Volume Title: The Roles of Immigrants and Foreign Students in US Science, Innovation, and Entrepreneurship

Volume Authors/Editors: Ina Ganguli, Shulamit Kahn, Megan MacGarvie, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-69562-4 (cloth); 978-0-226-69576-1 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/roles-immigrants-and-foreign-students-us-science-innovation-and-entrepreneurship

Conference Date: April 27, 2018

Publication Date: February 2020

Chapter Title: New Data and Facts on H-1B Workers across Firms

Chapter Author(s): Anna Maria Mayda, Francesc Ortega, Giovanni Peri, Kevin Shih, Chad Sparber

Chapter URL: https://www.nber.org/books-and-chapters/roles-immigrants-and-foreign-students-us-science-innovation-and-entrepreneurship/new-data-and-facts-h-1b-workers-across-firms

Chapter pages in book: (p. 99 – 121)

4

New Data and Facts on H-1B Workers across Firms

Anna Maria Mayda, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber

4.1 Introduction

Several researchers are using administrative data on petitions for H-1B workers (also known as I-129 forms) in their analyses of high-skilled immigrants in the United States. While potentially very useful, to date there has been no systematic analysis of the validity of these data. Such an exercise is important because these data are released without a detailed codebook and were not originally designed for use in academic research.

We obtained microdata from United States Citizenship and Immigration Services (USCIS) through a Freedom of Information Act (FOIA) request. These data contain the universe of approved petitions for H-1B workers, along with a substantial (though incomplete) number of denied petitions received during the period 1997–2012. The data set contains 3.72 million cases corresponding to roughly 300,000 companies.

Previous studies (e.g., Kerr and Lincoln 2010; Ghosh, Mayda, and Ortega

Anna Maria Mayda is currently senior advisor to the chief economist at the US State Department. She is on leave from Georgetown University, where she is an associate professor of economics. Francesc Ortega is the Dina Axelrad Perry Professor in Economics at Queens College CUNY. Giovanni Peri is professor and chair of the Department of Economics at the University of California, Davis, and a research associate of the National Bureau of Economic Research. Kevin Shih is assistant professor of economics at Queens College CUNY. Chad Sparber is professor of economics at Colgate University.

We thank Ina Ganguli, Jennifer Hunt, Shulamit Kahn, and Megan MacGarvie and all participants at the NBER conference on the Role of Immigrants and Foreign Students in Science, Innovation, and Entrepreneurship for helpful comments. All views contained herein are the authors' own. We are grateful to the National Science Foundation (award 1535561) for generously funding this project. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/chapters/c14106.ack.

2014, among others) have relied on data on labor condition applications (LCAs), which need to be filed by any company intending to hire H-1B workers. In contrast to our I-129 data set, LCA data are publicly available from the Department of Labor. While LCAs are a useful proxy for a firm's general interest in hiring H-1B workers, they are much less useful as a measure of how many H-1B petitions that firm files or how many approvals it eventually obtains. The reason is that firms can file LCAs at virtually no cost, and there is an advantage in keeping LCA applications even if hiring foreign workers is simply one of many options. There is no LCA filing fee, for example, and LCA approval does not commit firms to subsequently conduct a job search. As a result, many companies submit LCA paperwork requesting approval to hire far more H-1B workers than they actually intend to hire.[1] In contrast, the H-1B data are worker-specific and necessarily imply that a firm has performed a job search and identified suitable candidates. Hence it is much closer to the concept of "vacancy" or "labor demand" for a firm. Moreover, each petition is accompanied by a positive (and substantial) marginal cost in the form of an I-129 filing fee.[2]

This chapter has three goals. First, we examine the validity of the administrative USCIS microdata on petitions for H-1B workers by comparing these data to the aggregate totals published in the USCIS annual reports on *Petitions and Characteristics of the H-1B Population*. After showing that the microdata are highly consistent with the aggregate statistics, we use string-matching techniques to build a longitudinal, company-level data set for approved H-1B petitions. This turned out to be a very arduous process, and our results in this chapter represent a preliminary summary of work in progress. Nonetheless, we describe a number of important facts in these data, distinguishing between applications for initial employment and those for continuing employment at the firm level. Last, we match our data set on approved petitions to Compustat data on all publicly traded companies. The resulting panel data set contains a wealth of information on firm-level outcomes along with the number of yearly approved H-1B petitions. We use this data set to compare the characteristics of Compustat companies that received H-1B workers to those that did not and describe trends at the industry level in H-1B usage.

Our main findings are as follows. First, we show that the annual counts of petitions in the microdata closely match the totals in the USCIS reports for most, though not all, years. We also show that the microdata account fairly well for the total numbers of approved petitions, with a higher degree

1. The LCA data show multiple instances of companies that request the exact same number of applications every year for several years.

2. Originally, we intended to use the number of LCAs filed by a company in a particular year together with the number of approved H-1B petitions to build firm-specific annual success rates in order to exploit the randomization introduced by the lottery assignment. However, for the reasons outlined earlier, we abandoned such an approach.

of accuracy when focusing on issuances for initial employment (as opposed to continuing-employment applications).

Next we establish the following facts on the three million approved H-1B petitions in the period 2000–2012. First, 46 percent of all initial-employment H-1Bs were issued to workers in computer-related occupations. The bulk of the remaining approved petitions were issued to firms hiring managers, officials, and occupations in administrative specializations (13 percent); architects and engineers (11.3 percent); education-related occupations (9.9 percent); and workers in occupations in medicine and health (6.3 percent). Second, about 1 in 5 approved petitions for initial-employment originated in the metropolitan area of *New York/Northeastern New Jersey*. Other important metropolitan areas were *San Jose, CA*; *Washington, DC/MD/VA*; *Boston, MA/NH*; *Chicago, IL*; and *Dallas–Fort Worth, TX*. Together, these six metropolitan areas account for 60 percent of all initial-employment petitions. Third, our firm-level data set contains approximately 398,000 companies with an annual average for approved petitions of 1.6 for initial-employment and 1.9 for continuing employment. Fourth, we document a very large increase in the concentration of approved petitions. The data show a fourfold increase in the top-20 share for new-employment H-1B petitions over the period 2000–2012, with a sharp acceleration between 2008 and 2012. During this period, we also observe a clear trend toward a ranking dominated by global IT consulting companies. Fifth, public school districts and research universities enter the top-20 ranking in some years. Among not-for-profit institutions, in most years the top petitioner for initial-employment H-1B workers was the New York City Public School District.

Regarding publicly traded (Compustat) companies, our data reveal the following facts. Compustat companies account for about 13 percent of all approved petitions in our data set. Roughly 42 percent of Compustat companies had at least one approved petition over the period 2000–2012, and in any given year, only 20 percent of Compustat companies had at least one approved petition for an initial-employment H-1B. We also find that firms using the H-1B program are larger on average and have higher growth rates than nonusers. In our data, the main H-1B-receiving industries are business services, electronic equipment, and machinery and computers. The data also show the explosion in the number of new-employment H-1Bs received by the business services sector between 2009 and 2012. Moreover, this growth has been largely driven by an increase in the intensity of H-1B use (relative to overall employment in the industry) as opposed to an increase in the size of the industry. Between 2000 and 2008, the business services industry received about 1.5 initial-employment issuances per 1,000 employees. However, this intensity grew by 133 percent between 2008 and 2012.

This chapter is most directly related to the growing research on the economic effects of the H-1B program. Some studies have focused on the impact on innovation and patenting (Hunt and Gauthier-Loiselle 2010; Kerr and

Lincoln 2010; Kerr, Kerr, and Lincoln 2015). In our use of string-matching techniques, our chapter is closely related to the studies aimed at linking patenting data to other firm-level data sets (such as Compustat), as in Hall, Jaffe, and Trajtenberg (2001) and Bessen and Hunt (2007). Others have focused on labor market effects (Peri, Shih, and Sparber 2015; Mayda et al. 2018), company performance (Doran, Gelber, and Isen 2014; Ghosh, Mayda, and Ortega 2014), or educational and career choices (Kato and Sparber 2013; Amuedo-Dorantes and Furtado 2016; Shih 2016).

To our knowledge, this chapter, together with Mayda et al. (2018), is the first to utilize data on the universe of H-1B visas at the firm level (over several years). The other papers in the literature on H-1B visas either use aggregate data (e.g., state-level or city-level data) or use data on labor condition applications (LCAs) or focus on a small sample of firms. Hunt and Gauthier-Loiselle (2010) exploit cross-state variation for the United States and find that a 1 percentage point increase in the share of immigrant college graduates in the population leads to an increase in patents per capita of 9 percent to 18 percent—the main reason being that they disproportionately hold science, technology, engineering, and mathematics (STEM) degrees. Peri et al. (2015) use variation in the H-1B cap to identify the effect of increases in the population of STEM workers in a city on the wages of skilled and unskilled workers in the same city. This chapter finds that H-1B-driven increases in STEM workers are associated with increases in the wages paid to skilled workers (in both STEM and non-STEM occupations) and finds no evidence of effects on the wages of unskilled workers.

Kerr and Lincoln (2010) focus on the effects of H-1B visas on patenting activity and carry out the analysis, for the most part, at the city level—the firm-level analysis in Kerr and Lincoln (2010) is based on LCA data for a very small sample of companies (77 firms). On the other hand, Ghosh et al. (2014) use data on LCAs for the universe of publicly traded firms in the US (almost 4,000 firms) and investigate a different set of firm-level outcomes, including firm productivity, sales, profits, and total-factor productivity. Kerr, Kerr, and Lincoln (2015) exploit the same empirical strategy as Kerr and Lincoln (2010) to analyze the impact of hiring young skilled immigrants on the hiring and employment of several groups of skilled native workers. This chapter uses administrative microdata from the US Census Bureau, which is extremely accurate. However, as in Kerr and Lincoln (2010), the focus is on a subset of firms—specifically an unbalanced panel of 319 firms selected on the basis of employment and patenting activity.

An important recent contribution to the literature is the work by Doran et al. (2014), which exploits the visa lottery in fiscal years 2006 and 2007 to analyze the effects of H-1B visas on patenting and overall firm employment. This paper finds no evidence of an effect on patenting and at most a moderate effect on overall employment in the firm. Clemens (2013) analyzes

internal personnel data from an anonymous Indian-based IT firm to study the effects on earnings for workers who migrate to the US on H-1B status relative to those who remain in India. He finds a large effect stemming primarily from the change in location. It has been argued that H-1B status holders are tied to their employers and subject to some degree of exploitation. Depew et al. (2013) revisit this question by focusing on worker separations in a data set containing six large Indian IT firms. They show that quit rates are significant and procyclical, suggesting a substantial degree of mobility toward other employers.

The structure of this chapter is as follows. Section 4.2 describes our microdata on H-1B petitions. Section 4.3 describes the procedure to create the company-level data set on approved petitions. Section 4.4 summarizes the procedure to match the H-1B data to Compustat and presents the main facts arising from these data. Section 4.5 concludes.

4.2 H-1B Petitions for 1997–2012

4.2.1 Data Source

The starting point of our analysis is a microdata set provided by USCIS (through an FOIA request) on the universe of processed I-129 petitions for H-1B workers from 1997 to 2012. H-1B status provides foreign citizens a legal right to temporarily work in highly skilled specialty occupations in the United States. Although it is awarded to individuals, a person must have a qualifying job offer to receive H-1B status, and the I-129 petition for H-1B employment is filed by the employer. Thus the program creates a strong employer/employee link. This motivates us to create a firm-level data set on H-1B employment.

Our data set contains 3.72 million individual petitions for H-1B employment. Petitions for fiscal years 1997 and 1998 are severely incomplete for unknown reasons, and we do not use them in our analysis.[3] Each petition provides the date on which it was received as well as the status date and decision (i.e., if the H-1B application was approved, denied, rejected, pending, or administratively closed). In principle, all approved H-1Bs are included in our data set. We have limited information on nonapproved petitions, however. This is because new H-1B issuances have been subject to an annual cap since the program's inception. Cap exemptions exist for H-1B renewals and employees of universities and nonprofit research institutions. But USCIS stops processing and recording petitions for cap-bound new H-1B employment after the annual cap has been reached, so these unprocessed petitions

3. The 1,501 petitions for fiscal year (FY) 1997 and 21,324 for FY 1998 account for only 0.61 percent of all petitions in our data.

are not in our data set. Among the 3.64 million petitions processed in fiscal years 1999–2012, 82.4 percent (3 million) were approved.[4]

Our data set includes individual- and firm-level information for each petition. Firm-level information includes company name, state, and zip code. In theory, it also identifies whether the employer is a cap-exempt educational or nonprofit research organization. Individual-level information includes country of birth, age, education level, salary, occupation, and principal field of study. It also identifies whether the individual is requesting new H-1B status (24.4 percent), a change in status (24.1 percent), an extension of an existing H-1B status (49.6 percent), or an amendment (1.7 percent). Petitions can be for new employment (55.7 percent), continuation of employment (27 percent), change in previous approved employment (7.1 percent), change of employer (8.2 percent), or an amendment (1.5 percent).

We use this information to distinguish between petitions for new employment (which can be cap-bound) and for cap-exempt continuing employment. Specifically, we define a petition to be for *initial employment* when (a) the applicant's job status is new employment and (b) the petition is not requesting an extension or an amendment of an existing H-1B. Among the 3 million *approved* petitions, 1.60 million were for new employment. Among these, 251,000 petitions requested either an extension or an amendment of established H-1B employment. Thus according to our definition, 1.35 million approved petitions were for *initial employment*. We refer to all other approved H-1Bs (1.65 million) as pertaining to *continuing employment*.

4.2.2 Comparison with USCIS Reports

Validation. The data on petitions (I-129 forms) we obtained from USCIS lacked detailed documentation and had some awkward features. It is therefore important to check validity. To do so, we compare our microdata to the reports published annually by USCIS (*Petitions and Characteristics of the H-1B Population*). We restrict our comparison to fiscal years 2000–2012.

The figures in the annual reports correspond to the figures of USCIS in terms of H-1B petitions, filings, and approvals. The timing of their data is not directly linked to the lotteries or application deadlines in any given year. In our microdata, for each petition we know the receipt date and a status date. The latter probably corresponds to the time the last recorded decision on that petition was made. It is not obvious which of these two dating conventions best matches the data in the annual reports. It seems natural that *receipt date* should be the best criterion for classifying petitions filed. However, we believe *status date* is probably best to classify approvals because we understand that when a petition being processed is turned into an approval, that will be the status date reported. We think this dating convention matches

4. Among the remaining petitions, 16.2 percent were denied, 0.35 percent rejected, 0.64 percent pending, and 0.44 percent administratively closed.

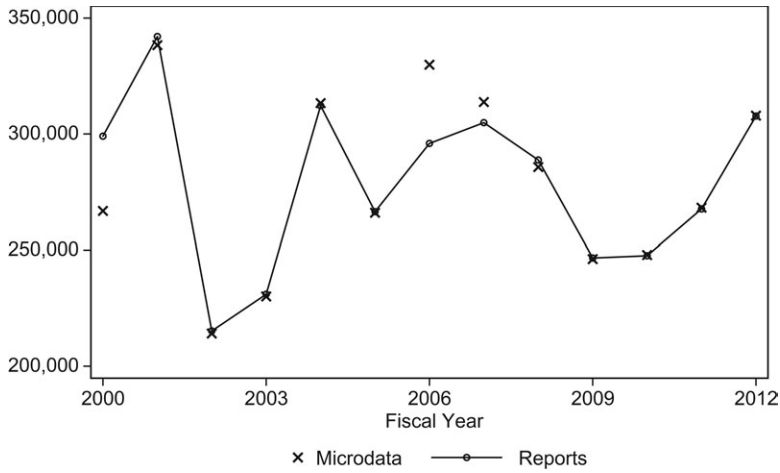

Fig. 4.1 I-129 H-1B petitions

Note: Microdata sorted by receipt year. "Reports" refers to the annual USCIS reports on Petitions and Characteristics of H-1B Workers. The R^2 of this simple linear regression is 0.88.

the spirit of the output of USCIS in terms of H-1B workers in a particular quarter, and we use it in our analysis in this section.

Counting Petitions. First, we aggregate all petitions in our microdata by fiscal (receipt) year and compare them to the annual aggregates reported in the USCIS reports. As seen in figure 4.1, in many cases the microdata exactly fits the total in the reports. However, there are significant discrepancies in years 2000, 2006, and 2007. The overall goodness of fit is 0.88, and the average ratio of petition counts in the microdata relative to the report is 1, although it varies from 0.89 to 1.11 in the years in our sample.

Approved Petitions. The data set includes petitions that were approved as well as petitions in another status (e.g., denied, rejected, or pending). So now we turn to approved petitions sorted by status date. Figure 4.2 reports the result. As before, the fit is fairly good (with an R^2 of 0.89). However, the counts for approved petitions based on our microdata are uniformly lower than the total in the reports. The ratio of approved petitions in the microdata relative to the report ranges from 0.76 to 0.94 and takes a value of 0.88 in an average year. We suspect that the larger figure in the USCIS reports may be due to the fact that when an application is amended, it might be counted as an additional processed item, even though in our microdata it might simply be recorded as a status update to an existing petition.

Approved Petitions for Initial Employment. We now turn to initial-employment petitions as defined in the previous section. As shown in figure 4.3, the match is somewhat improved relative to all approvals, but we still observe a uniformly lower count in our microdata relative to the published totals in the USCIS reports. The ratios between counts in the micro-

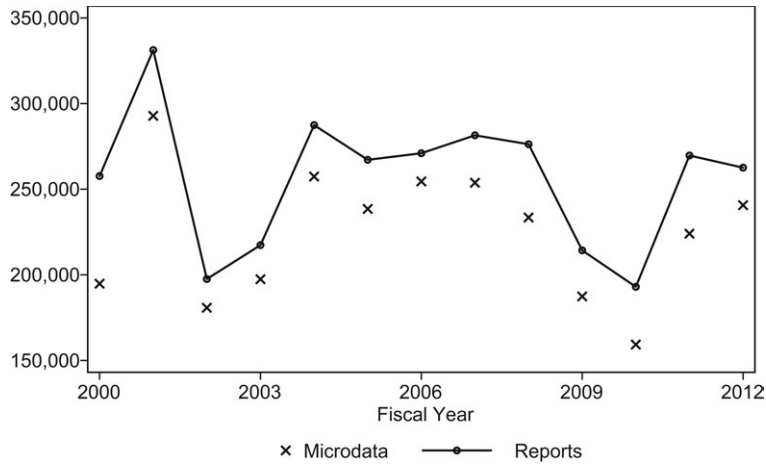

Fig. 4.2 Approved petitions for H-1B workers (I-129s)
Note: Microdata sorted by status year. Reports refer to the annual USCIS reports on Petitions and Characteristics of H-1B Workers. The R^2 of this simple linear regression is 0.89.

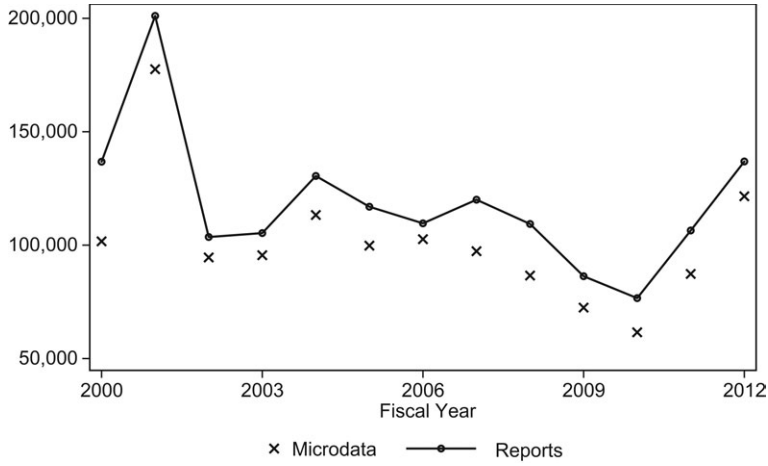

Fig. 4.3 Approved H-1B petitions for initial employment
Note: Microdata sorted by status year. Initial employment petitions (*jobstatus* = 1) excluding those referring to extensions or amendments (*request* = 3,4). "Reports" refers to the annual USCIS reports on Petitions and Characteristics of H-1B Workers. The R^2 of this simple linear regression is 0.94.

data and reported totals range between 0.74 and 0.94 and take the value 0.85 on average (the R^2 is 0.94). Obviously, the undercount of initial-employment approved petitions can be reduced by using a broader definition—that is, by defining initial-employment as any petition listing the applicant's job status as new employment regardless of whether it is simply requesting an exten-

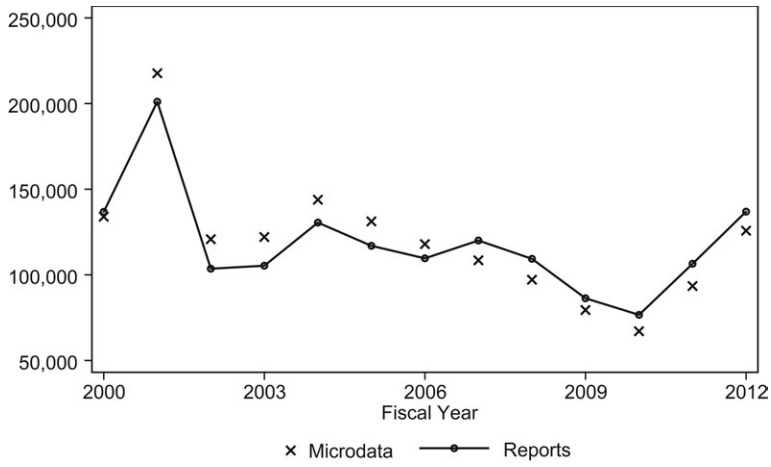

Fig. 4.4 Approved H-1B petitions for initial employment (2)
Note: Microdata sorted by status year. All initial employment petitions (*jobstatus* = 1). "Reports" refers to the annual USCIS reports on Petitions and Characteristics of H-1B Workers. The R^2 of this simple linear regression is 0.89.

sion or amendment. Clearly, in this case (figure 4.4) the number of approved initial-employment petitions increases, and we obtain a better fit of the totals in the annual reports. Nonetheless, we think that the narrower definition is more relevant for our analysis.[5]

Verdict. In summary, our comparison between our I-129 microdata and the aggregate figures in the USCIS annual reports turns out to be quite successful. Our data contain all filed petitions for most years. However, there is a small degree of discrepancy in the status of the petitions. The total approved petitions according to the annual reports is somewhat higher than what is implied by the microdata, but the two variables co-move very strongly. Agreement between the two sources of data improves when we restrict the sample to approved petitions for initial employment. Altogether, our microdata are strongly validated by the totals in the annual USCIS reports, although some discrepancies exist between the two sources.

4.3 Firm-Level Panel for Petitions

4.3.1 Aggregation

The largest data challenge we face is the aggregation of individual H-1B petitions to the company level. For each individual case, we know the

5. The average gap is now nonexistent, ranging between undercounts in some years (0.87) and overcounts (1.16) in others, with an R^2 of 0.89.

name and zip code of the company submitting the application, but we lack the exact address or, more importantly, a numerical identifier such as the employer identification number (EIN). Thus we need to rely on the company name to link individual cases within and across years. This is a challenging endeavor because a single firm will often file separate I-129 petitions under several name variants with a high prevalence of typos and misspellings. For example, there are 52 separate variants of the name "MICROSOFT" in Redmond, Washington, including "MICROSOFT CORP," "MICROSOFT COPORATION" [*sic*], "MICROSOFT CO," and just "MICROSOFT." We need to inspect the data and employ a harmonization routine to assign a common firm name to these separate entries.

We proceed in two steps. First, we conduct an extensive process of manual name harmonization in which we review the entries with company names that clearly pertain to the top H-1B-receiving firms. Specifically, we harmonize common words (e.g., "INCORPORATED," "GLOBAL," "RESEARCH") for all petitions. In addition, we manually assign a common company name to the petitions that appear to correspond to the same company in the top 3,000 firms in terms of filed petitions.[6] For instance, we aggregate records with company names "INFOSYS T," "ILNFOSYS T," and "INFORSYS TECH LIMITED" under the common name "INFOSYS TECH LIMITED." When collapsing the petitions by the harmonized name, the 3.72 million petitions in the raw data go down to 1.35 million company-year observations.

The second step conducts automatic name harmonization applied to all companies. Specifically, we parse company names to separate the company's official name from other names included in the same field (such as doing-business-as and formerly-known-as names), standardize the entity type (e.g., INC, CORP, etc.), and create numerical identifiers for groups of observations with similar names.[7] We then collapse observations using the numerical identifier, which results in 1.23 million company-year observations. When restricting to (status) fiscal years 2000–2012, the number of observations falls to 1.17 million.

An important caveat is how to deal with affiliates. We aggregate petitions under a common name in cases where company names indicate clear affiliation. For instance, we combined "IBM" with its foreign affiliate "IBM India" under the common name "INTL BUSINESS MACHINES CORP." Likewise, we also aggregated clearly recognizable affiliates within the country,

6. This ranking was built on the basis of the petitions filed in fiscal years 2008 and 2009. In these years, all new H-1B issuances were assigned through a lottery. These 3,000 firms account for more than 60 percent of all petitions filed in those years.

7. The parsing of company names is done using *Stata*'s command *STND_COMPNAME*. String-grouping is conducted using *Stata*'s STRGROUP command (Reif 2010) on the standardized name field. The command computes the Levenshtein distance between all bilateral pairs of standardized names. Pairs with a distance normalized by the number of characters corresponding to the shorter name string in the pair that is lower than 10 percent are grouped together under a common numerical identifier.

such as "AMAZON CORPORATE," "AMAZON DIGITAL," "AMAZON FULFILLMENT," "AMAZON TECH," and "AMAZON WEB," which were aggregated under the common name "AMAZON." However, we do not have systematic information on affiliates that do not share similar names.

The resulting longitudinal, firm-level data set for approved petitions contains almost 400,000 companies and 1.17 million company-year observations for the fiscal years 2000–2012. For short, we will refer to these data as the *H-1B Data Set*. For each of these companies, we have constructed the number of H-1B workers (approved I-129s) received annually in period 2000–2012, distinguishing between approvals referring to initial employment and those referring to continuing employment.[8]

4.3.2 Facts on H-1B Petitioners

Let us now examine the main facts pertaining to the *H-1B Data Set* for the period 2000–2012.

Occupation. Across all years and companies, 46 percent of all initial-employment H-1Bs were awarded to workers in *computer-related occupations*. The other most important occupations are *managers, officials, and occupations in administrative specializations* (13 percent); architects and engineers (11.3 percent); education-related occupations (9.9 percent); and occupations in *medicine and health* (6.3 percent). Together, these groups account for 87 percent of all initial-employment H-1Bs.

Metropolitan Area. It is also interesting to examine the geographical distribution of H-1B workers. This is based on the zip code listed in the I-129 form, which we matched with the corresponding metropolitan area. In many cases, this will identify the area of employment of the worker, but in others, this might simply be the headquarters of the company. Among initial-employment issuances, we observe a large concentration (21 percent) in *New York / Northeastern New Jersey*. The remaining H-1Bs are distributed much more uniformly, with 6.3 percent in *San Jose, CA*; 6.3 percent in *Washington, DC/MD/VA*; 4.7 percent in *Boston, MA/NH*; 4.5 percent in *Chicago, IL*; and 4.5 percent in *Dallas–Fort Worth, TX*. Together these six metropolitan areas account for 60 percent of all initial-employment issuances.

Rankings. Collapsing our data by company and year renders 0.82 million observations (corresponding to approximately 398,000 companies), with an annual average of 1.6 new-employment petition approvals and 1.9 continuing-employment approvals. However, there is a large degree of dispersion. Across years and companies, approved new-employment H-1Bs range between 0 and 9,483. It is also interesting to examine the rankings for a few selected years. Table 4.1 reports the top 20 receivers of new (initial-

8. These data could be used to estimate the stocks of H-1B workers at the firm level and their evolution over time. However, doing so requires making some assumptions regarding the depreciation of these stocks. For relevant information in this respect, see Depew et al. (2013) and Clemens (2013).

Table 4.1 **Ranking by approved initial employment: years 2000, 2004, 2008, and 2012**

Rank	2000 Firm	2000 Initial	2004 Firm	2004 Initial	2008 Firm	2008 Initial	2012 Firm	2012 Initial
1	TATA CONSULT	983	INFOSYS TECH	4,406	INFOSYS TECH	2,706	COGNIZANT TECH SOL	9,483
2	MICROSOFT	819	SATYAM COMPUTER	2,190	WIPRO LIMITED	2,683	TATA CONSULT	7,727
3	MOTOROLA	672	TATA CONSULT	1,879	TATA CONSULT	1,274	INFOSYS TECH	6,808
4	TEKEDGE	555	WIPRO	1,430	SATYAM COMPUTER	1,209	WIPRO LIMITED	4,002
5	INFOSYS TECH	652	COGNIZANT TECH SOL	1,196	MICROSOFT CORP	1,063	ACCENTURE LLP	3,548
6	INTEL	519	PATNI COMPUTER INC	941	ACCENTURE LLP	712	HCL AMERICA	2,133
7	ACE TECH	505	IBM GLOBAL INDIA	766	CISCO SYSTEMS INC	471	LARSEN & TOUBRO INFOTECH	1,703
8	MASTECH	493	MICROSOFT CORP	646	COGNIZANT TECH SOL	417	IBM INDIA PRIVATE	1,427
9	CISCO SYS	477	TATA INFOTECH	607	IBM GLOBAL INDIA	401	SATYAM COMPUTER	1,293
10	ALPHASOFT SVC	425	NYC PUBLIC SCHOOLS	540	LARSEN & TOUBRO INFOTECH	398	MICROSOFT	1,231
11	WIPRO	403	MPHASIS CORPORATION	516	INTEL CORP	377	PATNI AMERICAS	1,227
12	HTC GLOBAL SVC	394	LARSEN & TOUBRO INFOTECH	461	QUALCOMM	238	SYNTEL CONSULT	1,104
13	TECHSPACE SOL	368	SYNTEL	418	MPHASIS CORP	229	DELOITTE CONSULT	909
14	DATA CONVERSION	354	DELOITTE & TOUCHE LLP	346	BALTIMORE PUBLIC SCHOOLS	229	TECH MAHINDRA AMERICAS	826
15	BIRLASOFT	346	HEXAWARE TECH INC	298	PRINCE GEORGE PUBLIC SCHOOLS	213	MPHASIS	704
16	PEOPLE COM CON	342	DELOITTE CONSULTING	270	UST GLOBAL INC	199	AMAZON CORP	611
17	COGNIZANT TECH	327	POLARIS SOFT LAB INDIA	267	ERNST & YOUNG LLP	198	INTEL	552
18	SYNTEL	298	PWC	253	VERINON TECH SOL	176	GOOGLE	512
19	LUCENT TECH	291	UNIVERSITY OF PENN	238	GOOGLE	174	PRICEWATERHOUSECOOPERS	487
20	SATYAM COMPUTER	271	CAMBRIDGE RESOURCE GROUP	229	TERRA INFOTECH	166	UST GLOBAL INC	445
	TOTAL	112,071		109,662		86,470		116,099
	Share top 10	5.4%		13.3%		13.1%		33.9%
	Share top 20	8.5%		16.3%		15.7%		40.3%
	Share IT top 20	4.5%		15.0%		12.3%		37.7%

Note: The names of some of the companies have been shortened slightly to accommodate the formatting of the table. The row TOTAL refers to the total number of initial-employment visas approved in the corresponding year taking into account all petitions (not just those submitted by the firms in the ranking presented here. The bottom row reports the number of initial-employment visas that went to business and technology consulting firms that are part of the ranking presented in the table, as a share of the TOTAL initial-employment visas.

employment) H-1B issuances in the years 2000, 2004, 2008, and 2012. The top 3 companies by approved (initial-employment) visas in year 2000 were TATA CONSULTANCY SERVICES, MICROSOFT, and MOTOROLA. From 2004 onward, the top 3 companies have been business and information-technology consulting companies based in India, alternating between INFOSYS TECH, SATYAM COMPUTER SERVICES, WIPRO LIMITED, and TATA CONSULTANCY SERVICES. In addition, the number of H-1B visas obtained by these firms has grown enormously as a result of growing demand for their services. More generally, with the exception of MICROSOFT, AMAZON, INTEL, and GOOGLE, all other companies in the 2012 top-20 ranking by approved petitions for initial-employment issuances were business and technology consulting firms.

Increased Concentration. Between 2000 and 2012, the data show a sharp increase in the concentration of new visas in the hands of a small number of companies. In 2000, the top 20 receivers obtained 8 percent of the 112,071 issuances for initial employment granted in that year. In 2004, the degree of concentration increased further, with the top 20 firms receiving 16 percent of the 109,662 H-1Bs for new employment granted in that year. The share of these workers being granted to the top 20 companies remained at 16 percent in 2008 despite the lower total of 86,470 H-1Bs. However, there was another sharp increase in concentration in 2012, with the top-20 share increasing to 40 percent for a total of 116,099 H-1Bs granted in that year. In sum, the data reveal a fourfold increase in the top-20 share for new-employment H-1Bs over the period 2000–2012.

The rise in concentration has been fundamentally driven by business and IT consulting companies. As can be seen at the bottom of table 4.1, the IT consulting companies among the top 20 receivers accounted for 4.5 percent of new-employment visas in year 2000, slightly over half of the share among all the top 20 receivers. However, in year 2012, IT companies among the top 20 receiving companies accounted for 37.7 percent of all new-employment H-1B visas, or 94 percent of the visas awarded to the top 20 receivers.

Educational and Research Institutions. We also note that public school districts (e.g., New York City Public Schools) and universities (e.g., University of Pennsylvania) enter the top-20 ranking in some years. In table 4.2, we present the top-10 ranking of petitioners for initial-employment H-1Bs in years 2004, 2008, and 2012, distinguishing between for-profit and nonprofit organizations. This distinction is important because the latter are generally exempt from the annual cap. In the three selected years, the top petitioner of initial-employment H-1B issuances was the New York City Public School District. In addition, leading research universities are also part of the top 10, such as Yale, Stanford, University of Michigan, and University of Pennsylvania.

In recent work, we show that the above facts are consistent with evidence based on a triple difference estimation strategy. In particular, Mayda et al.

Table 4.2 **Ranking by approved new-employment H-1B petitions by exemption status**

Year 2004	Year 2008	Year 2012
Cap-bound		
INFOSYS TECH LIMITED	INFOSYS TECH LIMITED	COGNIZANT TECH SOLUTIONS US CORP
SATYAM COMPUTER SERVICES LIMITED	WIPRO LIMITED	INFOSYS TECH LIMITED
MICROSOFT CORP	MICROSOFT CORP	WIPRO LIMITED
TATA CONSULTANCY SERVICES LIMITED	SATYAM COMPUTER SERVICES LIMITED	TATA CONSULTANCY SERVICES LIMITED
WIPRO LIMITED	COGNIZANT TECH SOLUTIONS US CORP	MICROSOFT CORP
COGNIZANT TECH SOLUTIONS US CORP	TATA CONSULTANCY SERVICES LIMITED	ACCENTURE LLP
PATNI COMPUTER SYSTEMS INC	CISCO SYSTEMS INC	LARSEN & TOUBRO INFOTECH LIMITED
CISCO SYSTEMS INC	IBM CORP	HCL AMERICA INC (HCL TECH)
IBM GLOBAL SVCS IGS INDIA PVT	INTEL CORP	IBM INDIA PRIVATE LIMITED
INTEL CORP	ORACLE USA INC	INTEL CORP
Year 2004	Year 2008	Year 2012
Cap-exempt		
NEW YORK CITY PUBLIC SCHOOLS	NEW YORK CITY PUBLIC SCHOOLS	NEW YORK CITY PUBLIC SCHOOLS
YALE UNIVERSITY	PRINCE GEORGE'S COUNTY PUBLIC SCHOOLS	STANFORD UNIVERSITY
UNIVERSITY OF MICHIGAN	YALE UNIVERSITY	UNIVERSITY OF MICHIGAN
COLUMBIA UNIVERSITY	UNIVERSITY OF MICHIGAN	YALE UNIVERSITY
UNIVERSITY OF FLORIDA	UNIVERSITY OF PENNSYLVANIA	CLEVELAND CLINIC FOUNDATION
STANFORD UNIVERSITY	COLUMBIA UNIVERSITY	JOHNS HOPKINS UNIVERSITY
BAYLOR COLLEGE OF MEDICINE	HOUSTON INDEPENDENT SCHOOL DISTRICT	UNIVERSITY OF PITTSBURGH
DUKE UNIVERSITY MED CENTER & HOSPITAL	UNIVERSITY OF PITTSBURGH	COLUMBIA UNIVERSITY
UNIVERSITY OF PA	STANFORD UNIVERSITY	DUKE UNIVERSITY UNIVERSITY MED CENTER & AFFIL
HOUSTON INDEPENDENT SCHOOL DISTRICT	DUKE UNIVERSITY UNIVERSITY MED CENTER & AFFIL	EMORY UNIVERSITY

Note: Ranking on the basis of new-employment approved I-129 petitions in our *H-1B Data Set*.

(2018) empirically analyze the intended and unintended effects of the 2004 reduction in the H-1B quota. The policy change created a sudden discontinuity in the maximum supply of H-1B visas for the "treated" group of new H-1B workers of for-profit firms relative to the "control" group of experienced H-1B workers of for-profit firms and of (new and experienced) H-1B workers of nonprofit firms. We find that the cap restrictions significantly reduced the aggregate employment of new H-1B workers in for-profit firms relative to what would have occurred in an unconstrained environment. In addition, our results show that the quota reduction implied no change in employment of H-1B workers, respectively, in computer-related occupations, from India, and in firms that employ more than ten H-1B workers. As a consequence, the quota reduction redistributed H-1Bs toward computer-related occupations, Indian-born workers, and firms using the H-1B program extensively. This, in turn, produced a much higher concentration of the H-1B visas in the hands of a few employers, as shown in Lorenz curves of the inequality of H-1B issuances across firms (see figure 4.5: H-1B Concentration in Firms in Mayda et al. 2018).

4.4 H-1Bs among Publicly Traded Firms

Unfortunately, our *H-1B Data Set* does not contain any firm-level information beyond its name and geographic location. In order to learn more about the trends regarding the demand for H-1B workers as a function of firm-level characteristics, we merge our data set with Compustat. Once again, this needs to be done on the basis of company name.

4.4.1 Merging with Compustat

After some basic cleaning, our Compustat data contains 7,067 companies.[9] As noted earlier, the *H-1B Data Set* contains nearly 400,000 companies. To match the companies in this data set to the companies in Compustat, we make use of probabilistic record-linking techniques.[10] In essence, we examine all pairs (*n,m*), where *n* refers to the name in Compustat and *m* to the name in the *H-1B Data Set*. As before, for each pair of names, we compute a measure of similarity between the two character strings.

The code produces more than 11,000 potential matches, with associated scores ranging between 0.60 and 1. There are 3,070 perfect matches with a (perfect) score of 1. Clerical review of the potential fuzzy matches is time

9. We restricted the Compustat sample to companies with nonmissing, nonzero employment in 2012, which results in 7,067 companies. Interestingly, only 5,294 of these companies have an employer identification number (EIN), and in fact, several of the top recipients of H-1B workers, such as INFOSYS, SATYAM, WIPRO, or ERICSSON, lack an EIN. Hence some degree of record-linking error based on company names is unavoidable.

10. The specific record-linking protocol we use is Stata's *reclink2* command. This code is an extension of Blasnik's (2010) procedure carried out by Wasi and Flaaen (2014).

Table 4.3 **Record-linking *H-1B Data Set* and *Compustat***

	1	2	3	4	5	6
Threshold RLSC	1	0.99	0.98	0.97	0.96	0.95
perfect matches	3,070	3,070	3,070	3,070	3,070	3,070
potential fuzzy matches	.	900	991	1,101	1,237	808
accepted fuzzy matches	.	454	327	207	223	68
success rate	.	0.50	0.33	0.19	0.18	0.08
total matches	3,070	3,524	3,851	4,058	4,281	4,349
collapsed by firm	2,169	2,489	2,687	2,823	2,957	3,002
share of Compustat comp. with approved I-129s	0.31	0.35	0.38	0.40	0.42	0.42

Note: The RLSC (record-linking score) is the key output of the *reclink2* probabilistic record-linking routine. It is a measure of similarity between the two company name strings. The similarity score is based on the number of characters that need to be changed in one of the strings in order to perfectly match the other string. The shares of the last row are computed on the basis of the 7,067 Compustat companies (with nonmissing, nonzero employment in 2012). Column 1 considers only perfect matches. Columns 2–6 also include fuzzy matches, with a gradually decreasing threshold for the record-linking score in order to be considered.

consuming—it takes about one hour to review 500 candidate pairs. As a result, we conduct clerical review in stages, gradually lowering the similarity score threshold.[11] As reported in table 4.3, there are 3,070 pairs with a perfect match by company name (column 1). The next column also includes the (roughly 900) potential matches with a similarity score above 0.99. After manually reviewing each of them, we conclude that 454 of those are correct, amounting to a 50 percent success rate. We then proceed to review the candidate pairs with scores above 0.98, which results in a 33 percent success rate. Columns 4 to 6 gradually lower the similarity score threshold to 0.97, 0.96, and 0.95. As expected, the success rates decline to 19 percent, 18 percent, and 8 percent, respectively. At this point we deem the success rate to be too low to merit further clerical review. We have matched 4,349 pairs of company names. However, some of these pairs refer to the same firm. When collapsing by firm, we end up with 3,002 Compustat firms having approved I-129s, which amounts to 42 percent of all Compustat firms (with nonzero, nonmissing employment).

It is also worth noting that Compustat companies are only a small fraction

11. Some pairs have very similar names, which is why they are over the similarity threshold, but it is unclear whether they refer to the same company. For example, (ANDERSON,ANDERSONS) could very well refer to two different companies, so we verify that they exist. Typically, in ambiguous cases where both companies exist, we do not accept the match. We only assume there was a typo when the name for the I-129 data entry corresponds to a company that does not seem to exist according to Google searches. We are fairly confident of the quality of our matches. Keep in mind that some pairs will have been rejected despite being true matches. This type of measurement error is, by construction, random and should not bias our estimates.

Table 4.4 **Examples of the evolution of approved initial-employment petitions**

Year	gvkey	Approved petitions: initial employment	Approved petitions: continuing employment
GOOGLE			
2000	160329	6	2
2001	160329	16	9
2002	160329	11	9
2003	160329	36	31
2004	160329	71	52
2005	160329	184	120
2006	160329	148	149
2007	160329	178	217
2008	160329	174	180
2009	160329	252	215
2010	160329	298	388
2011	160329	573	495
2012	160329	512	579
COGNIZANT			
2000	111864	327	131
2001	111864	451	222
2002	111864	185	197
2003	111864	599	273
2004	111864	1,197	685
2005	111864	817	482
2006	111864	586	1,457
2007	111864	663	1,347
2008	111864	417	1,329
2009	111864	1,308	1,319
2010	111864	4,050	2,510
2011	111864	4,963	3,501
2012	111864	9,484	6,152

Note: Based on approved I-129 forms for initial-employment H-1B issuances on the basis of our USCIS microdata merged with Compustat. To save on space, we have shortened the company names.

of all companies based in the United States. Summing over all years in our data, Compustat firms account for roughly 412,000 approved H-1B petitions for H-1B, with 40 percent of these referring to initial-employment issuances. This figure accounts for only 13 percent of the three million approved H-1B petitions over the period 2000–2012.

Next we report two specific examples of companies that have substantially increased their use of H-1B workers over our period of analysis. The top panel in table 4.4 reports the data for GOOGLE. In year 2000, GOOGLE obtained merely six and two initial and continuing employment workers, respectively. Over the next 12 years, GOOGLE has received an increasing number of initial-employment issuances, peaking at 573 in 2011. The bottom panel reports the data for COGNIZANT. This company obtained a few

Table 4.5 Approved H-1B petitions: sum of initial and continuing employment

Approved H-1B	2000 percentage	2012 percentage
None	77	80
1 to 10	18	15
11+	5	5
Total firms	3,419	7,067

Note: Distribution of Compustat companies in year 2000 (or 2012) with nonmissing employment over the number of approved H-1B petitions (pooling initial and continuing employment). The lower number of firms in 2000 is due to the fact that our Compustat sample conditions on nonmissing, nonzero employment in year 2012.

hundred initial-employment issuances every year between 2000 and 2008. From 2009 onward, the growth in the number of this type of H-1B has been exponential. In 2012, COGNIZANT received 9,484 initial-employment H-1Bs compared to only 327 in year 2000.

4.4.2 Facts on Compustat H-1B Petitioners

As noted earlier, our matched H-1B-Compustat data set is a longitudinal data set containing 7,067 companies and 12 years.[12] We were able to match about 42 percent of the firms in Compustat through our string-matching algorithm, and we imputed zero issuances to the unmatched firms.

4.4.2.1 Characteristics of H-1B-Using Companies

The first exercise we carry out is a comparison between the matched (i.e., H-1B users) and unmatched Compustat companies. We focus on employment, revenue, and market value, both in levels and in growth rates.

Our starting point is to build the distribution of Compustat companies by usage of the H-1B program. Specifically, we consider the companies with nonmissing, nonzero employment in 2000 (as well as in 2012) and classify them in three groups: companies with no approved petitions in 2000, companies with 1 to 10 approved petitions (for initial or continuing employment), and companies with 11 or more approved petitions in 2000. The resulting distribution is summarized in table 4.5: 77 percent, 18 percent, and 5 percent, respectively, among the 3,419 companies satisfy the restrictions. The table also presents the H-1B usage distribution for the 7,067 firms with nonzero, nonmissing employment in 2012, with 80 percent of firms with no approved H-1B petitions in year 2012, 15 percent with 1 to 10 approved petitions, and 5 percent with 11 or more approved petitions in that year.[13]

12. The time dimension is restricted by the availability of data on H-1B petitions, which ranges from year 2000 to 2012. Among Compustat companies we have restricted to those that have nonmissing, nonzero employment in year 2012.

13. As noted earlier, there may be some unmatched firms that did receive H-1B workers. However, the size of this group is likely to be very small based on the statistics reported in table 4.3.

Table 4.6 **Characteristics of H-1B usage**

Year	2000	2000	2000	2012	2012	2012
H-1B	None	1 to 10	11+	None	1 to 10	11+
Employment (M)	10	16	44	8	13	35
Revenue ($MM)	2,462	3,744	12,593	3,103	4,296	17,330
Market value ($MM)	1,765	4,830	29,783	1,803	3,528	21,851
Growth employment	11.2%	12.4%	15.2%	6.0%	6.4%	8.8%
Growth revenue	32.4%	61.1%	85.3%	20.5%	19.3%	20.7%
Growth market value	.	.	.	30.1%	62.3%	40.1%
Mode SIC2d	60, 28	73, 36	73, 36	60, 73	73, 36	73, 36

Note: Employment counts are in thousands of employees. Revenue and market value are in millions of dollars (at current prices). The last row reports the top two mode industries (to-digit Standard Industrial Classification [SIC] code) in each column. The relevant SIC codes are as follows: chemicals and allied products (28); electronic and other electrical equipment and components, except computer equipment (36); depository institution (finance; 60); business services (73); engineering, accounting, research, management, and related services (87); and industrial and commercial machinery and computer equipment (35). In the bottom three rows, for year 2000, the growth rate is computed as the annualized 1997–2000 growth rate. For year 2012, the growth rate is computed as the annualized 2009–12 growth rate. To compute these growth rates, we restrict to companies with initial year (1997 or 2000) values of at least 1,000 employees and $1MM revenue and market values.

Size and Market Value. Next we compare the three groups of companies on the basis of H-1B usage. As reported in table 4.6 (columns 4–6), in year 2012, the average employment for Compustat companies that did not receive any (initial or continuing employment) H-1Bs in year 2012 was 8,000 workers. In comparison, companies that had 1 to 10 or 11 or more approved petitions had average employment of 13,000 and 35,000, respectively. Thus firms employing H-1B workers are much larger than nonusers. The same size gradient is also present in terms of revenue and market value. In year 2012, the average revenue among non-H-1B users in Compustat was $3.1 billion compared to $4.3 billion and $17.3 billion among moderate and heavy users of the program. These relationships are also confirmed when we focus on year 2000 (columns 1–3).

Firm Growth. The bottom part of the table examines firm-level growth rates by H-1B usage, which suggests there exists a positive relationship as well between the number of approved H-1B petitions and firm growth (over the previous three years). More specifically, the 2009–12 annualized growth rate in terms of employment was 6.0 percent among firms that did not receive any H-1B workers in year 2012 (measured by approved petitions for either initial or continuing employment). In comparison, moderate and heavy users of the program exhibited average employment growth rates of 6.4 percent and 8.8 percent, respectively. Revenue growth in this period was practically the same for the three groups of firms at around 20 percent per year. In terms of growth in market value, once again we see substantially

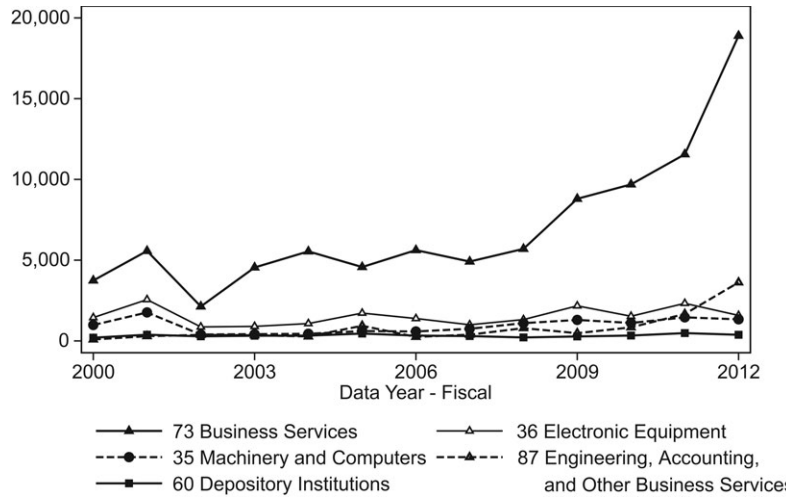

Fig. 4.5 Approved initial-employment H-1B petitions by industry

Note: Approved initial-employment H-1B petitions by two-digit SIC industry code. We plot only the data for the top five receiving industries in 2012.

higher growth rates among users of the H-1B program (40–60 percent) relative to nonusers (30 percent). The 1997–2000 growth rates also confirm these patterns, with clearer evidence of a monotonic relationship between H-1B usage and firm growth.

Clearly, these are purely descriptive facts. To a large extent, the differences in level and growth as a function of H-1B usage reflect differences in terms of industry composition. The last row in table 4.6 reports the mode two-digit Standard Industrial Classification (SIC) code by H-1B usage. The mode industries in the sample that did not receive approved petitions in 2012 were depository institutions (finance; 60) and business services (73). Among H-1B users, the mode industries were business services (73) and electronic and other electrical equipment and components, except computer equipment (36).

4.4.2.2 Industry Trends

Counts of Approved Petitions. In order to better understand industry trends in H-1B usage, we collapse our H-1B-Compustat data set by two-digit SIC industries. Figure 4.5 plots the counts of approved initial-employment H-1B petitions for the top five receiving industries. The top receiving industry is business services (73), followed by electronic equipment (36), machinery and computers (35), engineering, accounting and other business services (87), and depository institutions (60). Business services is by far the industry receiving the largest number of workers. Between years 2000 and 2008, Compustat companies in this industry received about 5,000 initial-employment H-1Bs annually. However, there has been an explosion in this figure since

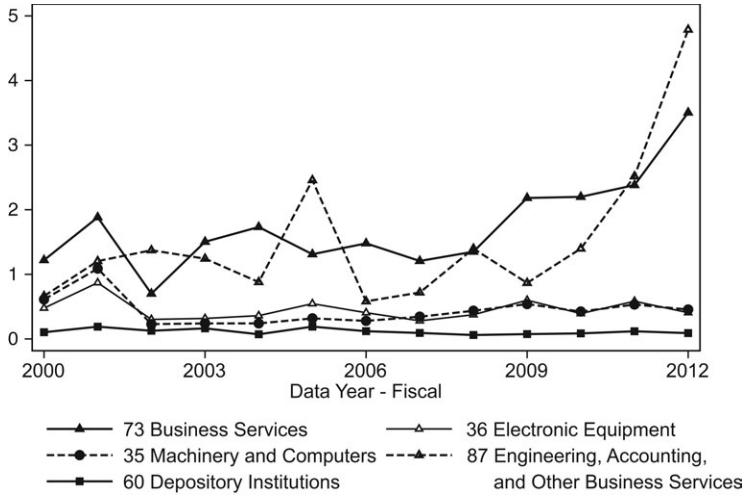

Fig. 4.6 H-1B Intensity at the industry level: approved initial-employment H-1B petitions per 1,000 employees by industry

Note: Approved initial-employment H-1B petitions by two-digit SIC industry code per 1,000 employees. We plot only the data for the top five receiving industries in 2012.

2009. In 2012, these companies hired close to 20,000 initial-employment H-1B workers.

Intensity of Use of H-1B Visas. Naturally, this increase may simply reflect a rise in the size of the business services industry, keeping the intensity of H-1B use constant. To examine this hypothesis, we compute the industry-level intensity, defined as approved initial-employment issuances per 1,000 employees, and plot it in figure 4.6. The figure suggests that the bulk of the increase in H-1B usage in the business services industry is due to an increase in intensity. The intensity of initial-employment H-1Bs in the business services industry has remained practically unchanged throughout the 2000–2008 period (at around 1.5 initial-employment issuances per 1,000 employees). However, it grew by 133 percent between 2008 and 2012. Interestingly, the engineering, accounting, and other business services (87) industry exhibits very similar behavior. In fact, in 2012, the H-1B intensity in this industry is 5 initial-employment H-1Bs per 1,000 employees, compared to a 3.5 intensity for business services (73).

4.5 Conclusions

As is often the case in merging large data sets based on names of firms with automated or semiautomated matching techniques, the quality of the matches improves at each iteration, and a perfect match is often infeasible. This is also the case here. While we believe that the general facts presented here will persist, we also note that our data set will continue to evolve as

we continue improving the quality of our matching algorithm. False positives (matched firms that should not have been matched) and false negatives (unmatched firms that should have been matched) will continue to occur. Naturally, a nearly perfect match could be attained if USCIS agreed to release the employer identification number (EIN) associated with each petitioning firm, which so far has not been the case.

Possibly, the single most important fact regarding the aggregate economic effects of the current H-1B program is the large increase in the concentration of H-1Bs in the hands of a small number of global technology consulting companies. With little doubt, the large expansion of these firms derives from a pronounced trend toward outsourcing of information technology services. This trend may be fundamentally driven by technological developments in information and communication systems that have triggered this change in the boundaries of the firm. However, it is also possible that the increasing difficulty in obtaining and managing H-1Bs due to the increasing excess demand over the last few years has accelerated the tendency to outsource these tasks. At any rate, it is important to keep in mind that from its inception, the H-1B visa program has been intended as a vehicle for *trade in services*.[14]

Some recent papers (Peri et al. 2015) have argued that the H-1B program may have increased the productivity and wages of highly skilled *native* workers due to spillovers and increasing returns to innovation. However, the recent trend toward an increasing concentration of H-1B workers in the hands of companies engaged in outsourcing of information-technology services may reduce the scope for these spillovers even though it is likely to increase the profitability (and perhaps the productivity) of the firms contracting out IT services. Characterizing precisely the firm-level dynamics of H-1B users, which will be made possible by this data set and further iterations of it, is crucial to predict the potential impact of the H-1B visa program into the future.

References

Amuedo-Dorantes, Catalina, and Delia Furtado. 2016. "Settling for Academia? H-1B Visas and the Career Choices of International Students in the United States." IZA Discussion Papers 10166, August. Institute for the Study of Labor (IZA).

14. As part of the Uruguay Round (1986–1994) multilateral trade agreements, the US agreed to set up the H-1B visa program and committed to offering at least 65,000 visas. Similarly, the US also set aside specific numbers of visas during the negotiation of the North American Free Trade Agreement (NAFTA) and the US Free Trade Agreements with Chile (2,400 visas) and Singapore (1,600 visas). These agreements were also incorporated into the H-1B visa program. We thank Jennifer Hunt for pointing out the connection between the H-1B visa program and multilateral trade agreements.

Bessen, James, and Robert M. Hunt. 2007. "An Empirical Look at Software Patents." *Journal of Economics and Management Strategy* 16 (1): 157–89.
Blasnik, Michael. 2010. "RECLINK: Stata Module to Probabilistically Match Records." Working Paper.
Clemens, Michael A. 2013. "Why Do Programmers Earn More in Houston Than Hyderabad? Evidence from Randomized Processing of US Visas." *American Economic Review* 103 (3): 198–202.
Depew, Briggs, Peter Norlander, and Todd A. Sorensen. 2013. "Flight of the H-1B: Inter-firm Mobility and Return Migration Patterns for Skilled Guest Workers." IZA Discussion Papers 7456, June. Institute for the Study of Labor (IZA).
Doran, Kirk, Alexander Gelber, and Adam Isen. 2014. "The Effects of High-Skilled Immigration Policy on Firms: Evidence from H-1B Visa Lotteries." NBER Working Paper no. 20668. Cambridge, MA: National Bureau of Economic Research.
Ghosh, Anirban, Anna Maria Mayda, and Francesc Ortega. 2014. "The Impact of Skilled Foreign Workers on Firms: An Investigation of Publicly Traded U.S. Firms." IZA Discussion Papers 8684, November. Institute for the Study of Labor (IZA).
Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper no. 8498. Cambridge, MA: National Bureau of Economic Research
Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2 (2): 31–56.
Kato, Takao, and Chad Sparber. 2013. "Quotas and Quality: The Effect of H-1B Visa Restrictions on the Pool of Prospective Undergraduate Students from Abroad." *Review of Economics and Statistics* 95 (1): 109–26.
Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln. 2015. "Firms and the Economics of Skilled Immigration." In *Innovation Policy and the Economy*, vol. 15, edited by William R. Kerr, Josh Lerner, and Scott Stern, 115–52. Chicago: University of Chicago Press.
Kerr, William R., and William F. Lincoln. 2010. "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention." *Journal of Labor Economics* 28 (3): 473–508.
Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber. 2018. "The Effect of the H-1B Quota on the Employment and Selection of Foreign-Born Labor." *European Economic Review* 108:105–12.
Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. "STEM Workers, H-1B Visas, and Productivity in US Cities." *Journal of Labor Economics* 33 (S1): 225–55.
Reif, Julian. 2010. "STRGROUP: Stata Module to Match Strings Based on Their Levenshtein Edit Distance." Statistical Software Components S457151, Boston College Department of Economics, Revised August 14, 2010.
Shih, Kevin. 2016. "Labor Market Openness, H-1B Visa Policy, and the Scale of International Student Enrollment in the United States." *Economic Inquiry* 54 (1): 121–38.
Wasi, N., and A. Flaaen. 2015. "Record Linkage Using Stata: Preprocessing, Linking, and Reviewing Utilities." *Stata Journal* 15 (3): 672–97.