these periods, many of the hiring and productivity decisions of firms in a boom could also shift preference and demand for H-1B visas. If H-1B visas are preferred in negative downturns as a wage saving mechanism, negative vacancy count values would be correlated to higher demand for the program. Due to this possibility of omitted variable bias, we cannot use the present coefficient to make conclusions on the effect of vacancy count. At the same time, we can also say as a silver lining that the inclusion of vacancy count would control for this firm and occupation heterogeneity in our estimation the effect of vacancy duration, if any bias is present. In this sense, even though the effect of vacancy count cannot be interpretable, the estimation of vacancy duration becomes more robust to the unobservables correlated with vacancy count.

5.4 Robustness

As a first assessment of robustness, I control for the decreased sample size in Regressions 1 and 3 of Table 7 due to incomplete data on domestic OES estimates and education. I replicate Table 7 by estimating these regressions using only observations with complete data. Table 8 contains these results. For STEM occupations, the coefficient for Regression 3, without controls, is smaller than the estimates obtained with the full sample (0.0083 vs. 0.01). This discrepancy suggests that the decrease in the coefficients when the controls were included in Regression 6 - Table 7 is also caused by a negative sample effect. For non-STEM occupations, the coefficient on vacancy count at Q1 increases from -0.24 to -0.32. Moreover, the estimate for vacancy count at Q4 changed from -0.19 to -0.18. For these occupations, the sample was considerably reduced when the control variables were introduced. The fact that the coefficients did not change dramatically gives confidence in the interpretation of these results. The estimates in Table 8 suggest there would be slightly different estimates for Regressions 3 and 6 if wage and education data, the control variables, were available for the entire sample. However, since the difference observed is relatively small, these discrepancies should not cause major concern about the existence of a considerable sample bias in the regressions with control variables.

Another source of concern for our results is measurement error in the vacancy duration. Even though measuring average vacancy duration gives us a more easily interpretable variable, there is a higher possibility of outlier bias in its measurement. As a further assessment of robustness, I replicate the original regressions with an alternative vacancy variable, the share of positions "filled". As discussed in the Data section, instead of taking the mean vacancy duration, instead I can measure the proportion of positions removed or "filled" before 60 days. Table 9 and Appendix Table B.2 show the results for the regressions with this alternative variable, and we can see that the results are very similar to the ones originally obtained with vacancy duration. First of all, the signs of the coefficients of the share variable are negative because a tighter job market is related to a smaller value of this variable, as there is a smaller proportion of jobs being filled before the

Table 8: Effect of Vacancy Duration by at Code Aggregation with Quarter Vacancy, Restricted Sample

	(1) Panel	(2) Panel	(3) W. Panel	(4) W. Panel
Duration Q1	0.0021	0.0016	0.0039	0.0038
	(0.00194)	(0.00135)	(0.00330)	(0.00270)
Duration Q2	-0.0013	-0.00088	0.0083***	0.0077***
	(0.00188)	(0.00127)	(0.00187)	(0.00195)
Duration Q3	-0.00072	-0.00027	0.0018	0.0014
	(0.00192)	(0.00158)	(0.00118)	(0.000863)
Duration Q4	0.00082	0.00092	-0.0000034	0.00032
	(0.00111)	(0.00139)	(0.00236)	(0.00231)
Vacancy Count Q1	-0.32***	-0.24***	0.19+	0.15+
	(0.0631)	(0.0605)	(0.108)	(0.0822)
Vacancy Count Q2	0.084	0.13+	0.026	-0.0044
	(0.0785)	(0.0687)	(0.149)	(0.0809)
Vacancy Count Q3	-0.078	-0.0044	-0.061	0.0070
	(0.0715)	(0.0730)	(0.0977)	(0.0756)
Vacancy Count Q4	-0.18**	-0.16**	-0.080	-0.099
	(0.0655)	(0.0529)	(0.0860)	(0.0904)
Education Years		-0.030		-0.099
		(0.0828)		(0.0798)
Lower Quartile		0.17		-0.027
		(0.248)		(1.505)
Median		-1.58**		-0.61
		(0.505)		(1.260)
Upper Quartile		-1.08**		1.01
		(0.328)		(0.771)
Constant	10.9***	12.4***	8.08***	8.27*
	(0.811)	(0.987)	(1.893)	(3.202)
Observations	2508	2508	1315	1315
Adjusted R^2	0.131	0.340	0.581	0.591

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

limit. Due to this relationship, the coefficients of this proportion variable should have the opposite sign of vacancy duration. Overall, the yearly vacancy variable is significant at the 0.1% level for the weighted and unweighted regressions, as seen in Appendix Table B.2. The significance of the unweighted regression is again evidence of the robustness of the effect previously found. Table 9 shows the disaggregation for STEM and non-STEM occupations. We can find the same sharp effect at Q2 for STEM occupations, and no apparent significance of the control variables. For the non-STEM counterparts, there is again a significant effect of vacancy count at Q2 and Q4, with no apparent change in magnitude due to the new variable specification. The results in terms of magnitude and significance are identical to the ones found with vacancy duration. With the results of these last two tables, we can rule out a significance effect of outlier bias in the vacancy duration variable, and thus we can put more weight on the conclusion reached with the previous estimates.

A usual criticism levied against the H-1B program is the large proportion of outsourcing firms that apply for this visa, a phenomenon explored in the Initial Results section. It is interesting to analyze as another robustness assessment if the previous results are changed dramatically with the exclusion of H-1B applications from the outsourcing industry. In Table 10, I estimate again the same regressions as in Table 7, now removing from the H-1B dataset applications from major outsourcing firms. For STEM-occupations, vacancy duration at Q2 is again significant at the 1% as seen in Regression 3 and Regression 4. Interestingly, these coefficients have a larger magnitude compared to the previous results (0.010 vs 0.011, 0.0077 vs 0.0092). This increase would indicate that vacancy duration has a slightly stronger effect for firms not associated with outsourcing. For non-STEM applications, there is no significant effect of vacancy duration but vacancy count has an effect at Q1 and Q4, similar to the previous findings. An interesting difference appears for this group in our wage control variable. While none of the previous robustness checks caused a change in the control variables, with the removal of demand from outsourcing firms, the coefficient of the median wage ratio decreases in magnitude from -1.58 to -0.38 and becomes insignificant. For H-1B demand beyond outsourcing firms, the increase in the median ratio is not associated with a decrease in H-1B numbers, which indicates a lack of sensitivity to wage increases. This result could be evidence of significant wage pressure or incentives for outsourcing firms, compared to the general sample. As mentioned in Section 5.1, an important caveat is that this sample of outsourcing firms do not capture completely H-1B outsourcing demand, only demand from the most prevalent

Table 9: Effect of Share Filled with Quarter Vacancy and Controls, by STEM status

	(1) Panel	(2) Panel	(3) W. Panel	(4) W. Panel
Share Filled Q1	-0.73*	-0.21	-0.099	-0.44
	(0.338)	(0.205)	(0.278)	(0.320)
Share Filled Q2	0.061	0.033	-1.40**	-1.29***
	(0.381)	(0.253)	(0.396)	(0.274)
Share Filled Q3	0.066	0.029	0.12	-0.19
	(0.230)	(0.228)	(0.267)	(0.132)
Share Filled Q4	0.081	-0.060	0.20	-0.13
	(0.183)	(0.236)	(0.337)	(0.359)
Vacancy Count Q1	-0.25***	-0.24***	0.093	0.16*
	(0.0599)	(0.0584)	(0.0667)	(0.0777)
Vacancy Count Q2	0.050	0.13+	0.039	-0.019
	(0.0707)	(0.0660)	(0.159)	(0.0755)
Vacancy Count Q3	0.022	-0.0069	0.046	0.0049
	(0.0818)	(0.0695)	(0.0644)	(0.0749)
Vacancy Count Q4	-0.18*	-0.16**	-0.14	-0.088
	(0.0739)	(0.0507)	(0.0919)	(0.0944)
Education Years		-0.030		-0.10
		(0.0831)		(0.0743)
Lower Quartile		0.17		-0.0071
		(0.247)		(1.513)
Median		-1.59**		-0.65
		(0.506)		(1.283)
Upper Quartile		-1.07**		1.03
		(0.329)		(0.787)
Constant	10.1***	12.6***	9.81***	10.4**
	(0.552)	(0.895)	(1.962)	(2.998)
Observations	3325	2508	1556	1315
Adjusted R^2	0.0830	0.339	0.471	0.592

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

firms. Still, however, the removal of outsourcing demand does not appear to change significantly our previous estimates of the effect of vacancy duration. This result is evidence of a true domestic effect of job market tightness that exists beyond outsourcing demand.

I will now discuss the strengths and shortcomings of the present identification strategy. Our panel empirical specification allows us to control for possible unobserved sources of bias. In the Vacancy section, I examined the possible effect of an increase of the algorithm efficiency as a source of bias, but after corroborating an occupation distribution similar to JOLTS's, I can assume this effect will be picked up by year fixed effects. Another possible omitted variable concern is the issue of H-1B approval probability. It could be argued that if employers know the chances of obtaining an H-1B visa are becoming slimmer each year, they would be more inclined to find more workers and submit more applications in order to actually be able to access them. Since the number of H-1B applications has grown substantially in the last few years, this is possibly a significant concern. However, since the probability of H-1B approval does not depend on the industry or occupation of the employee, this probability factor will be common across units of occupation. Controlling for year fixed effects will account for changes in the common probability of approval.

A similar possible concern is that, due to H-1B policy, certain industries and occupations are allowed to apply to a "masters" pool, separate from the general H-1B lottery group. Moreover, institutions like non-profits and universities are also excluded from the lottery process by the program's design. These policy considerations create an additional source of probability heterogeneity. Occupations that require master's degrees, and occupations usually hired by universities like professors will have a different probability of approval. However, since these given lottery exemptions did not change in the given timeframe of analysis, the probability heterogeneity previously described will be included in the unit fixed effects of these professions.

Continuing with the analysis of identification threats, another bias concern could be heterogeneity in hiring practices by industry or occupation. Common hiring trends to all occupations, including effects of increased technology in hiring application, screening and management, will be included in the year fixed effects, as well as common macroeconomic phenomena. Furthermore, occupation fixed effects controls for constant H-1B hiring heterogeneity in a given profession, assuming these hiring characteristics remain fixed during the regression timeframe. If, however, there are other occupation-specific changing variables affecting the hiring of H-1B workers, the coeffi-

Table 10: Effect of Vacancy Duration by STEM status, without Outsourcing demand

	(1) Panel	(2) Panel	(3) W. Panel	(4) W. Panel
Duration Q1	0.0041^{+}	0.0019	0.0016	0.0029
	(0.00223)	(0.00130)	(0.00204)	(0.00235)
Duration Q2	-0.000033	-0.00055	0.011**	0.0092***
	(0.00179)	(0.00114)	(0.00393)	(0.00221)
Duration Q3	-0.0024	-0.0012	-0.0016	-0.00018
	(0.00173)	(0.00134)	(0.00143)	(0.00140)
Duration Q4	0.000085	0.00088	0.00018	0.0021
	(0.000785)	(0.00157)	(0.00155)	(0.00153)
Vacancy Count Q1	-0.22***	-0.19***	0.081	0.17
	(0.0525)	(0.0509)	(0.0863)	(0.109)
Vacancy Count Q2	0.10	0.12	0.097	-0.00023
	(0.0633)	(0.0732)	(0.170)	(0.0882)
Vacancy Count Q3	0.039	-0.037	0.055	0.053
	(0.0815)	(0.0791)	(0.0811)	(0.0701)
Vacancy Count Q4	-0.18**	-0.15*	-0.13	-0.12
	(0.0617)	(0.0589)	(0.0871)	(0.0819)
Education Years		-0.047		-0.090
		(0.0699)		(0.0999)
Lower Quantile		0.099		2.58^{+}
		(0.510)		(1.488)
Median		-0.38		-1.48
		(0.601)		(0.933)
Upper Quantile		-1.43***		0.19
		(0.331)		(0.550)
Constant	8.44***	11.0***	7.25***	6.40*
	(0.844)	(0.881)	(1.899)	(3.089)
Observations	2987	2366	1547	1308
Adjusted R^2	0.219	0.413	0.514	0.638
STEM	No	No	Yes	Yes

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

cient of vacancy duration could be biased if these variables are also related to job market tightness. As described in the conclusion of the previous subsection, the inclusion of vacancy count controls partly for firm and occupation heterogeneity that is correlated with expansionary hiring periods.

The usual panel strategy assumes common trends across units of analysis, that is, a lack of long term differences in the growth of the dependent variable. The unobservable variant heterogeneity described above would violate the common trend assumption if it significantly extends across the years of analysis. As a related Robustness assessment, I introduce individual cluster trends that relaxes the assumption of common H-1B demand growth across occupations from different clusters. This new specification would encapsulate additional factors that drive H-1B growth and that could be correlated with vacancy duration. As such, the introduction of these trends is a stronger test on the significance of vacancy duration. The reason why cluster trends are introduced instead of occupation trends is a matter of regression power, as the number of observations prevents us from estimating accurately a single trend per occupation. Therefore, a non far-fetched assumption in place here is that occupations within a cluster share common trends. Table 11 summarizes the results with cluster time trends. We find again the effect of vacancy duration at Q2 for STEM occupations, with coefficients of magnitude 0.0070 in Regression 3. However, the p-value of the coefficient decreases and it is now only significant at the 5% level. For Non-STEM occupations, we find similar results as before, with slightly smaller significant coefficients for vacancy count at Q1 and Q4.

One caveat of introducing cluster time trends is that it requires the assumption that the trend effect for H-1B numbers is linear. This assumption could be problematic, especially within the context of a post-recession period when most macroeconomics trends of recovery follow non-linearity. As an alternative to cluster time trends, we can also analyze the effect of cluster-specific year fixed effects. A downside of this non-linear control of trend effects is the rapidly decreasing power of the regression due to the introduction of a large number of extra fixed effects. In terms of the general sample, we can conclude using Appendix Table B.4 that a significant yearly vacancy duration effect is present in both the weighted and unweighted specifications, even after including the specific cluster time effects. The coefficient is 0.0018 in the unweighted specification and 0.0028 in the weighted one, both significance at the 5% level. This result is further evidence of the robustness of our conclusion concerning the general vacancy duration effect found in section 5.2. In terms of

Table 11: EEffect of Vacancy Duration at Code Aggregation, by STEM Status, with Cluster Time Trends

	(1) Panel	(2) Panel	(3) W. Panel	(4) W. Panel
Duration Q1	0.0045^{+}	0.0020^{+}	0.0032^{+}	0.0036
	(0.00239)	(0.00112)	(0.00181)	(0.00258)
Duration Q2	0.00020	-0.00035	0.0070*	0.0068**
	(0.00215)	(0.00126)	(0.00306)	(0.00228)
Duration Q3	-0.0028^{+}	-0.0011	-0.000073	0.0022^{+}
	(0.00158)	(0.00141)	(0.00117)	(0.00121)
Duration Q4	0.0020*	0.0018	-0.0013	0.00054
	(0.000945)	(0.00133)	(0.00236)	(0.00120)
Vacancy Count Q1	-0.15***	-0.19*	0.13^{+}	0.12
	(0.0396)	(0.0785)	(0.0710)	(0.0877)
Vacancy Count Q2	0.034	0.097	0.18	0.13
	(0.0651)	(0.0582)	(0.225)	(0.146)
Vacancy Count Q3	0.079	-0.030	0.0084	-0.046
	(0.0819)	(0.0751)	(0.0438)	(0.0512)
Vacancy Count Q4	-0.18*	-0.14*	-0.053	0.013
	(0.0759)	(0.0537)	(0.0400)	(0.0758)
Education Years		-0.044		0.033
		(0.112)		(0.0467)
Lower Quartile		0.11		0.48
		(0.264)		(1.441)
Median		-1.46**		-1.05
		(0.460)		(1.329)
Upper Quartile		-0.88**		1.25
		(0.306)		(0.758)
Constant	-69.1*	-10.6	-183.8***	-235.8***
	(26.43)	(21.84)	(29.76)	(26.73)
Observations	3325	2508	1556	1315
Adjusted R^2	0.267	0.439	0.578	0.714

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Effect of Vacancy Duration at Code Aggregation, by STEM Status, with Cluster Time Dummies

	(1) Panel	(2) Panel	(3) W. Panel	(4) W. Panel
Duration Q1	0.0057**	0.0031**	0.0043+	0.0022
	(0.00207)	(0.00116)	(0.00222)	(0.00234)
Duration Q2	0.0023	0.0011	0.0062	0.0025+
	(0.00332)	(0.00183)	(0.00423)	(0.00144)
Duration Q3	-0.0029*	-0.00033	-0.00012	0.0025*
	(0.00140)	(0.00126)	(0.00168)	(0.00114)
Duration Q4	0.00063	0.0011	-0.0027	0.00023
	(0.000782)	(0.00217)	(0.00311)	(0.00129)
Vacancy Count Q1	-0.15**	-0.17	0.098	-0.021
	(0.0476)	(0.104)	(0.0760)	(0.0699)
Vacancy Count Q2	-0.049	0.039	0.24	0.094
	(0.0740)	(0.0767)	(0.300)	(0.115)
Vacancy Count Q3	0.11	-0.0074	0.0038	-0.018
	(0.132)	(0.0684)	(0.0523)	(0.0703)
Vacancy Count Q4	-0.17*	-0.16**	-0.045	0.086
	(0.0751)	(0.0568)	(0.0287)	(0.0903)
Education Years		-0.11		0.043
		(0.129)		(0.0642)
Lower Quartile		0.14		0.44
		(0.289)		(1.416)
Median		-1.53**		0.28
		(0.574)		(1.532)
Upper Quartile		-0.97*		0.91
		(0.441)		(1.169)
Constant	8.74***	12.7***	6.46*	5.52+
	(0.888)	(1.464)	(2.679)	(2.894)
Observations	3325	2508	1556	1315
Adjusted R^2	0.339	0.520	0.634	0.791

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

STEM and non-STEM dynamics, Table 12 summarizes these results for the weighted regressions. The effect of vacancy duration at Q2 for STEM applications is 0.0062 in Regression 3, but it is not statistically significant most likely due to the loss of regression power. In Regression 4, the coefficient decreases in magnitude to 0.0025 and it is significant at the 10% level. Even though the point estimate (0.0062) indicates a similar effect as the ones previously found, with this specification we cannot conclude a significant effect of vacancy duration. A different dynamic occurs with non-STEM occupations. While in all of our previous estimates, we found no significant effect of the vacancy duration variables, in Regression 1 and 2 of this table we can find marked vacancy duration effects at Q1. Similar vacancy count effects described before are also found. Non-STEM regressions are not as affected as the STEM group by the loss in power due to the larger sample size, which allows for certain precision to remain in the standard errors. To summarize, the duration effect at Q2 of STEM applications is similar in magnitude and significance after including cluster time trends, but its effect cannot be distinguished from the null under the cluster-specific fixed effects specification. For Non-STEM applications, the vacancy count effects at Q1 and Q4 are found again with trend control. Moreover, the introduction of the trend control mechanisms points out to a significant effect at Q1, a job tightness effect not picked up before. This could potentially indicate that both STEM and non-STEM professions are equally affected by vacancy duration at the earlier periods, although this result is not as robust as our previous conclusions.

As a final note on the possible threats of the estimation, it could also be argued that this identification strategy presents a problem of simultaneous causality. Assuming vacancy duration causes an increase in the number of necessary H-1B workers present, this foreign labor supply could impact the domestic labor market and possibly the vacancy measures themselves. I counter this argument with two points. First, given the timing specification outlined in the Method section, the vacancy measures always correspond to periods before the submission of the bulk of the H-1B applications. A small percent of applications does exist in the in-between months of the application cycle, but in any case these numbers do not represent true demand at this point in time because of the bunching that occurs during March. Secondly, after the submission of applications, the visas are not approved until October. Therefore, the workers would not be introduced into the labor supply pool at least after a year of our latest vacancy time period, Q4. The increase of H-1B workers cannot affect retroactively tightness of the labor market. There could be a year to year effect in vacancy duration, as the domestic labor market becomes less tight due to the influx of H-1B workers. However, at any point in time, the decision to apply for H-1B workers cannot affect contemporaneously the present condition of labor market tightness. As such, the temporal specification of the identification strategy removes major simultaneity concerns.

6 Conclusion

The present paper attempts to find a link between domestic labor market tightness and the number of H-1B visas demanded. I find a significant effect of the vacancy duration during a given application cycle's previous year. Moreover, it appears only STEM occupations are affected by vacancy duration, while non-STEM applications appear to respond negatively to the number of domestic vacancies. After accounting for outsourcing firm demand, outliers in our vacancy measures and possible deviations from the common trend assumption, the results remain significant.

These results have several implications for policy. It gives strong evidence of a H-1B visas serving a domestic need giving given labor market constraints, specifically for STEM occupations. While the increase in outsourcing demand is concerning, this phenomenon does not appear to change our estimations of the necessity of the program. Given the continuous increase in demand for H-1B visas, a domestic policy that offers greater access to these visas would be satisfying an authentic requirement in domestic markets. At the same time, enforcement agencies and regulators should pay more attention to the outsourcing phenomenon. An effective reform of the program would aim to curb and regulate the demand from these firm while ensuring firms are available to satisfy their labor necessities appropriately.

Further avenues of research remain in regards to this research question. The availability of employer information could allow us to control for employer's heterogeneity at the firm level. This employer focus could give greater evidence of the effect of the program at the firm level, and would also generate greater dialogue with the present literature of H-1B effects on the firm. Similarly, it is also possible to analyze the effect of regional market conditions instead of national aggregate relations, by using the geographical information on vacancies and the location information of the H-1B dataset. This set-up could allow to test whether there are regional variables affecting demand beyond simply the national market characteristics. The regional approach would also overcome the degree of freedom limitation of some of the present regressions, and would also contribute with present research of the effects of skilled labor in metropolitan areas.

References

Baert, Stijn, Bart Cockx, Niels Gheyle, and Cora Vandamme, "Is There Less Discrimination in Occupations Where Recruitment is Difficult?," *ILR Review*, May 2015, *68* (3), 467–500.

Bound, John, Gaurav Khanna, and Nicolas Morales, "Understanding the Economic Impact of the H-1B Program on the U.S.," Working Paper 23153, National Bureau of Economic Research February 2017.

Clifford, Robert, "Demand for H-1B Visas in New England: An Analysis of Employer Requests for Highly-Skilled Guest Workers," October 2014.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger, "The Establishment-Level Behavior of Vacancies and Hiring," Working Paper 16265, National Bureau of Economic Research August 2010.

Dayton, Toby, "LinkUp Data Versus Conference Board's HWOL Data Series & Help Wanted Analytics/CEB Data," August 2016.

Doran, Kirk, Alexander Gelber, and Adam Isen, "The Effects of High-Skilled Immigration Policy on Firms: Evidence from H-1B Visa Lotteries," Working Paper 20668, National Bureau of Economic Research November 2014.

Ghosh, Anirban, Anna Maria Mayda, and Francesc Ortega, "The Impact of Skilled Foreign Workers on Firms: An Investigation of Publicly Traded U.S. Firms," December 2014.

Gorter, Cees, Wolter H. J. Hassink, and Giovanni Russo, "The Structure of Hiring and Labour Market Tightness," *Economics Letters*, July 2003, *80* (1), 99–103.

Hira, Ron, "Top 10 users of H-1B guest worker program are all offshore outsourcing firms," February 2014.

IAOP, "The Global Outsourcing 100."

Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln, "Skilled Immigration and the Employment Structures of U.S. Firms," Working Paper 19658, National Bureau of Economic Research November 2013.

Kerr, William R. and William F. Lincoln, "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention," Working Paper 15768, National Bureau of Economic Research February 2010.

Kight, Stef, "H-1B visa application cap hit within 5 days," *Axios*, April 2018.

Kirkegaard, "Outsourcing and Skill Imports: Foreign High-Skilled Workers on H-1B and L-1 Visas in the United States," Technical Report, Peterson Institute for International Economics December 2005.

Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber, "The Effect of the H-1B Quota on Employment and Selection of Foreign-Born Labor," Working Paper 23902, National Bureau of Economic Research October 2017.

Moss, Phillip and Chris Tilly, "How Labor-Market Tightness Affects Employer Attitudes and Actions Toward Black Job Applicants: Evidence From Employer Surveys," in "Prosperity For All?: The Economic Boom and African Americans," Russell Sage Foundation, 2000.

Mukhopadhyay, Sankar and David Oxborrow, "The value of an employment-based green card," *Demography*, February 2012, *49* (1), 219–237.

O*NET, "O*NET OnLine Help: Browse by Career Cluster."

Park, Haeyoun, "How Outsourcing Companies Are Gaming the Visa System," *The New York Times*, November 2015.

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.

Raphelson, Samantha, "Trucking Industry Struggles With Growing Driver Shortage," *NPR*, January 2018.

Rothwell, Jonathan, "Still Searching: Job Vacancies and STEM Skills," Technical Report, Brookings Institute July 2014.

Rothwell, Jonathan T. and Neil Ruiz, "H-1B Visas and the STEM Shortage: A Research Brief," SSRN Scholarly Paper ID 2262872, Social Science Research Network, Rochester, NY May 2013.

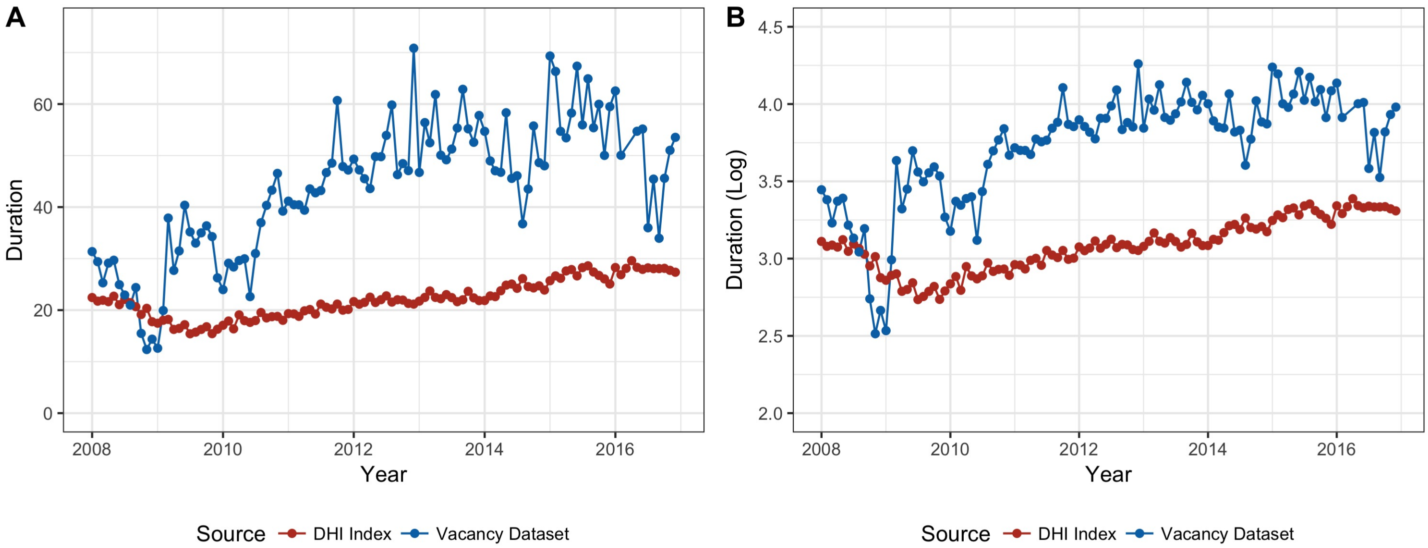
Torres, Nicole, "The H-1B Visa Debate, Explained," *Harvard Business Review*, May 2017.

USCIS, "H-1B Specialty Occupations, DOD Cooperative Research and Development Project Workers, and Fashion Models," April 2018.

Wilson, Jill, Neil Ruiz, and Shyamali Choudhury, "The Search for Skills: Demand for H-1B Immigrant Workers in U.S. Metropolitan Areas," July 2012.

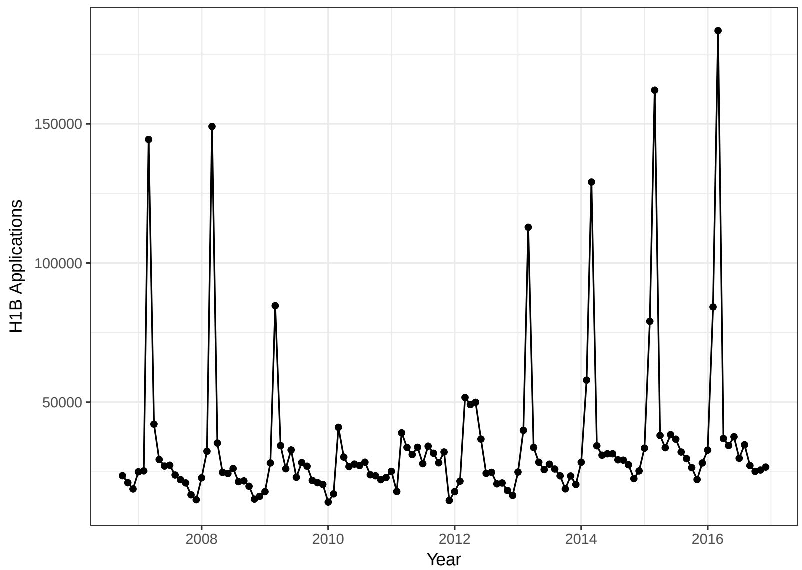
Appendix A Figures

Figure A.1: Comparison of Mean Online Job Posting and DHI-DHF Index, Normal and Log Scale



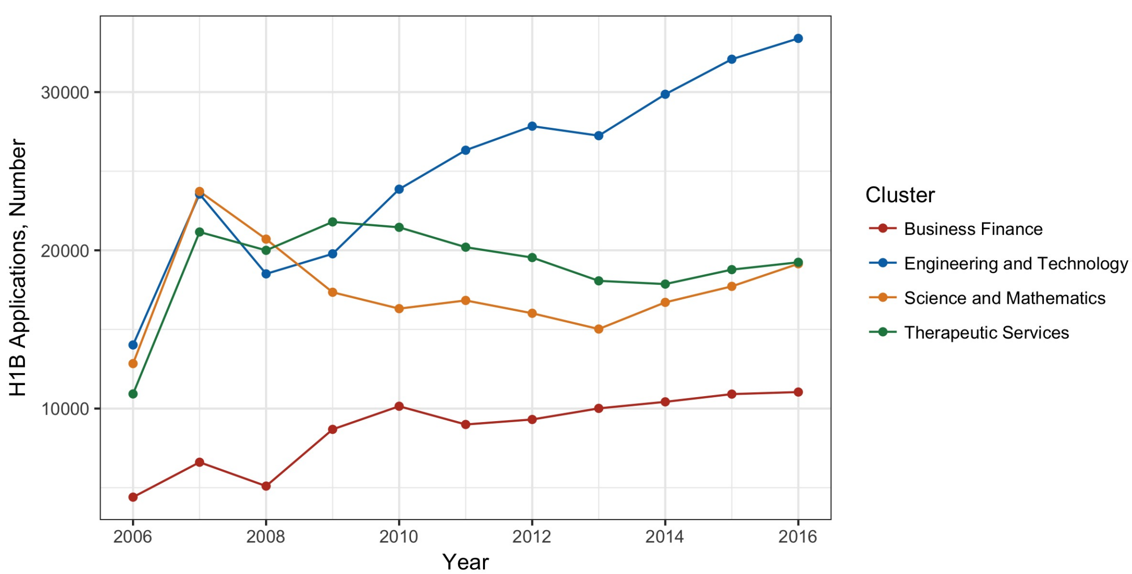
Note: Data on online job postings obtained from LinkUp from 2008 to 2016. Each data point is the mean value of the duration of total job postings that were started in a given month. The DHI-DHF Vacancy Duration Indicator is a measurement of vacancy duration from the DHI Group, based on the model of Davis, Faberman and Haltiwanger. Graph A show the measure of vacancies in their absolute values. Graph B shows the same measures in log scale.

Figure A.2: Number of H-1B Applications by Month of Submission



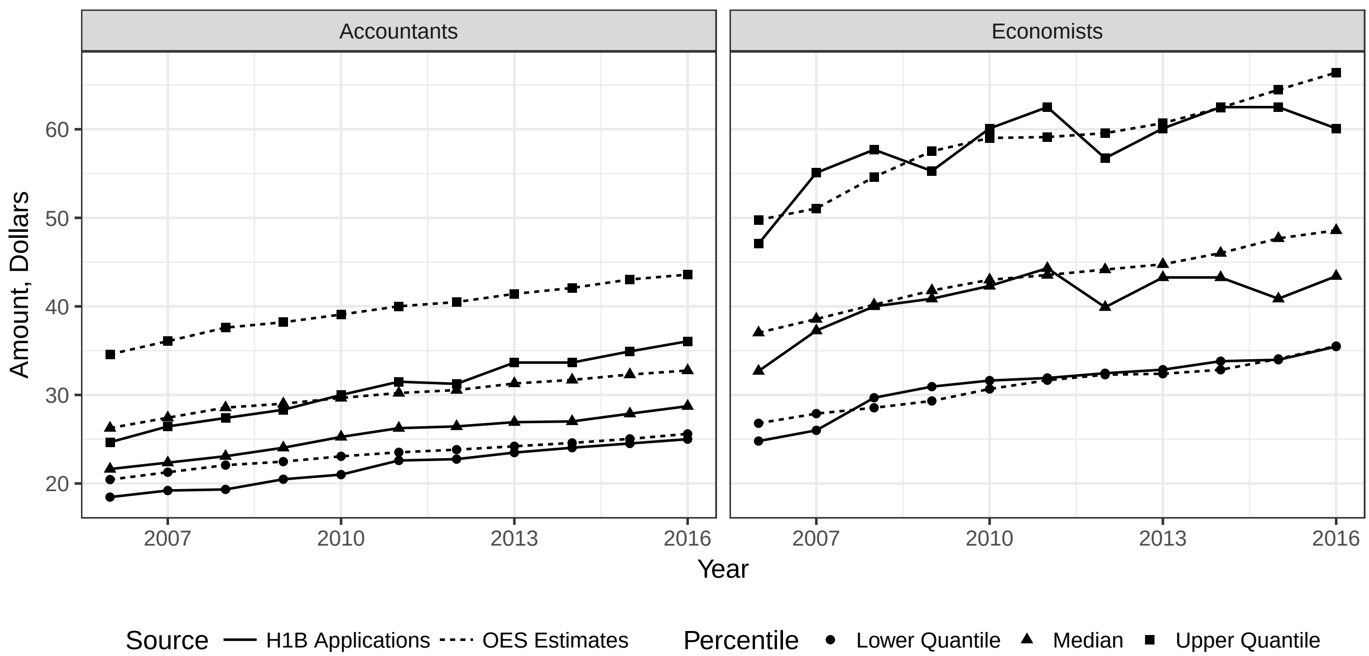
Note: Each data point is the number of total Labor Condition Applications submitted in a given month between January 2008 and December 2016. Data on LCAs is available publicly on the website of the Office of Foreign Labor Certification.

Figure A.3: Number of H-1B Applications by Cluster, Selected



Note: Each data point is the number of Labor Condition Applications between April 1st of a given year and March 31st of the next year. Labor Condition Applications are publicly available on the Office of Foreign Labor Certification website. Using each observation's SOC code, the LCAs were classified into clusters using the "Career Pathway" classification of O*NET. The graph shows data for 4 of the most representative clusters.

Figure A.4: Distribution of H-1B Proposed Wages and Domestic Wages, Selected



Note: Each Labor Condition Application specifies a "Proposed" wage. The full lines represent the yearly "Proposed" wage quartiles for Accountants and Economists in the LCA dataset, from 2006 to 2016. The dashed lines are estimates for domestic wage quartiles for Accountants and Economists. The estimates are obtained from the Office of Employment Statistics.

Appendix B Tables

Table B.1: Sample Occupations with their respective Codes and Clusters

Cluster	Code	Occupation
Agribusiness Systems	13-1021.00	Buyers and Purchasing Agents, Farm Products
Agribusiness Systems	11-9013.02	Farm and Ranch Managers
Agribusiness Systems	13-1074.00	Farm Labor Contractors
Agribusiness Systems	11-9013.00	Farmers, Ranchers, and Other Agricultural Managers
Agribusiness Systems	11-9013.01	Nursery and Greenhouse Managers
Animal Systems	45-2021.00	Animal Breeders
Animal Systems	19-1011.00	Animal Scientists
Animal Systems	45-2093.00	Farmworkers, Farm, Ranch, and Aquacultural Animals
Animal Systems	39-2021.00	Nonfarm Animal Caretakers
Environmental Service Systems	17-3025.00	Environmental Engineering Technicians
Environmental Service Systems	17-2081.00	Environmental Engineers
Environmental Service Systems	19-4091.00	Environmental Science and Protection Technicians
Environmental Service Systems	47-4041.00	Hazardous Materials Removal Workers
Environmental Service Systems	37-2021.00	Pest Control Workers
Environmental Service Systems	53-7081.00	Refuse and Recyclable Material Collectors
Environmental Service Systems	51-8031.00	Water and Wastewater Treatment Operators
Environmental Service Systems	17-2081.01	Water/Wastewater Engineers

Table B.2: Effect of Share Filled at Code Aggregation with Quarter Vacancy, Controls

	(1) Panel	(2) Panel	(3) Panel	(4) W. Panel	(5) W. Panel	(6) W. Panel
Percent Filled	-0.40***			-0.73***		
	(0.113)			(0.205)		
Vacancy Count	-0.045			-0.28***		
	(0.0727)			(0.0614)		
Share Filled Q1		-0.37***	-0.29**		-0.45+	-0.44+
		(0.0990)	(0.0893)		(0.229)	(0.224)
Share Filled Q2		0.016	-0.091		-0.77*	-0.86**
		(0.120)	(0.118)		(0.304)	(0.272)
Share Filled Q3		-0.056	-0.12		0.15	-0.042
		(0.150)	(0.146)		(0.256)	(0.159)
Share Filled Q4		0.0050	-0.030		0.026	-0.27
		(0.0903)	(0.0966)		(0.242)	(0.350)
Vacancy Count Q1		0.0058	0.0018		-0.036	-0.019
		(0.0210)	(0.0246)		(0.0383)	(0.0525)
Vacancy Count Q2		0.014	0.042		-0.037	-0.041
		(0.0467)	(0.0646)		(0.0781)	(0.0474)
Vacancy Count Q3		-0.017	-0.057		0.069	-0.018
		(0.0654)	(0.0381)		(0.0542)	(0.0606)
Vacancy Count Q4		-0.023	-0.027		-0.27**	-0.26***
		(0.0333)	(0.0310)		(0.0806)	(0.0602)
Education Years			-0.026			-0.12*
			(0.0641)			(0.0566)
Lower Quartile			0.14***			0.12
			(0.0317)			(0.734)
Median			0.033			-0.89
			(0.211)			(0.714)
Upper Quartile			-0.87***			0.52
			(0.149)			(0.618)
Constant	4.63***	4.50***	6.03***	12.1***	12.0***	14.0***
	(0.652)	(0.653)	(0.372)	(0.771)	(0.715)	(1.710)
Observations	4956	4883	3823	4943	4881	3823
Adjusted R^2	0.0855	0.0975	0.186	0.347	0.355	0.459
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: Effect of Vacancy Duration at Code Aggregation with Quarter Vacancy, without Outsourcing Demand

	(1) Panel	(2) Panel	(3) Panel	(4) W. Panel	(5) W. Panel	(6) W. Panel
Vacancy Duration	0.0018**			0.0051**		
	(0.000647)			(0.00159)		
Vacancy Count	-0.051			-0.21***		
	(0.0835)			(0.0575)		
Duration Q1		0.0022***	0.0017**		0.0032*	0.0028+
		(0.000565)	(0.000605)		(0.00124)	(0.00157)
Duration Q2		-0.00019	0.00036		0.0061*	0.0058**
		(0.000703)	(0.000861)		(0.00278)	(0.00216)
Duration Q3		-0.00024	0.00079		-0.0018	-0.00072
		(0.00117)	(0.000707)		(0.00141)	(0.00114)
Duration Q4		-0.00036	0.000075		0.00083	0.0025
		(0.000596)	(0.000796)		(0.00115)	(0.00162)
Vacancy Count Q1		-0.0071	0.017		-0.041	0.0066
		(0.0266)	(0.0226)		(0.0432)	(0.0490)
Vacancy Count Q2		0.035	0.038		0.042	-0.011
		(0.0512)	(0.0590)		(0.0820)	(0.0507)
Vacancy Count Q3		-0.010	-0.068+		0.062	0.0023
		(0.0694)	(0.0351)		(0.0667)	(0.0457)
Vacancy Count Q4		-0.045	-0.029		-0.23**	-0.24**
		(0.0319)	(0.0294)		(0.0848)	(0.0762)
Education Years			-0.025			-0.093
			(0.0614)			(0.0830)
Lower Quantile			1.04***			2.00+
			(0.173)			(1.112)
Median			-0.62***			-1.45+
			(0.136)			(0.740)
Upper Quantile			-0.85***			-0.039
			(0.160)			(0.462)
Constant	3.98***	3.85***	4.96***	9.95***	9.24***	10.1***
	(0.759)	(0.735)	(0.330)	(0.610)	(0.703)	(2.034)
Observations	4606	4536	3674	4593	4534	3674
Adjusted R^2	0.0670	0.0686	0.163	0.411	0.420	0.527
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4: Effect of Vacancy Duration at Code Aggregation with Quarter Vacancy, with Cluster Time Dummies

	(1) Panel	(2) Panel	(3) Panel	(4) W. Panel	(5) W. Panel	(6) W. Panel
Vacancy Duration	0.0018*			0.0028*		
	(0.000730)			(0.00142)		
Vacancy Count	-0.029			-0.011		
	(0.0363)			(0.168)		
Duration Q1		0.0022***	0.0016*		0.0049***	0.0037**
		(0.000571)	(0.000673)		(0.00118)	(0.00118)
Duration Q2		0.00062	0.00043		0.0039+	0.0031
		(0.000798)	(0.000922)		(0.00233)	(0.00191)
Duration Q3		0.00011	0.00085		-0.00086	0.0013
		(0.00123)	(0.000834)		(0.00164)	(0.00169)
Duration Q4		-0.00057	-0.00034		-0.0021	0.00021
		(0.000548)	(0.000561)		(0.00197)	(0.000879)
Vacancy Count Q1		-0.018	-0.038		0.017	-0.064+
		(0.0333)	(0.0430)		(0.0618)	(0.0374)
Vacancy Count Q2		-0.0054	0.021		0.14	0.15*
		(0.0326)	(0.0563)		(0.190)	(0.0736)
Vacancy Count Q3		0.022	-0.017		0.020	-0.095*
		(0.0571)	(0.0348)		(0.0671)	(0.0420)
Vacancy Count Q4		0.0021	-0.0034		-0.099	-0.048
		(0.0374)	(0.0352)		(0.0655)	(0.0782)
Education Years			-0.022			-0.0079
			(0.0589)			(0.0913)
Lower Quartile			0.17***			0.24
			(0.0392)			(0.754)
Median			0.050			-0.14
			(0.202)			(0.452)
Upper Quartile			-0.69***			0.48
			(0.142)			(1.010)
Constant	3.35***	3.12***	4.43***	8.73***	8.07***	8.39*
	(0.293)	(0.318)	(0.592)	(1.668)	(2.051)	(3.381)
Observations	4956	4883	3823	4943	4881	3823
Adjusted R^2	0.227	0.237	0.316	0.543	0.554	0.666
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$