NBER WORKING PAPER SERIES

THE EFFECT OF HIGH-SKILLED IMMIGRATION ON PATENTING AND EMPLOYMENT: EVIDENCE FROM H-1B VISA LOTTERIES

Kirk Doran
Alexander Gelber
Adam Isen

Working Paper 20668
http://www.nber.org/papers/w20668

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2014

We thank U.S. Customs and Immigration Services for help with the H-1B lottery data. We thank Sunil Vidhani for outstanding research assistance. We thank Notre Dame and the Wharton School of the University of Pennsylvania for research support. We are grateful to Lee Fleming for sharing the patent data with us. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the U.S. Treasury Department, any other person associated with the U.S. Treasury Department, or the National Bureau of Economic Research. All errors are our own.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w20668.ack

The Effect of High-Skilled Immigration on Patenting and Employment: Evidence from H-1B Visa Lotteries
Kirk Doran, Alexander Gelber, and Adam Isen
NBER Working Paper No. 20668
November 2014
JEL No. J18,J21,J23,J24,J44,J48,J61,O3,O32,O34,O38

ABSTRACT

We study the effect of winning an additional H-1B visa on a firm's patenting and employment outcomes. We compare firms randomly allocated H-1Bs in the Fiscal Year 2006 and 2007 H-1B visa lotteries to other firms randomly not allocated H-1Bs in these lotteries. We use Department of Homeland Security administrative data on the winners and losers in these lotteries matched to administrative data on the universe of approved U.S. patents, and matched to IRS administrative data on the universe of U.S. employment. Winning an H-1B visa has an insignificant average effect on patenting, with confidence intervals that rule out moderate-sized effects and that are even more precise in many cases. Employment data generally show that on average H-1B workers at least partially replace other workers in the same firm, with estimates typically indicating substantial crowdout of other workers.

Kirk Doran
438 Flanner Hall
University of Notre Dame
Notre Dame, IN 46556
kdoran@nd.edu

Alexander Gelber
Goldman School of Public Policy
University of California at Berkeley
2607 Hearst Ave
Berkeley, CA 94720
and NBER
agelber@berkeley.edu

Adam Isen
Office of Tax Analysis
U.S. Department of the Treasury
1500 Pennsylvania Ave., NW
Washington, DC 20220
adam.isen@gmail.com

1. Introduction

What are the benefits and costs of high-skilled immigration for the economy receiving the immigrants? This question has inspired debate among economists and policymakers for decades. The debate has reached a fever pitch in the last several years, with prominent voices from government, the business community, the labor community, academia, and the media discussing major changes to U.S. immigration law. While extensive literature has examined how high-skilled immigration affects wages, employment, and innovation, this literature has not reached a consensus. One hurdle is the wide variety of sources of variation and research designs that the literature has relied on for identification, including visa caps, supply-push instruments, and other natural experiments (e.g. Card 1990; Altonji and Card 1991; Borjas, Freeman, and Katz 1997; Card 2001; Friedberg 2001; Borjas 2003; Edin, Fredriksson, and Åslund 2003).

Our paper addresses identification by using randomization to estimate the causal impact of high-skilled immigration on the receiving firm. Specifically, we exploit lotteries for visas given through the largest high-skilled immigration program: H-1B visas for temporary immigration. We use administrative microdata on these lotteries from the U.S. Citizenship and Immigration Services (USCIS), matched to U.S. Patent and Trademark Office (USPTO) data on the universe of U.S. patents, and matched to Internal Revenue Service (IRS) microdata on the universe of employment at U.S. firms. We use these data to examine the effect of winning an additional H-1B on firms' patenting and employment outcomes.

The previous literature has found that H-1Bs have substantial positive effects on patenting and employment (Kerr and Lincoln 2010, Hunt 2011, Peri, Shih, and Sparber 2014, and Pekkala Kerr, Kerr and Lincoln forthcoming). Immigrants with H-1B visas may have exceptional skills that cannot easily be obtained any other way. Under this scenario, a firm that gains an H-1B worker could be more likely to develop new techniques or new knowledge, some of which it may wish to patent. Furthermore, such new techniques, and/or complementarity between H-1B workers and other workers, could cause the firm to increase its employment of other workers as well. This is the scenario exemplified by former Microsoft Chairman Bill Gates' statement in congressional testimony that Microsoft hires four additional employees to support each new high skill immigrant worker hired on the visa (Gates 2008). More generally, receiving an extra H-1B worker may lead to an increase in employment at a firm, unless H-1B workers fully replace other workers.

On the other hand, economic theory predicts that firms will apply to hire an H-1B worker if doing so increases the firm's profit—which is, of course, distinct from increasing the firm's rate of patenting. It could be that both with and without the extra H-1B worker, the firm does not patent. Moreover, if H-1B workers are extremely substitutable with other workers, then we may see small or negligible changes in employment when a firm "wins" an H-1B worker. An H-1B worker could even replace a native worker who would have otherwise patented more (or less, or equally)—but the firm still chooses to hire the H-1B worker because the wage paid to the H-1B worker is lower relative to the worker's marginal product than the wage of the native relative to the native's marginal product. Although prevailing wage regulations are intended to require firms to pay H-1B workers the same amount as natives with similar skills, these regulations may not achieve their intended effect. In such scenarios, hiring an additional H-1B visa worker would not necessarily increase the rate of knowledge generation and innovation in the firm. This is suggested in the case studies of H1-B-induced job displacement in Matloff (2003), who argues that H-1B visas tend to replace older workers with higher salaries.

To investigate these questions, we use the Fiscal Year (FY) 2006 and FY2007 H-1B visa lotteries to evaluate the impacts of an additional H-1B visa immigrant at the firm level. In these years, when firms submitted H-1B visa applications precisely on the date when USCIS reached the maximum number of H-1B visa applications allowed for a given year and visa type, the applications submitted on these dates were subject to a lottery. Some visa applications were randomly chosen by USCIS to win the lottery, while the remaining visa applications were randomly chosen to lose the lottery. Across both years and across lotteries for two visa types (for those with and without advanced degrees), 3,050 firms applied for 7,243 visas, of which 4,180 won the lottery. Our results speak directly to an important issue: the effects of increases in the cap on the number of visas that applies to firm-sponsored visas (as opposed to H-1B visas not subject to the cap, such as those for educational institutions).

Across our specifications, which examine the impacts of an additionally approved H1-B visa on the firm's approved patents over the seven years following the start of the visa, the estimated effects cluster around zero and are never significantly positive. Our confidence intervals allow us to rule out moderate-sized effects, and in many cases they are even more precise. The results are particularly precise when we focus on small firms, where the impact of one additional employee is likely to be most clearly distinguishable from the baseline in a statistical sense; one additional programmer, for example, may have a large impact relative to the baseline in a firm with ten programmers, but would represent a "drop in the bucket" at a firm with one thousand programmers. The results suggest that plausible changes in the H-1B visa cap would have at most a small effect on firm patenting relative to the baseline.

On employment, our paper is the first to our knowledge to document evidence that H-1Bs displace other workers.[2] In most specifications, the estimates indicate substantial and statistically significant crowdout of other workers within one year of the start of the visa. Thus, over this time frame our findings generally rule out the scenario in which one additional H-1B visa immigrant leads to an increase in total firm employment of greater than one, and they generally rule out the claim that an additional approved H-1B visa has no negative effect on the employment of other workers at the same firm.

Our paper is closely related to other literature on the innovation or labor market impacts of the H-1B program, including Kerr and Lincoln (2010), Hunt (2011), Peri, Shih, and Sparber (2013), Peri, Shih, and Sparber (2014), and Pekkala Kerr, Kerr and Lincoln (forthcoming). In contrast to our results, these papers have found that the H-1Bs have large positive impacts on innovation and productivity and have found no clear evidence of displacement of other employment. In preliminary work, Peri, Shih, and Sparber (2014) examine the winners of H-1B visa lotteries, but because they do not have access to the list of random lottery losers their paper does not leverage randomized variation.[3] Our paper's results are not fully comparable to much previous literature on the

[2] Kerr and Lincoln (2010) find no evidence that H-1Bs displace other workers. Pekkala Kerr, Kerr, and Lincoln (forthcoming) find mixed evidence on the effect of H-1Bs on total firm size. Peri, Shih, and Sparber (2014) find that H-1Bs increase employment of natives.

[3] Specifically, Peri, Shih, and Sparber (2014) examine the effects of H-1B visas on local labor markets using the FY2008 and FY2009 H-1B visa lotteries. However, in these years, USCIS did not record the list

effect of H-1Bs, in part because we examine data at the firm level and most previous literature on H-1Bs has examined aggregate variation in large geographic areas. Our paper also relates to previous work on the effects of immigration on innovation or productivity, including in contexts other than H-1Bs (*e.g.* Hunt and Gauthier-Loiselle 2010; Borjas and Doran 2012; see the survey in Kerr 2013). Finally, our paper relates to the long line of literature that focuses on the labor market impacts of immigration in general, not specifically in the H-1B context (see surveys in Borjas 1994; Friedberg and Hunt 1995; Freeman 2006; Dustmann et al. 2008; and Pekkala Kerr and Kerr 2011). Relative to all of these studies on H-1Bs and other immigration programs, we are the first to our knowledge to leverage true randomized variation to estimate the effect of immigration on the outcomes of the receiving economy,[4] and we are one of the first that has used administrative data.

The paper is structured as follows. Section 2 describes the policy environment that gave rise to the H-1B lotteries we study. Section 3 describes our empirical specification. Section 4 describes the data we use. Section 5 demonstrates the validity of the randomization. Section 6 describes our empirical results on patenting. Section 7 describes our results on employment. Section 8 concludes.

2. Policy Environment

The H-1B visa is the largest program for temporary skilled migration to the United States. H-1Bs are sponsored by firms, which apply to the U.S. government to

---

of lottery losers (personal correspondence with USCIS, 2009). That paper attempts to reconstruct the list of lottery losers by using Department of Labor (DOL) records on Labor Condition Applications (LCA), which must be submitted before firms can submit an H-1B application to USCIS. That paper's identification strategy assumes that conditional on having an LCA application that is approved by DOL, selection for an H-1B is random. However, many approved LCA applications end up *not* being subject to the H-1B lottery. When a firm is no longer interested in hiring the worker for which the firm had previously submitted the approved LCA application, the firm does not submit an H-1B application to USCIS. In FY 2008 and 2009, at least 20% of LCA applications are contaminated by these companies that chose to not apply for an H-1B visa (e.g. USCIS 2008, DOL 2014). This raises the concern that the analysis of that paper is confounded by demand shocks; for example, firms in areas that experience negative shocks might be less likely to submit H-1B applications to USCIS (conditional on having an approved LCA application), and one would expect that this negative shock would be correlated with subsequent economic outcomes.

[4] Clemens (2013) examines a different question using H-1B lottery data. He uses personnel records from a single firm that is a large sponsor of H-1Bs, in combination with information on the winners and losers of the FY2008 and FY2009 H-1B lotteries at this firm. He finds that winning an H-1B, and therefore working in the firm's U.S. affiliate rather than in the firm's Indian affiliate, raises the workers' wages. Edin, Fredriksson, and Åslund (2003) use variation that appears quasi-random.

obtain a visa for each H-1B worker they wish to hire. In their application for each visa, firms must specify the identity of the worker they wish to hire. An H-1B visa allows a skilled foreigner to enter the U.S. for three years, during which period the H-1B visa holder is supposed to remain at the firm (unless the worker obtains another visa or permanent residency). The H-1B is considered a "nonimmigrant" visa because it allows those with H-1Bs to stay in the U.S. only temporarily, rather than more permanently. After these three years, a number of possibilities may occur. First, the worker may leave the U.S. Second, the firm may seek to renew the worker's H-1B visa, or it can sponsor the worker to be a permanent resident. Third, the worker could exit the firm but stay in the U.S.

The firm submitting the H-1B LCA to DOL must attest, among other things, that: "(a) The employer pays H-1B non-immigrants the same wage level paid to all other individuals with similar experience and qualifications for that specific employment, or the prevailing wage for the occupation in the area of employment, whichever is higher. (b) The employment of H-1B non-immigrants does not adversely affect working conditions of workers similarly employed."[5]

We obtained data from U.S. Customs and Immigration Services (USCIS) on the lotteries for H-1B visas that were conducted for visas granted in FY2006 and FY2007. We study these lotteries in particular because USCIS did not keep lottery data for other years we have sought.[6] Specifically, the data contain information on which firms were subject to each lottery, and those that won and lost each lottery. Of the winners, the data also identify which visa applications were approved or denied.[7]

Visas given for FY2006 allowed a worker to work from October 2005 to October 2008, and visas given for FY2007 allowed a worker to work from October 2006 to October 2009. The total number of H-1B visas awarded to firms in a given year is subject to a maximum number or "cap." This cap is different for visas given to workers who have

[5] See http://www.doleta.gov/regions/reg05/Documents/eta-9035.pdf (accessed October 17, 2014).
[6] Personal communication with USCIS (06/01/2011).
[7] In order to rely on random variation, it is necessary to know which firms won and lost the lottery, as opposed to knowing simply which lottery participants had approved or denied visas. Firms with denied visas may be systematically different than those with approved visas, which would contaminate the random variation with cross-sectional variation.

only a B.A. (the "Regular" H-1B visa) and for visas given to workers who have a masters degree (the Advanced Degree Exemption (ADE) H-1B visa). The cap for regular H-1B visas was 65,000 in each year for FY2006 visas and for FY2007 visas, and the cap for ADE H-1B visas was 20,000 in each year for FY2006 visas and for FY2007 visas. In each year and for each of the two types of H-1B visa, visas are allocated by lottery to firms that applied on the date when the total number of applications reached the cap. In a given lottery, firms are allowed to apply for multiple visas. In those cases in which firms applied for multiple visas in a given lottery, the probability that the firm won each visa was independent and equal to the number of lottery winners divided by the number of lottery entrants. The lotteries were conducted by USCIS. In each of these lotteries, the total number of applications that won the lottery was equal to the number of remaining visas necessary to reach the cap. The cap does not apply to visas given for work at U.S. educational institutions, and so these visas are excluded from the lotteries (and educational institutions are excluded from the sample of firms in our lottery data).

Firms did not know in advance the date at which the cap would be reached, and they did not know the probability that firms applying on this date would be selected for an H-1B. For the FY2006 regular visa, the cap was reached on August 10, 2005; for the FY2007 regular visa, the cap was reached on May 26, 2006; for the FY2006 ADE visa, the cap was reached on January 17, 2006; and for the FY2007 visa, the cap was reached on July 26, 2006 (personal correspondence with USCIS, 2009). These dates were not announced in advance but instead were an implication of the number of applications that happened to occur on different dates in these years, making it effectively impossible for firms to successfully game the system and apply for more than they desire.[8] Each of the lotteries was conducted within a month of reaching the cap for that lottery.

For a given lottery year (*i.e.* FY2006 or FY2007), we refer to the calendar year that the lottery occurred (*i.e.* 2005 in the case of the FY2006 lottery, and 2006 in the case of the FY2007 lottery) as "Year 0." The year before this calendar year is "Year -1"; the year after Year 0 is "Year 1"; and so on. We refer to the first quarter when an H-1B

[8] These were also the first two years USCIS used a lottery to allocate H-1B slots, and it was not announced in advance that lotteries were going to be run for FY2006. Our discussions with executives at firms hiring H-1Bs have indicated that firms apply for the number of H-1Bs they desire, rather than gaming the system by applying for more than the number that they desire in order to end up with the number they desire.

employee would begin work at a firm (*i.e.* the first quarter of FY2006 in the case of the FY 2006 lottery, or the first quarter of FY2007 in the case of the FY2007 lottery) as "Q1"; we refer to the next quarter as "Q2"; and so on. A fiscal year begins in October of the previous calendar year; for example, Q1 of FY2006 corresponds to the fourth quarter of calendar year 2005 (i.e. October to December of calendar year 2005).

3. Empirical Strategy

Our empirical strategy exploits the random assignment of H-1B visas in the lotteries. Thus, we consider only the sample of firms that entered the FY2006 or FY2007 H-1B lotteries. Our main outcomes of interest are patenting and number of employees.

After a firm wins an H-1B lottery, its application may be approved, denied, or withdrawn. For example, the application may not have met the eligibility criteria, leading to a denial, or the applicant firm may go out of business, leading to a withdrawal. As a result, the total number of H-1B visas approved in any given year from the sample that applies for the lottery depends also on the fraction of those that win the lottery that also are approved, which represents potentially endogenous variation. Thus, we exploit the lottery to provide an instrument for approved H-1B visas.

Our strategy must accommodate firms that applied for multiple H-1B visas. If a firm submits n visa applications to a lottery in which p percent of lottery applicants won an H-1B visa, and W is the number of H-1B visas awarded to the firm, the expected number of H-1B visas awarded to the firm is $E[W]=pn$. If the actual number of visas won is w, then the number of unexpected wins $u=w-pn$ reflects the random realization of the net number of *unexpected wins* (or losses) and will be orthogonal to the error in the regression we specify below. Thus, our main instrument for the number of approved H-1B visas is the random variable U, the net number of unexpected wins (or losses) (whose realization is u).

In order to determine the causal effect of an approved H-1B on the outcome, we run a two-stage least squares model:

$$A_{iT} = \alpha_0+\alpha_1 U_{iT}+\nu_{iT} \qquad (1)$$

$$Y_{itT} = \beta_0+\beta_1 A_{iT}+\eta_{itT} \qquad (2)$$

Here t is defined as the number of calendar years since the lottery in question occurred; for example, $t=0$ corresponds to Year 0, *i.e.* 2005 in the case of the FY2006 lottery, or 2006 in the case of the FY2007 lottery. We run this regression separately for different choices of t. T indexes the year of the lottery in question, *i.e.* FY2006 or FY2007. A_{iT} represents the number of H-1B visas approved for this firm in the lottery that occurred in year T. In the first stage (1), we regress approved H-1B visas A_{iT} for firm i in lottery T on U_{iT}, which represents the number of unexpected wins in a firm in a given year (i.e. the year 2006 or 2007). Y_{itT} represents the time period t level of an outcome (*e.g.* patenting) in firm i that participated in a lottery in year T. In the second stage (2), we regress Y_{itT} on approved H-1B visas A_{iT} (instrumented using U_{iT}). We interpret the coefficient β_1 as a local average treatment effect of an extra approved H-1B visa among the compliers (i.e. those induced by winning the lottery to have an extra approved H-1B visa).

In those cases in which a firm participates in more than one lottery in a given fiscal year T (*e.g.* a firm participates in both the 2006 regular and ADE lotteries), we calculate U_{iT} by summing the total number of unexpected wins across both of the lotteries that the firm enters in year T (except for specifications in which we run separate regressions for the Regular and ADE lotteries).[9] We seek as much statistical power as possible, and so we pool the FY2006 and FY2007 ADE and regular lotteries in our main specification. (We also investigate the results separately in different combinations of lotteries.) In these pooled regressions, for a given firm, we stack data corresponding to the FY2006 lottery and data corresponding to the FY2007 lottery, so that we can capture

[9] We verified that winning a slot in one lottery does not affect the probability of applying for subsequent H-1B visas. For example, in both the case of FY2006 and FY2007 visas, the Regular visa lottery chronologically occurred on a date before the ADE cap was reached. We also verified that unexpected wins in earlier lotteries have no significant effect on the probability of applying for or obtaining subsequent H-1B visas. To give a sense of these results, when we pool FY2006 and FY2007 and regress total ADE H-1B visa approvals in a given year on unexpected lottery wins in the Regular lottery in that year, we find a coefficient on unexpected lottery wins of -0.20, with a standard error or 0.18 (insignificant at conventional levels, p=0.26). We also find that unexpected lottery wins in 2006 have no effect on approved 2007 visas; for example, when regress total FY2007 Regular and ADE approvals (summed) on unexpected lottery wins in the FY2006 Regular and ADE lotteries combined, we obtain a coefficient on unexpected lottery wins of -0.05, with a standard error of 1.45 (insignificant at conventional levels, p=0.97). Finally, winning one lottery also does not affect the probability of winning a subsequent lottery conditional on entering the subsequent lottery, both according to USCIS and as we have verified empirically.

the effects of winning the lottery in Year 0 on employment in each subsequent year (measured consistently as number of years since the lottery in question occurred). ν_{it} and η_{it} are error terms. We cluster our standard errors at the level of the firm to account for intra-cluster correlations (including those resulting from stacking the data).

Although the randomization implies that U_i should be orthogonal to the error in (1), it is also possible to control for various pre-determined covariates (as many papers involving randomized experiments do). For example, we can control for a lagged value of an outcome variable at the firm (*e.g.* in the case in which the dependent variable is the number of patents, we can control for $Y_{i,-3 \text{ to } -1,T}$, the number of patents in firm i observed from Year -3 to Year -1 (inclusive), where year is measured relative to lottery T); for the expected number of lottery wins pn; or other covariates.

Previous literature has not examined the level of patenting due to the volatility of this variable; instead, it has examined transformations of the number of patents that reduce volatility. Given the approximate lognormality of patents, one may wish to run a specification in which log patents forms the dependent variable (as in, for example, Kerr and Lincoln 2010). However, in our context, estimating exactly this specification would lead to a problem: we would like to include firms in the regressions that have zero patents, as a large fraction of firms have zero patents in our context, but the log of zero is undefined.[10] Thus, we approximate the log of the number of patents using the inverse hyperbolic sine of the number of patents. The inverse hyperbolic sine approximates the log function but is defined at zero and negative values (*e.g.* see related work in Burbidge, Magee, and Robb 1988, Pence 2006, or Gelber 2011). The inverse hyperbolic sine of patents Y is defined as:[11]

$$\sinh^{-1}(Y) = \ln(Y + \sqrt{1 + Y^2})$$

In the specifications in which the inverse hyperbolic sine of the number of patents is the dependent variable, the coefficient on approved H-1B visas approximately reflects the percent increase in patents caused by an extra H-1B visa.

[10] This is not a problem in the context of Kerr and Lincoln (2010); they examine patents at the city level, where patents are greater than zero.
[11] A more general form of the inverse hyperbolic sine function adds a scaling parameter; our results are similar when we use other scaling parameters.

Another way of reducing noise is to investigate a binary outcome, specifically a dummy for whether the firm patented. In this case, rather than controlling for the number of patents from Year -3 to Year -1, we control instead for a dummy for whether the firm patented between Year -3 and Year -1. Since we investigate a panel of data, when we investigate a binary outcome, we run a linear probability model to avoid an incidental parameters problem.[12]

In the case of the employment outcome, we run a related set of specifications. As in the patenting context, previous literature has not examined the level of employment as a dependent variable, but has instead examined transformations employment, such as the log, that reduce volatility (*e.g.* Pekkala Kerr, Kerr, and Lincoln forthcoming). As we show, the employment outcome is much more volatile (i.e. has a much larger standard deviation) than the patenting outcomes we investigate. As a result, noise in the dependent variable is an especially important issue in the employment context, given that the variation in the dependent variable (employment) is very large relative to the variation in the key independent variable (unexpected lottery wins). Our main way of addressing this issue is by running median regressions in our baseline specification in the employment context. In these median regressions, we are unable to run quantile instrumental variables regressions because of a practical consideration: they typically did not converge. As a result, we run "reduced form" median regressions, in which we perform a median regression of the outcome directly on unexpected lottery wins:

$$Y_{itT} = \beta_0+\beta_1 U_{iT}+\varepsilon_{itT} \qquad (3)$$

As before, we are able to add various controls to this regression. In interpreting these "reduced form" regressions, it is worth noting that the first stage regressions corresponding to equation (1) that we show later are extremely strong, with first stage F-statistics ranging from 239.94 to 993.51 in baseline specifications, and have first stage coefficients near 1 (ranging from 0.86 to 0.88).

Our second method of addressing noise in the employment variable involves two-stage least squares regressions with winsorization. Just as unexpected lottery wins are orthogonal to the error when Y_{itT} is the dependent variable, they are also orthogonal to the

[12] We would run into an incidental parameters problem with logits or probits in the case of binary outcomes, or with negative binomial or Poisson regressions in the case of count outcomes.

error when the first difference ΔY_{itT} is the dependent variable. We run the following two-stage least squares regressions, where the regression in each stage is run using ordinary least squares:

$$A_{iT} = \alpha_0+\alpha_1 U_{iT}+\nu_{iT} \qquad (4)$$

$$\Delta Y_{itT} = \beta_0+\beta_1 A_{iT}+\eta_{itT} \qquad (5)$$

The first difference ΔY_{itT} is taken from before the lottery, in Year -1 (*i.e.* the first quarter of 2005 for FY2006 visa applicants and the first quarter of 2006 for FY2007 visa applicants), to time period t after the lottery. The 95th percentile of the first difference in employment is 352, and the 5th percentile is -109, which are very large in absolute value relative to the variation in unexpected lottery wins; to help in reducing noise, we winsorize the dependent variable at the 95th percentile before running these regressions. Winsorization is common in administrative data (*e.g.* Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan 2011) and in survey data (*e.g.* the topcoding in the Current Population Survey). In these regressions, we also typically additionally control for the lagged dependent variable, specifically employment in firm i observed in year -1 relative to lottery T, $Y_{i,-1,T}$, which in practice helps in reducing the variance introduced by the first-differencing.[13]

One potential concern about the winsorized regressions is that if an extra H-1B worker can lead to a large increase in employment at the firm, this will not be captured in the winsorized version of the regressions. However, in practice when we run the version of (4)-(5) without winsorizing, the point estimate of the coefficient on H-1B visas is negative and insignificant (as it is in the quantile regressions), which lessens the worry that the winsorization dulls an actual positive effect. We have also found that winning an

[13] Of course, if we did not winsorize, running regressions (4)-(5) while additionally controlling for Year -1 employment (as we often do) is equivalent to simply controlling for Year -1 employment with the Year t level (rather than first difference) of employment as the dependent variable, since the coefficient on Year -1 employment mechanically changes by exactly 1 from the specification with the Year -1 control to the specification without. However, given that we do winsorize the dependent variable, (4)-(5) give different results than those obtained from controlling for Year -1 with the year t level of employment as the dependent variable. We winsorize the first difference of employment and control for lagged employment, rather than winsorizing the level of employment in period t after the lottery and controlling for lagged employment, again because in the context of examining firms of all sizes, winsorizing the first difference is more effective in removing large outliers than is winsorizing the level of employment. When we limit the sample to smaller firms, the two specifications show very similar results, with similar point estimates and confidence intervals.

extra H-1B visa has an insignificant effect on the probability that the change in employment is outside the 95th percentile. Nonetheless, because of these potential concerns about the winsorized specification, we consider the quantile regressions to be our primary regressions in the employment context.

A third way of addressing noise in the employment variable is to estimate the effect on the (first-differenced) inverse hyperbolic sine of employment. Again, we do not estimate the effect on the log of employment because employment sometimes takes a zero value, and the log of zero is undefined.

4. Data

Match between USCIS Data and patenting data

We merge a number of administrative datasets. First, we use USCIS administrative data on the H-1B lotteries for FY2006 and FY2007. The data contain information on each H-1B visa application that entered in the lottery in each of these years, for both regular and ADE H-1Bs. These data contain information on Employer Identification Number (EIN); the exact date the firm applied for a visa; the type of H-1B (regular or ADE); the name of the firm that applied for the H-1B; whether the H-1B application won or lost the lottery; and whether the H-1B application was ultimately approved or denied by USCIS.

We obtained data on U.S. patents from the Patent Dataverse from 1975 to 2013.[14] This database contains data on the universe of U.S. patents granted in these years, based on USPTO data. We use data from the Patent Dataverse on firm name and the number of patents granted in each calendar year. (The Patent Dataverse does not contain data on firm EINs.) Patents are classified by the calendar year in which a firm applied for the patent. Thus, for example, our measure of the number of patents at a given firm in Year 0 reflects the number of patents the firm applied for in Year 0 that were approved by 2013. The time to develop a patent can range from months to years, with substantial variance.

[14] We thank Lee Fleming for sharing these data with us. These data build upon the Harvard Business School Patent Dataverse, which contains data from only 1975 to 2010, by updating the sample to 2013. The original data covering patents granted through 2010 may be found at https://thedata.harvard.edu/dvn/dv/patent/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/15705&studyListingIndex=1_403d45eba801962a7a6ca2b83323 (accessed Sept. 20, 2014).

In a typical case, a patent is approved in a matter of two to three years—for example, the mean approval time reported by USPTO for patents filed in FY2008 is 32.2 months (USPTO 2012)—although there is again substantial variance. Since it may take a number of years for firms to develop patents and apply for them, or for these patents to be approved, we separately examine patenting over the full sample period of seven years (Years 0 to 6); over the first three years after the H-1B lottery (Years 0 to 2); and over the subsequent four years (Years 3 to 6). We ultimately find comparable results over all of these time periods. Our data will allow us to estimate the effect on an important set of patents—i.e. those that could have been developed and approved within seven years of the initial H-1B arrival at the firm—but the effect on patents that may be approved in the future is unobserved.

Since the Patent Dataverse does not contain EIN, but does contain firm name, we matched data from the Patent Dataverse to the USCIS lottery data using firm names. As we describe in greater detail in the Appendix, to match firms between these two datasets, we performed an intentionally liberal automatic matching procedure between these datasets in order to obtain all plausible matches between companies and patents. We then searched through the matches by hand in order to detect and remove all matches that appeared spurious. We classified firms into three categories: (1) 392 firms that definitely matched between the Patent Dataverse and the USCIS data; (2) 63 firms that possibly matched (*i.e.* it was ambiguous whether they matched); and (3) the remaining firms that definitely did not match. In our main results, we exclude the 63 possible matches from the list of matched companies. In the Appendix, we show that the results are robust to assuming that the possible matches were in fact true matches. In general, our results are robust to alternative assumptions and similar alternative matching procedures.

Match between USCIS data and IRS data

Using firms' EIN, we also merged the USCIS lottery data to IRS data on the universe of U.S. employment. These IRS data contain information on overall firm employment (among other outcomes) for each EIN. We are not able to link individual employees from the USCIS data to the IRS data. Employment as measured in our IRS data in a given quarter reflects employment at the firm in that quarter, from IRS form

941. In our data, the measure of employment in this quarter refers to the "number of employees who received wages, tips, or other compensation for the pay period including…Dec. 12 (Quarter 4)."[15] As a result, our measure of employment in Q1 will be influenced by hiring decisions that firms can make until December of that quarter. Thus, between the time when a firmed learned that it won or lost the lottery in June to August of Year -1, and the end of Q1, when workers generally begin working at the firm and after which employment is measured, firms had a number of months to react by hiring other worker(s), or not. For example, firms were notified of the FY2007 regular visa lottery results in June of 2006, which gave firms over six months until the last month of the first quarter of FY2007, which occurred in December of calendar year 2006. However, in the sole case of the FY2006 ADE lottery, the lottery was held on January 17, 2006, *after* Q1 of FY2006 ended. Thus, in the employment regressions, we drop data corresponding to Q1 of the FY2006 ADE lottery, since firms' hiring decisions in Q1 could not have been influenced by the results of the lottery.

We use data from 2004 to 2007. The first IRS data available from form 941 are in the first quarter of calendar year 2004. We lack form 941 data on the second through fourth quarters of calendar year 2004, and thus we measure employment in calendar year 2004 using the first quarter of calendar year 2004. We are able to examine outcomes until up to one year after the initial date an H-1B worker is first employed at a firm, which occurs in the last quarter of calendar year 2007 in the case of the FY2007 H-1B lottery.

2.0 percent of the firms in the USCIS data did not match to the EIN master list in the IRS data. We drop these firms. Pooling over all quarters, 4.5 percent of the remaining firms in the USCIS data did not match to the quarterly firm employment in the IRS data; we likewise treat this data as missing. We make additional restrictions in the employment data: of the remaining firms, 17.9 percent have missing employment data in Year -1, which makes it impossible to run our specifications (in which we control for Year -1 employment), and we drop these data for the purposes of the employment specifications. Of the remaining observations, pooling over Q1-Q4, 2.2 percent are missing in a given quarter. We verify in Appendix Table 4 that appearing as missing (conditional on the

[15] See http://www.irs.gov/pub/irs-pdf/f941.pdf (accessed October 16, 2014).

other restrictions) is unrelated to exogenous variation in H-1Bs, and we verify in Table 2 that the other sample restrictions are also unrelated to this exogenous variation in H-1Bs.

Summary statistics

Table 1 shows summary statistics. We use data on 3,050 firms.[16] The mean number of approved patents per firm in this sample is 37.74. The standard deviation of patents is very large, 390.95, due to a small number of firms—typically very large firms—that patent in large numbers. 9.3 percent of firms in this sample have approved patents. The mean (0.33) and standard deviation (1.28) of the inverse hyperbolic sine of the number of patents are much lower. Due to the large standard deviation of patenting in this full sample, and because an extra H-1B worker represents only a small fraction of average employment at a firm in the full sample, it will also prove illuminating to examine patenting in smaller firms. There are 1,276 firms with 30 or fewer employees. 3.3 percent of these firms patent; the mean number of patents is 1.92; the standard deviation of number of patents, 61.74, is much lower than in the full sample; and the mean (0.064) and standard deviation (0.37) of the inverse hyperbolic sine of number of patents is still lower. Moving to still smaller firms, there are 749 firms with 10 or fewer employees. 2.5 percent of these firms patent; the mean number of patents is 0.19 (or 0.027 patents per year); the standard deviation is 2.87; and the mean and standard deviation of the inverse hyperbolic sine of number of patents is are 0.048 and 0.34, respectively.

Another key outcome is employment. The mean number of employees over Q1-Q4 in the full sample of firms is 1,877.84, and the standard deviation is very large, 39,721.31. In firms with 30 or fewer employees in Year -1, the mean and standard deviation of Q1-Q4 employment are much lower but still large: 43.09 and 1,904.34, respectively. Finally, in firms with 10 or fewer employees in Year -1, the mean of Q1-Q4 employment is lower (9.64), but the standard deviation is still large (55.63). These summary statistics make clear that in the sample of firms with 10 or fewer employees, an extra H-1B worker represents a substantial fraction of mean employment in the sample.

[16] "Firm" refers to an EIN.

As a result, in much of our results, we focus on smaller firms, in which we might *a priori* expect that an extra H-1B worker might have a noticeable effect on the outcomes.

As we discussed in the Empirical Specification section, median regressions are our baseline specification in the employment context. The median number of employees in the sample of all firms over Q1-Q4 is 31. Among those in Year -1 with 30 or fewer employees, or 10 or fewer employees, the median number of employees over Q1-Q4 is unsurprisingly much smaller (10 and 6, respectively).

When considering whether an H-1B affects a firm's change in employment from before to after the new H-1B, we examine the winsorized first difference of employment (from the first quarter of calendar Year -1 to a given quarter of Year 0). This specification is also motivated by the large standard deviations in employment noted above. Despite winsorizing, which reduces the mean and variance by orders of magnitude, the mean (27.28) and standard deviation (92.39) of the winsorized first difference is still large in the full sample (and is also large relative to the standard deviation of the number of H-1B visas). The standard deviation of winsorized employment is substantially lower when we consider firms with 30 or fewer, or 10 or fewer, employees in Year -1, although they are still substantially larger than the standard deviation of patents in these samples.

The next rows of Table 1 show data at the level of the visa application, rather than showing data at the level of the firm or firm-quarter. The sample contains 7,243 visa applications, with an average of 2.37 H-1B applications per firm over both years, or 1.19 applications per firm per year. We show the fraction winning each of the four lotteries. For the FY2006 regular visa, 2,687 H-1B applications entered the lottery, of which 103 (3.8 percent) won the lottery. For the FY2006 ADE visa, 305 applications entered the lottery, of which 51 (17 percent) won the lottery. For the FY2007 regular visa, 3,955 applications entered the lottery, of which 3,863 (98 percent) won. Finally, for the FY2007 ADE visa, 295 firms entered the lottery, of which 163 (55 percent) won. Thus, in the FY2006 regular lottery the vast majority of firms lost the lottery, and in the FY2007 regular lottery the vast majority of firms won the lottery, whereas the ADE lotteries have a more even fraction of winners and losers; this will not pose a problem for us, as the standard errors we estimate will determine how precise the estimates are. The average firm that entered at least one lottery won 0.57 H-1B visas when aggregating across both years, or 0.29 per year.

Finally, the mean of the number of unexpected lottery wins (defined above) is 0.00, as expected, and its standard deviation is 0.33. The range of this variable runs from -2.65 to 2.96.

5. Validity of Randomization

Table 2 verifies the validity of the randomized design by regressing various pre-determined variables that could not be affected by the lottery on unexpected lottery wins. The table confirms that none of the pre-determined variables is significantly related to unexpected lottery wins. Given the random nature of the lottery, this is to be expected.

We begin by assessing whether our match of firms from the USCIS lottery data to other datasets is balanced between lottery winners and losers. Among lottery participants, we separately regress several dummy variables on unexpected lottery wins: a dummy for whether the USCIS lottery data have information on the firm's EIN (27 firms do not); a dummy for whether a firm's EIN in the USCIS data matches to the EIN of a firm in the IRS master file on the universe of U.S. EINs; and a dummy for whether a firm's EIN in the USCIS data matches to the EIN of a firm in the IRS quarterly employment data. In all cases, we find insignificant coefficients on unexpected lottery wins, with small standard errors.

Variables measuring the lagged dependent variable also show no significant correlation with unexpected lottery wins. We regress three measures of patents on unexpected lottery wins: total approved patents from a placebo period of three years prior to receiving the H-1B, Year -3 to Year -1 (inclusive); the inverse hyperbolic sine of the number of patents over this period; and a dummy for whether the firm patented over this period. These are insignificant when we use all firms in the sample, those with 30 or fewer employees in Year -1, and those with 10 or fewer employees in Year -1.

Using regression specifications parallel to those we implement for the employment outcomes, we also demonstrate that pre-determined measures of firm employment are not significantly correlated with unexpected lottery wins. When we investigate the pre-period in the employment context, we examine only Years -1 and -2, rather than examining a longer pre-period such as all years from Year -3 to Year -1 (as in the case of the patenting data), because the IRS quarterly employment data begin in the first quarter of year 2004, which we refer to as Year -2. We perform quantile regressions of employment in Year -2 on unexpected wins and Year -1 employment, and we also winsorize employment in Year -2 at the 95th percentile and regress this on unexpected wins. We control for Year -1 employment here in order to parallel the control for Year -1 employment in our main employment regressions in Table 5. Across all firm size cutoffs (all firms, those with 30 or fewer, and those with 10 or fewer in Year -1) and all outcomes, we find insignificant coefficients on unexpected wins. We also find an insignificant coefficient on unexpected wins when the dependent variable is the first-difference of employment from Year -2 to Year -1, regardless of the controls that we use.

In order to examine a period closer to Year 0, we also show that employment in Year -1 is uncorrelated with unexpected lottery wins in the sample of all firms.[17] These regressions also fail to yield significant coefficients on unexpected wins, albeit with more imprecision relative to the regressions in which we investigate the effect on Year -2 employment controlling for Year -1 employment. When Year -1 employment is the dependent variable and we control for Year -2 employment (not shown), we estimate an insignificant effect with precision similar to the regressions in which we investigate the effect on Year -2 employment and control for Year -1 employment.

Finally, we find that a dummy for whether the firm has North American Industry Classification System (NAICS) code 54—representing professional, scientific, and technical services—is insignificantly related to treatment. Firms in this sample represent 56.43 percent of the sample. In general, the sample of firms that entered the lotteries are similar to the full set of firms that receive H-1Bs; for example, in the full set of firms with approved H-1B visas, 57.47 are in are in professional, scientific, and technical services.[18] We also regressed lottery wins on dummies for all two-digit NAICS codes and

---

[17] In the specifications in which employment in Year -1 is the dependent variable, we clearly cannot control for Year -1 employment—thus increasing the standard error in the regressions relative to those in which we investigate the effect on Year -2 employment controlling for Year -1 employment. When Year -1 employment is the dependent variable, we only investigate the results in the sample of firms of all sizes because selecting this sample based on Year -1 employment could lead to biased and inconsistent results.
[18] See
http://www.uscis.gov/sites/default/files/USCIS/New%20Structure/2nd%20Level%20(Left%20Nav%20Par

perform an F-test for joint significance of these dummies; this test showed insignificant results (for example, when using the sample of all firms, p=0.96).

6. Patenting Results

Having verified the validity of the randomization, we estimate the effect of approved H-1B visas on patenting outcomes. As described above, without having seen the results we present below, one might think that a firm receiving an extra H-1B could plausibly have a noticeable effect on outcomes, because among small firms the extra H-1B worker can represent a substantial fraction of the workers at the firm. To evaluate how the effects vary across firms of different sizes (where a substantial effect on patenting might be *a priori* expected to a greater or lesser extent), we investigate the effect in the sample of firms with 10 or fewer employees (which represents roughly the 25th percentile of firm size in our sample); in the sample of firms with 30 or fewer employees (which represents roughly the 50th percentile of firms in our sample); and in the sample of firms of all sizes.

We investigate three main outcomes. First, we investigate the effect on the inverse hyperbolic sine of patents. Second, we investigate the effect on the number of patents the firm files. Finally, we investigate the effect on a dummy for whether the firm filed for a patent. We investigate each of these outcomes separately over Years 0 to 6 (inclusive); Years 0 to 2 (inclusive); and Years 3 to 6 (inclusive). We focus the most attention on the inverse hyperbolic sine of patents. This measure has the virtue of reflecting changes in patenting at the intensive margin (as opposed to the patenting dummy), while accounting for the approximate lognormal distribution of the patenting variable and allowing greater precision than the large-standard-deviation measure of number of patents. Moreover, among firms of all sizes the results are not sensitive to outliers when the inverse hyperbolic sine of patents is the dependent variable, which is unsurprising given that the inverse hyperbolic sine transformation should reduce the influence of such outliers.

For each of these outcomes, we show the results with two alternative sets of controls: (a) controlling for the number of patents from Year -3 to Year -1; and (b)

---

ents)/Resources%20-%202nd%20Level/h1b_fy07_characteristics_report_30mar09.pdf (accessed September 25, 2014).

controlling for the number of patents from Year -3 to Year -1, as well as the expected number of lottery wins (conditional on the number of H-1B applications and the probability of winning the lottery in question to which the firm applied). We take specification (b), with the larger set of controls, as our baseline, though the results are similar either way. The results are nearly identical when we control for additional or alternative controls, such as controlling additionally for the two-digit NAICS code of the firm, controlling for the firm's number of H-1B lottery applications *n*, and/or controlling for dummies for each of the four lotteries (2006 regular, 2006 ADE, 2007 regular, and 2007 ADE). We also find very similar results when our pre-period measure of patents measures the number of patents over other time periods.

Panel A shows the results for firms with 10 or fewer employees. We estimate precise, insignificant effects in all specifications. Row A shows the effect in Years 0 to 6. The point estimates are near zero across all six specifications, with some positive and some negative, but more negative than positive. In the baseline specification with the broader set of controls, the upper end of the 95 percent confidence interval rules out a large effect in most cases. When the dependent variable is the inverse hyperbolic sine of the number of patents, the upper end of the 95 percent confidence interval enables us able to rule out an increase in patents of more than 2.1 percent, relative to a "base" mean number of patents of 0.027 per year.[19] When the dependent variable is the number of patents, the upper end of the 95 percent confidence interval is 0.057, indicating that an extra H-1B does not raise the total number of patents over these seven years by more than 0.057 (or by more than 0.0081 per year). When we normalize the number of patents by the standard deviation of patents in the control group—so that the dependent variable is the number of patents divided by the standard deviation of patents—the upper end of the 95 percent confidence interval is 0.020, indicating that an extra H-1B does not raise the number of patents by more than two percent of a standard deviation. Turning to the final columns, where the dependent variable is the dummy for whether the firm patented, the upper end of the 95 percent confidence interval in the baseline is 0.010, indicating that an

[19] We calculate the "base" of 0.027 by taking the mean number of yearly patents in Years 0-6 in a "control group," specifically firms whose number of unexpected wins was less than or equal to zero. In samples of firms below other size thresholds, we calculated the "base" number of patents analogously.

extra H-1B does not raise the probability of patenting over seven years (*i.e.* Years 0 to 6) by more than 1.0 percentage point. All of these results are similar when controlling only for prior patents. Rows B and C show that we estimate comparable results when we limit the period over which we observe the outcome to Years 0 to 2, or to Years 3 to 6. While the point estimates are sometimes below zero, we would not conclude from these estimates that H-1Bs actually decrease patenting, as we can never rule out a decrease of zero at any standard significance level; of course, this is why confidence intervals are useful in determining what we can rule out with a standard degree of statistical certainty.

Panel B shows the results for firms with 30 or fewer employees. These results also show small coefficients with narrow confidence intervals, although the confidence intervals are somewhat wider than in Panel A (which is unsurprising given the much larger standard deviation of patents among these firms). The point estimates are again a mix of positive and negative estimates. The inverse hyperbolic sine results show that in the baseline we can bound the increase in patents over the full period below 3.4 percent, relative to a yearly mean number of patents of 0.27. The point estimates and confidence intervals are also in the same range when we consider Years 0 to 2 or Years 3 to 6.

Panel C shows the results for all firms. In the baseline specification over Years 0 to 6 in the inverse hyperbolic sine context, the upper end of the 95 percent confidence interval rules out an increase greater than 1.3 percentage points, on a yearly mean of 4.87. When we consider the number of patents, the results are extremely imprecise, which is unsurprising since the standard deviation of patents in this sample is so large, and since an extra H-1B worker represents only a small fraction of mean employment in the full sample of firms. The positive point estimate in this context is very sensitive to outliers; for example, when we winsorize the number of patents at the 99th percentile in the sample of all firms, we obtain negative point estimates, but the estimates are similarly imprecise and insignificant. Moreover, aside from when number of patents is the dependent variable, the other point estimates are universally negative in our other specifications on the sample of firms of all sizes (*i.e.* when the dependent variable is the inverse hyperbolic sine of patents or the patenting dummy). When the dependent variable is the probability of patenting, the upper end of the 95 percent confidence interval rules out an increase greater than 2.5 percentage points.

Our choices of the number of employees in our size thresholds (*i.e.* 10 or fewer, or 30 or fewer) could certainly be varied. To examine the consequence of varying this threshold, we show Figure 1, which plots the coefficient and confidence interval on approved H-1B visas when the dependent variable is the inverse hyperbolic sine of number of patents, as a function of the size of the employer. We show the results for employers of each size from under 10 to under 500, in increments of 10.[20] The upper end of the 95 percent confidence interval ranges from near 0 to around 0.05, indicating that across all 50 choices of the employer size threshold, in the most positive case we are able to rule out an increase in patents more than around 5 percent (for employers with 20 or fewer employees in Year -1). Notably, the point estimate is positive in only one out of 50 cases (again for employers with 20 or fewer employees in Year -1), though this coefficient is not significantly different from zero. Out of 50 employer sizes (*i.e.* each threshold from 10 to 500 in increments of 10), only two show an upper end of the confidence interval above 0.03. While a few of the estimates are negative and barely significant at the 5 percent level, this is not a robust finding as the substantial majority show estimates that are insignificant at the 10% (or 5%) level.[21]

Appendix Table 2 repeats the specifications from Table 3, showing that the results are similar when we assume that those companies that possibly matched between the USCIS and patenting database in fact did actually match, rather than assuming that they did not match as in our baseline.[22]

Heterogeneity

Table 4 examines whether there is heterogeneity in the effect on patents across type of lottery or type of industry, using our baseline specification and examining effects

[20] Disclosure concerns (*i.e.* the necessity of keeping a sufficiently large number of firms in each category, to prevent the potential identification of any given firm) prevent us from going beyond 500 employees in increments of 10.

[21] When we investigate the patenting dummy, the results are similar to those shown across the entire set of firm sizes from 0 to 500. When we investigate the number of patents, the results unsurprisingly grow increasingly imprecise at larger firm sizes.

[22] When we examine the sample of all firms, the estimated effect for the full period is negative and significant at the 10 percent level when we include the full set of controls, although we do not consider this a robust finding: (a) it is not significant at more conventional levels (*i.e.* 5% or 1%); (b) it is not robust to other specifications such as using the less extensive list of controls or removing the "possible" matches; and (c) it is not matched by any significant estimate when we investigate other dependent variables, including the patenting dummy.

on patents in Years 0 to 6 combined. Column 1 examines the Regular H-1B lottery. The results are typically similar to those in the full sample—with point estimates that cluster near zero, and the upper end of the 95% confidence interval ruling out large effects—which should not be surprising since 85.96% of the full sample participates in the Regular lottery. The confidence intervals rule out changes in patents greater than 2.8 percent, 5.0 percent, and 3.2 among firms with 10 or fewer employees, 30 or fewer employees, or firms of all sizes, respectively.

Column 2 examines the ADE lottery. The point estimates are all negative and insignificantly different from zero. Among firms with fewer than 10 or fewer than 30 employees, the confidence intervals rule out large effects. However, the confidence intervals are larger and have a larger upper end of the 95% confidence interval than in the case of the Regular lottery—consistent with the loss of statistical power due to the fact that the ADE lotteries have much smaller sample sizes than the Regular lotteries.

It is also of interest to investigate how the effect varies by industry. Over half of H-1B visas are through firms in NAICS code 54 (professional, scientific, and technical services), where the effect on patenting is particularly relevant since the bulk of patents occur in this industry. We find no evidence of an effect on patenting in this industry, with negative point estimates and confidence intervals that are larger but in the same range as those in Table 3. In firms outside of NAICS code 54, we again find comparably small point estimates and confidence intervals.

When we investigate the effect separately in each year of the lottery (*i.e.* separating the FY2006 lotteries from the FY2007 lotteries), we again estimate insignificant effects in each year separately, with comparable point estimates to those in the full sample, though again with modestly larger confidence intervals. Separating the results further into four regressions, one for each of the four lotteries (FY2006 Regular, FY2006 ADE, FY2007 Regular, and FY2007 ADE), we again estimate comparable point estimates to those in the full sample, though again with somewhat larger confidence intervals (particularly in the case of the ADE lotteries).

In Table 4, we examine only our main outcome, the inverse hyperbolic sine of patents. When we examine the patenting dummy or the number of patents as the outcome, the results are similar in showing no systematic or significant patterns across different samples.

7. Effect on employment

Table 5 shows estimates of the effect of extra H-1B visas on total employment at the firm level. Like the previous tables, Table 5 shows coefficients and 95% confidence intervals on the number of approved H-1Bs. One test of interest is a two-sided test of whether the coefficient is significantly different from 0. If a coefficient were positive and significant, it would indicate that the extra H-1B worker increases total employment at the firm, as opposed simply replacing a worker that the firm would have otherwise hired. (In principle, an extra H-1B worker could decrease employment at the firm, for example if the H-1B worker works more hours or works harder than others and therefore replaces more than one other worker.) Another question of interest is a two-sided test of whether the coefficient on approved H-1B visas is significantly different from 1. If the coefficient were greater than 1, this would indicate that an H-1B worker leads to employing a greater number of other workers, for example if these workers are complementary to H-1B workers in the production process. If the coefficient is less than one, this indicates that an extra H-1B worker on average to some extent replaces other worker(s) who would otherwise have worked at the firm.

Table 5 shows a variety of specifications. The first two columns show median regressions where the dependent variable is total employment (corresponding to model (3) above), and the final two columns show two-stage least squares regressions where the dependent variable is the first difference in employment winsorized at the 95^{th} percentile (corresponding to model (4)-(5) above). We show both sets of regressions with the alternative sets of controls analogous to those in Table 3 (employment in the pre-period, or alternatively employment in the pre-period and expected lottery wins). We also show all of the regressions in all firm size categories (≤10 employees, ≤30 employees, and firms of all sizes).

We show the regressions for up to a year after the first quarter when the H-1B worker may join the firm under the H-1B visa in question. We have also investigated longer time periods, but this leads to less precise results. It is not surprising that the results gradually lose precision as we move forward further in time. As firms are subject to more shocks over time, the variance of the change in employment from Year -1 to a subsequent period increases. (Our regression specification includes Year -1 employment as a control, rather than explicitly regressing the first difference on the independent variables, but the analogous point implies that the variance of the residual increases as we move further forward in time.) Moreover, during the first year, there is a minimal amount of attrition of H-1B workers who return to their home countries or leave the firm for another reason (as discussed in North 2011), but this attrition increases with time. It is therefore of interest to determine whether there is an effect on employment within four quarters of the start of the H-1B's tenure, to determine whether there is evidence that firms immediately decrease other (non-H-1B) employment in response to receiving an extra H-1B. We show the results when pooling data from Q1, Q2, Q3, and Q4, as well as in each of these quarters separately. The effect in each of these quarters separately is of interest to document the time path of the effects. In addition, in many cases the precision of the estimates becomes weaker as we move farther from the initial quarter, so it is of interest to observe the effect in the initial quarters when the results are as precise as possible, and to document the gradual loss of precision. In quarters beyond Q4, the point estimate of the effect of unexpected H-1Bs is generally similar to those shown; is almost never positive; is sometimes significantly different from 1; and in no case does it rise above 1 (*i.e.* the point estimates continue to indicate crowdout). Thus, while our results for future quarters are less precise, our point estimates of the effect on employment in the more distant future continue to indicate crowdout.

In Table 5, we are typically able to rule out that an extra approved H-1B visa leads to an increase in employment of 1 or greater at the 1% level. However, the estimates are never significantly different from zero across all of the specifications in each of the quarters, indicating that we cannot rule out that the extra H-1B causes no increase in employment (*i.e.* we cannot rule out that an extra H-1B worker fully replaces a different worker at the firm). Moreover, the top end of the 95% confidence interval is always positive, indicating that we cannot rule out a modest to moderate positive effect. While the point estimates are typically below zero, we do not again conclude that H-1Bs actually decrease employment, because our confidence interval does not rule out a decrease of zero; of course, this is again why confidence intervals are useful in determining what we can rule out with a standard degree of statistical certainty.

Beginning with the median regressions in firms with less than 10 employees, in the baseline specification with the more extensive set of controls and pooling over all quarters (Q1 to Q4), the top end of the 95% confidence interval is 0.11, indicating that an extra H-1B leads to an increase in total employment of at most 0.11 workers. Similarly, in this specification in firms with 30 or fewer employees, the top end of the confidence interval is 0.37. When we apply this specification to the full sample of firms, we can rule out an increase greater than 0.57. Moreover, when we break down the results by quarter, we can typically bound the coefficient below one in each quarter separately.

Turning next to the two-stage least squares regressions, we again mostly find that the coefficient is significantly different from 1, at the 1% level. In a baseline specification applied to firms with 10 or fewer employees pooling across Q1 to Q4, the top end of the 95% confidence interval is 0.68. In this baseline in the sample with 30 or fewer employees, we are able to rule out a coefficient of 0.71 or greater. In this specification applied to the full sample of firms pooled across quarters, the results are extremely imprecise, and we are unable to rule out a coefficient well above 1. When we examine each quarter separately, the coefficients are insignificantly different from 1 when we examine firms of all sizes, though they remain significantly different from 1 in many other cases when we examine firms with 10 or fewer, or 30 or fewer, employees. In a two-stage least squares, linear probability model, approved H-1B visas also have an insignificant impact on the probability that a firm has zero employees, with small confidence intervals.

As in the patenting context, our choices of the number of employees in our size thresholds can be further varied. Figure 2 plots the coefficient and confidence interval on unexpected lottery wins when we run median regressions and the dependent variable is the number of employees in the firm, as a function of the size of the employer. The point estimates are never positive, though they are also insignificantly different from zero. The upper end of the 95 percent confidence interval ranges from 0.1 to 0.6, indicating that across all 50 choices of the employer size threshold, in the most positive case we are able to rule out an increase in total employees of more than 0.6. In all cases, the estimate is significantly less than 1 at the 1% level. In no case is the estimate significantly negative. Thus, we are robustly able to rule out an increase in employment due to H-1B visas that is substantially less than one-for-one (specifically, with a coefficient typically 0.6 or under).

We run a number of other specifications as robustness checks in Appendix Table 3. When we winsorize at the 99th percentile in Column 1 of Appendix Table 3, rather than at the 95th percentile as in Table 5, the coefficient is still significantly different from 1 in the sample of firms with 10 or fewer employees ($p<0.05$) and in the sample with 30 or fewer employees ($p<0.10$). Our employment specifications are not the same as those run in the patenting context (i.e. model (3), or model (4)-(5), is different than model (1)-(2)), but for parallelism with the patenting results and in order to investigate robustness, we show additional specifications in Appendix Table 3. In Column 2 of Appendix Table 3, the dependent variable is the inverse hyperbolic sine of the first difference in employment (as in the patenting context), and the coefficient is again significantly different from 1 in the sample of firms with 10 or fewer employees ($p<0.05$) and in the sample with 30 or fewer employees ($p<0.10$).[23] In Column 3, we winsorize the inverse hyperbolic sine of the first difference in employment at the 99th percentile (to address occasional outliers that appear even in the inverse hyperbolic sine) and find coefficients that are significantly different from 1 at the 5 percent level in the case of both 10 or fewer, and 30 or fewer, employees. In Column 4, we winsorize the inverse hyperbolic sine of the level of Q1-Q4 employment at the 99th percentile and find nearly identical results to those in Column 3.

[23] In the case of these inverse hyperbolic sine specifications, before testing whether a coefficient is equal to 1, we transform the coefficient from the regression (which reflects the percentage increase in employment, rather than the increase in the absolute level of employment) by multiplying it by the mean level of employment. We then test whether this transformed coefficient is equal to 1. The test results that are reported above refer to this test. We note the limitation that in calculating the implied magnitude of the effect on the level of employment, we chose to apply the coefficient to the mean level of employment because it is illustrative, but other choices are possible. When we use smaller base levels of employment than the mean, the coefficients are even more significantly different from 1. When we use larger base levels of employment, the coefficients continue to be significantly different from 1 until we reach moderately higher base levels. In other words, we find significant evidence of crowdout except in the largest firms. It is because of this limitation that we present the inverse hyperbolic sine results in the Appendix, rather than in the main tables; we consider this to be a secondary specification in the employment context. (In the patenting context, our interest is instead in the mean effect of H-1Bs on patents, as opposed to testing whether this effect is different than a fixed specific number—as in the employment context, where we test for a difference from unity.)

Finally, in Column 5 we show that the results of the median regressions are very similar when the dependent variable is the first difference of employment and we include no controls (though the results of these regressions are also very similar when we include controls, such as controlling for expected wins).

If a firm goes out of business, the data may be missing or may show zero employees. As noted, conditional on the sample restrictions, pooling over Q1-Q4, 2.2 percent of observations are missing in a given quarter. We define a firm as "out of business" if it has either zero employees or is missing the number of employees, *and* it also has either zero total payroll or is missing total payroll. Appendix Table 4 shows that approved H-1B visas have no significant effect on the probability that a firm is out of business. This specification runs parallel to the specification in the patenting context where the dummy for having a positive number of patents is the dependent variable. The results are similar with other definitions of being out of business.

The rationale for the discrepancy between the specifications run in the patenting context and those run in the employment context is described in our Empirical Specifications section, but it is worth additionally describing additional results when we run exactly parallel specifications in both contexts. In the patenting context, the median number of patents is zero, so median regressions parallel to those in the employment context would show no effect on the median number of patents, and we omit these regressions. Regressions at higher quantiles in the patenting context show precisely-estimated zeroes. When the number of patents (or the inverse hyperbolic sine of patents) is the dependent variable and we winsorize at the 95th (or 99th) percentile, parallel to those in the employment context, our results are very similar to those shown in Table 3 but are more precise and allow us to bound the maximum increase in patenting at a still lower level. When we run the two-stage least squares employment regressions but do not winsorize the dependent variable, the results are extremely imprecise among firms of all sizes or among firms with 30 or fewer employees in Year -1, which is unsurprising given the very large standard deviation of employment and large outliers. However, when we do not winsorize and run this specification among firms with 10 or fewer employees in Year -1, the top end of the 95 percent confidence interval is 0.31, and we are able to rule out a coefficient of 1 (p=0.015). In sum, running parallel specifications does not change any of our conclusions, except that our results are unsurprisingly more imprecise when we examine the level of employment and do not winsorize, relative to when we do winsorize.

We cannot fruitfully estimate the extent to which H-1B visas replace foreigners or natives separately. Although citizenship status is available through IRS data on W-2 forms, these data only have information on the individual's *most recent* citizenship status, as opposed to being measured in the year in question in our regressions (*e.g.* Year 0 or Year 1). The data on past citizenship status is unavailable. This is an important issue because a large fraction of H-1Bs go on to become permanent residents and in many cases citizens (Lowell 2000). In our baseline employment specification, we are never able to rule out that there is no effect of unexpected lottery wins on the median number of citizens (as measured by most recent citizenship status), but we are almost always able to rule out that the median number of citizens decreases by one. This could mean that H-1Bs do not displace citizens, but it could also mean that H-1Bs in some cases go on to become citizens eventually. The effect on the median number of non-citizens (again as measured by most recent citizenship status) is again always insignificantly different from zero, but in all cases we are able to rule out that it is equal to one. However, this again could mean that H-1Bs displace other non-citizens, or it could mean that H-1Bs eventually become citizens in some cases.[24]

Heterogeneity

Table 6 investigates whether there is heterogeneity in the employment results, using our baseline employment specification with median regressions and the more extensive set of controls. Interestingly, the point estimates are more negative for the Regular lotteries than for the ADE lotteries, and they are more negative for scientific

[24] We also find no significant impact on a number of related outcomes. We find no significant effect of an extra H-1B visa on the firm's reported profits or wages per employee, though these regression results are extremely imprecise (as is unsurprising given the large standard deviation of these variables). It is also possible that an unexpected H-1B lottery win affects a firm's competitors, but we find no significant impact of unexpected H-1B lottery wins on any of the outcome variables among all other firms in that firm's 6-digit NAICS code, which is unsurprising given the large size of a six-digit industry. We also find no evidence that unexpected H-1Bs affect the probability that a firm changes EINs, for example by merging with another firm.

services (*i.e.* NAICS code 54) than for other industries. In fact, the point estimates are often positive and substantial in the case of the ADE lotteries, and in the case of scientific services—particularly when we examine firms of all sizes. However, as in the case of the patenting results, there are no significant differences across the different samples. The results are similar, though typically with larger confidence intervals, when we separately examine only the 2006 lotteries or only the 2007 lotteries. In Table 6, we again examine only our main specification, the median regressions. Other specifications are similar in showing no notable patterns across different samples.

Interpreting the estimates

It is possible to address a number of factors relevant to the interpretation of our estimates. If firms respond to an extra H-1B visa by reducing contracting work or outsourcing to other firms—neither of which shows up in our measure of employment at the firm itself—then we should see the total number of employees at a given firm increase when the firm receives an extra H-1B. In other words, if there are margins of substitutability that we do not observe by examining employment at the firm itself, then H-1Bs will appear to be *less* substitutable with non-H-1B employment than it actually is. This would make the coefficient on H-1Bs *more* positive than that which we would estimate if we could observe such broader employment impacts of the firm's unexpected H-1B. In light of this consideration, it is all the more notable that we are able to rule out a coefficient on H-1Bs of one or greater. Fraud has also been alleged in the context of H-1Bs;[25] this could lead to a larger coefficient on unexpected H-1Bs (if firms fraudulently obtain other types of visas for the workers who would have been H-1Bs if the firm had been awarded an H-1B) or a smaller coefficient (if the firm responds to not receiving an H-1B by hiring a worker off the books).

In principle, another limitation of our results is that we do not observe if the worker actually ended up at a firm (as opposed to having an approved H-1B visa, which we do observe). For example, after being approved by USCIS, some workers may die before being admitted to the U.S. to start their job, or the State Department may not

[25] For example, see http://www.bloomberg.com/news/2013-10-30/infosys-settles-with-u-s-in-visa-fraud-probe.html (accessed September 16, 2014).

approve their visa. However, in practice this is likely to affect our employment results only negligibly. In the employment context, we examine (among other things) the immediate impact on employment in the first quarter of Year 0; North (2011) estimates that 95% of those approved for H-1Bs end up being admitted.[26] This would not pose an issue for our employment results, where we are typically able to rule out an increase in employment in the initial quarters that is under 0.6 (*i.e.* well under 0.95). North (2011) also estimates that 82% of workers remain at the firms for the full three years, as some workers return home or depart for another reason. This is not relevant for interpreting our employment results, where our focus is shorter-term, but it is relevant to interpreting our patenting results. Note, however, that the patenting results we later estimate would likely be similar if they were scaled up by 22% (=1/0.82).

In rare cases, workers start working at the firms after the first quarter of the first year, which affects the interpretation of our results. As noted above, we use USCIS administrative data on the proposed start dates of each H-1B application that won the lottery in FY2006 and FY2007 to calculate that 91.87 percent of H-1Bs started working at the firms under this H-1B in Q1, and 100 percent had started working at the firms by Q2. Thus, nearly everyone had started working at the firms, and this does not represent a major issue. We are unable to instrument for the number of approved H-1B visas who were working at the firms in each quarter, because for those firms with multiple H-1B applications in a given lottery (e.g. the FY2007 Regular lottery), the data do not allow us to determine which H-1B application(s) (with a given start date) won the lottery and which lost the lottery; instead, the data only report the total number winning the lottery, as well as the proposed start date on each application. Nonetheless, this is unlikely to make a substantial difference to the employment results; our estimates are sufficiently precise that our estimates would generally be unaffected by estimating slightly different coefficient (*e.g.* multiplying all coefficients by 1/0.9187, as we would if we used a first stage for the regressions in Q1 that took account of the fact that a modest percentage of H-1B workers do not arrive until Q2).

[26] In North (2011), the fraction admitted is calculated by including those who were already in the U.S. and apply for a renewal of their H-1B. Excluding these individuals would not materially change our conclusions.

In the case of the median regressions in the employment context, which are "reduced form" regressions, the coefficients do not take account of the fact that some H-1B lottery winners do not have their applications approved. However, our first stage coefficient is extremely precise and quite close to 1 (specifically, it ranges from 0.86 to 0.88), so this consideration is also unlikely to change our conclusion that H-1B workers at least partially replace other workers at the firm. Moreover, we also estimate two-stage least squares specifications in this context that show comparable results.

Importantly, even ignoring any of the considerations raised in this section, we estimate a policy-relevant parameter. The effect of increasing the H-1B visa cap on employment—the question that the "reduced form" addresses—is of policy interest, for example in projecting the effect of immigration reform on employment (*e.g.* Congressional Budget Office 2013). In the case of the patenting results, the key issue of policy interest is how patenting would be affected by approving more H-1B visas or by raising the H-1B visa cap—and these parameters are precisely what we estimate.[27]

8. Conclusion

We investigate the effect of winning an H-1B visa on firms' patenting and employment outcomes. We find an insignificant effect of an extra approved H-1B visa on patenting, including among small firms in which the extra H-1B visa reflects a substantial percentage of initial firm employment. Among these smaller firms where the H-1B could in principle make a substantial difference, we are able to rule out large positive effects. For example, our 95 percent confidence interval allows us to rule out that in firms with 10 or fewer employees, an extra H-1B visa leads to an increase in the probability that a firm patents over Years 0 to 6 of more than 2.1 percent on a base of 0.027, or that the total number of approved patents increases by more than 0.0081 per year. Across a

[27] The effect of raising the H-1B cap on patenting would correspond to the "reduced form" version of the patenting regression results shown in Table 3 and elsewhere; since the first stage coefficients are around 0.87, the coefficients in these "reduced form" regressions could be calculated by multiplying the coefficients in the two-stage least squares regressions by around 0.87 (where the precise number depends slightly on the specification). The standard errors in the reduced form regressions are slightly smaller than those shown in the two-stage least squares specifications in the tables.

variety of specifications, the preponderance of evidence allows us to rule out moderate-sized effects.

It is worth considering the implications of our patenting results. While we readily acknowledge that we have no direct evidence that the results obtained on this lottery sample would generalize, we also note that the key policy question of interest is what would happen if the H-1B cap were marginally higher, which in this context would have allowed H-1Bs to precisely some of those firms that lost the lottery that our regressions investigate. In other words, our regressions address the key counterfactual policy question for these lotteries: the effect of marginally increasing the cap. Moreover, the firms subject to the lottery have similar characteristics to those in the full sample of firms receiving H-1Bs. Consider a plausible policy option within the range contemplated by policy-makers: increasing the H-1B cap by 5,000 workers per year. If each of these workers increased patents by 1.3 percent on a base of 4.87 patents per year—*i.e.* by the amount consistent with the top end of our 95 percent confidence interval when we examine firms of all sizes, applied to the mean yearly number of patents in the control group in this full sample[28]—then the number of patents per year would increase by 316.55. Since this is the top end of our 95 percent confidence interval, we are able to rule out any larger effect at this confidence level. Compared to an average of 233,120 patents per year in the U.S. over the seven years 2007-2013, this allows us to rule out an increase in patents greater than 0.136 percent.[29] While this represents a non-trivial increase in patenting, this result strongly contrasts with the much larger positive effects on patenting and innovation found in previous literature.[30] Moreover, if we include "probable"

---

[28] When we calculate the implied effect on the level of patents by transforming the effect on the inverse hyperbolic sine of patents by the mean level, we find the implied mean effect on the level, which is precisely the object of interest in this context. Again, the "control group" consists of firms whose number of unexpected wins was less than or equal to zero.

[29] The mean number of patents (4.87 per year) and therefore the implied effect on the level of patenting is heavily influenced by outliers in the number of patents, which tend to be very large firms in which an extra H-1B represents only a tiny fraction of total employment. For example, when winsorizing the number of patents at 1,000 (which is above the 99th percentile), the mean number of patents is only 1.94 per year. If there were a 1.3 percent increase—the maximum allowed by our 95 percent confidence interval—on a base of 1.94 patents per year, then increasing the cap by 5,000 would cause a yearly increase in total patenting of only 126, or only 0.054%.

[30] For example, Kerr and Lincoln (2010) find that a 10 percent growth in a city's H-1B population corresponded with a 0.3 percent to 0.7 percent increase in total patenting for each standard deviation growth in "city dependency," a measure of H-1B applications per capita in each city. Given the standard

matches in our patent data (as in Appendix Table 2), the top end of the 95 percent confidence interval rules out that an increase in the H-1B cap of 5,000 would cause a yearly increase in patents more than 17.5, or only a 0.00751 percent increase in patents.

Despite our null result on patenting, firms may have other good reasons for applying for H-1Bs. Firms should apply for the visas if doing so increases their profit. Relative to alternative workers, H-1B workers could have higher marginal products in comparison to their pay. For example, even if there is no positive effect on patenting, it could be that H-1B workers increase firm productivity in ways not captured by patenting.

Parallel to these patenting results, we also find that H-1B workers to some extent replace other workers within one year of the beginning of the visa. In median regressions we are generally able to rule out a coefficient of 0.6 or under, indicating substantial crowdout. Just as the evidence indicates that patenting does not greatly increase, so the preponderance of evidence indicates that employment does not greatly increase.

In interpreting these results, note that H-1B workers may not represent the same quantity of labor as non-H-1B workers do. For example, as noted, H-1B workers could work a greater or smaller number of hours than non-H-1B workers do. Hours worked is unobserved in our data, as is typical in administrative datasets. In order to estimate the parameters of a formal model—such as a model attempting to estimate the elasticity of substitution between H-1B and non-H-1B workers—one would wish to know the quantity of labor that each worker represents, but we do not observe this. In order to estimate such a model, one might also wish to estimate effect of H-1B workers on capital, of which we have only noisy and incomplete proxies. The degree of crowdout of natives should depend not only on the nature of the substitutability or complementarity of H-1B and non-H-1B workers (and/or labor and capital), but also a number of other factors, including the nature of the process that matches firms with workers (possibly including search frictions). If the firm faces frictions in finding a new employee that *limit* the degree of crowdout of non-H-1B workers, it would be all the more notable that we find

deviation of city dependency in their sample, this would imply an increase in patenting at least 10 times as large as the maximum effect allowed by our 95% confidence interval. Our 95% confidence interval rules out effects as large as those Kerr and Lincoln (2010) estimate, and their 95% confidence interval rules out effects as small as those we estimate.

that an H-1B worker *does* partially replace other workers in the median case, and that we cannot rule out that an H-1B worker has no effect on total employment.[31]

Our results are not perfectly comparable with much of the previous literature, for several reasons. We examine the effect on individual firms using microdata, as opposed to the aggregate data at the level of the local labor market used in many other studies that may incorporate spillover effects. At the same time, if an H-1B has little effect on patenting in the firm receiving the H-1B as in our results, one may question whether this could lead to strong spillovers.[32] Moreover, our results only apply to H-1B visas subject to the cap, as opposed to H-1B visas given for educational institutions, extension of prior H-1B visas, and other categories that were not subject to the lottery, introducing another reason that our results are difficult to compare with some prior literature—though our estimates speak to the policy-relevant question of the effects of raising the cap.

It is worth emphasizing that H-1Bs represent only one type of high-skilled immigration, and that other types of high-skilled immigration could have very different effects. The majority of H-1B workers—including those in our sample—do not have the advanced degrees that would be most closely associated with innovation. Many H-1Bs are not in scientific industries, and among the 56.43 percent that are in scientific industries, many H-1B workers perform jobs (*e.g.* technical support) that might not be expected to lead to patenting in the overwhelming majority of cases. Moreover, our sample sizes are generally larger in the case of the Regular H-1B lotteries than in the case of the ADE lotteries, implying that it is harder to draw precise lessons about the effects of the higher-skilled ADE lottery participants. It is also possible that other types of high-skilled immigration, such as O-1 visas given to those with "extraordinary abilities" or proposals to encourage advanced degree holders to stay in the U.S., have more positive effects on patenting or employment.

Although we can address several of the narratives in the current policy debate with the lotteries we exploit in this paper, the precision of the estimates could be increased by using a larger sample size. One way to do so would be to use lottery data

[31] One question related to ours is how firms react to losing or gaining a worker exogenously—not specifically in the case of H-1Bs—which is examined in Isen (2013).

[32] In principle, H-1B workers could also be patenting on their own, not through their firm.

from a larger set of years. It would be helpful for USCIS to begin saving the data on H-1B lottery winners and losers in other lottery year(s), so that the statistical power of the estimates could be increased in the future.

References

Altonji, Joseph, and David E. Card. "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives." In John Abowd and Richard B. Freeman, eds., *Immigration, Trade, and the Labor Market*. Chicago: U of Chicago Press, 1991. 201-34.

Borjas, George. "The Labor Demand Curve Is Downward-Sloping: Reexamining the Impact of Immigration on the Labor Market." *Quarterly Journal of Economics* 118 (2003): 1335-374.

Borjas, George. "The Economics of Immigration." *Journal of Economic Literature* 32 (1994): 1667–1717.

Borjas, George, and Kirk Doran. "Cognitive Mobility: Native Responses to Supply Shocks in the Space of Ideas." *Journal of Labor Economics* (forthcoming 2015).

Borjas, George, and Kirk Doran. "The Collapse of the Soviet Union and the Productivity of American Mathematicians." *Quarterly Journal of Economics* 127.3 (2012): 1143-203.

Borjas, George, Richard Freeman, and Lawrence Katz. "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity* (1997): 1-90.

Burbidge, John, Lonnie Magee, and A. Leslie Robb. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association* 83 (1988): 123-127.

Card, David. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43.2 (1990): 245-57.

Card, David. "Immigrant Inflows, Native Outflows, and the Local Market Impacts of Higher Immigration." *Journal of Labor Economics* 19.1 (2001): 22-64.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. "How does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *The Quarterly Journal of Economics* 126.4 (2011): 1593-1660.

Clemens, Michael. "Why Do Programmers Earn More in Houston than Hyderabad? Evidence from Randomized Processing of U.S. Visas." *American Economic Review Papers and Proceedings* 103.3 (2013): 198-202.

Congressional Budget Office. "The Economic Impact of S. 744, the Border Security, Economic Opportunity, and Immigration Modernization Act." Web.

http://www.cbo.gov/sites/default/files/44346-Immigration.pdf (accessed October 7, 2014).

Dustmann, Christian, Albrecht Glitz, and Tommaso Frattini. "The Labour Market Impacts of Immigration." *Oxford Review of Economic Policy* 24.3 (2008): 477-94.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. "Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment." *The Quarterly Journal of Economics* 118.1 (2003): 329-57.

Freeman, Richard. "People Flows in Globalization." *Journal of Economic Perspectives* 20.2 (2006): 145–70.

Friedberg, Rachel M. "The Impact of Mass Migration on the Israeli Labor Market." *The Quarterly Journal of Economics* 116.4 (2001): 1373-408.

Friedberg, Rachel, and Jennifer Hunt. "The Impact of Immigrants on Host Country Wages, Employment and Growth." *Journal of Economic Perspectives* 9.2 (1995): 23-44.

Gates, William H. "Testimony before the Committee on Science and Technology." U.S. House of Representatives, Washington D.C. 12 March. 2008. Address.

Gelber, Alexander M. "How Do 401(k)s Affect Saving? Evidence from Changes in 401(k) Eligibility." *American Economic Journal: Economic Policy* 3.4 (2011): 103-22.

Hunt, Jennifer. "Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa." *Journal of Labor Economics* 29.3 (2011): 417-57.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. "How Much Does Immigration Boost Innovation." *American Economic Journal:Macroeconomics* 2.2 (2010): 31-56.

Isen, Adam. "Dying to know: Are Workers Paid their Marginal Product?" University of Pennsylvania working paper (2013).

Kerr, William R. "U.S. High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence." Harvard Business School Working Paper (2013): 14-17.

Kerr, William R., and William F. Lincoln. "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention." *Journal of Labor Economics* 28.3 (2010): 473-508.

Lowell, B. Lindsay. “H-1B Temporary Workers: Estimating the Population.” UCSD Center for Comparative Immigration Studies Working Paper No. 12 (2000).

Norman, Maltoff. "On the Need for Reform of the H-1B Non-immigrant Work Visa in Computer-Related Occupations." *University of Michigan Journal of Law Reform* 36.4 (2003): 815-914.

North, David. "Estimating the Size of the H-1B Population in the U.S." *Center for Immigration Studies Memorandum* (2011).

Pekkala Kerr, Sari, and William R. Kerr. "Economic Impacts of Immigration: A Survey." *Finnish Economic Papers* 24.1 (2011): 1-32.

Pekkala Kerr, Sari, William R. Kerr, and William F. Lincoln. "Skilled Immigration and the Employment Structures of U.S. Firms." *Journal of Labor Economics* (forthcoming).

Pence, Karen. "The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving." *The B.E. Journal of Economic Analysis & Policy* 5.1 (2006): 1-24.

Peri, Giovanni, Kevin Shih, and Chad Sparber. "STEM Workers, H-1B Visas, and Productivity in US Cities." *Journal of Labor Economics* (forthcoming).

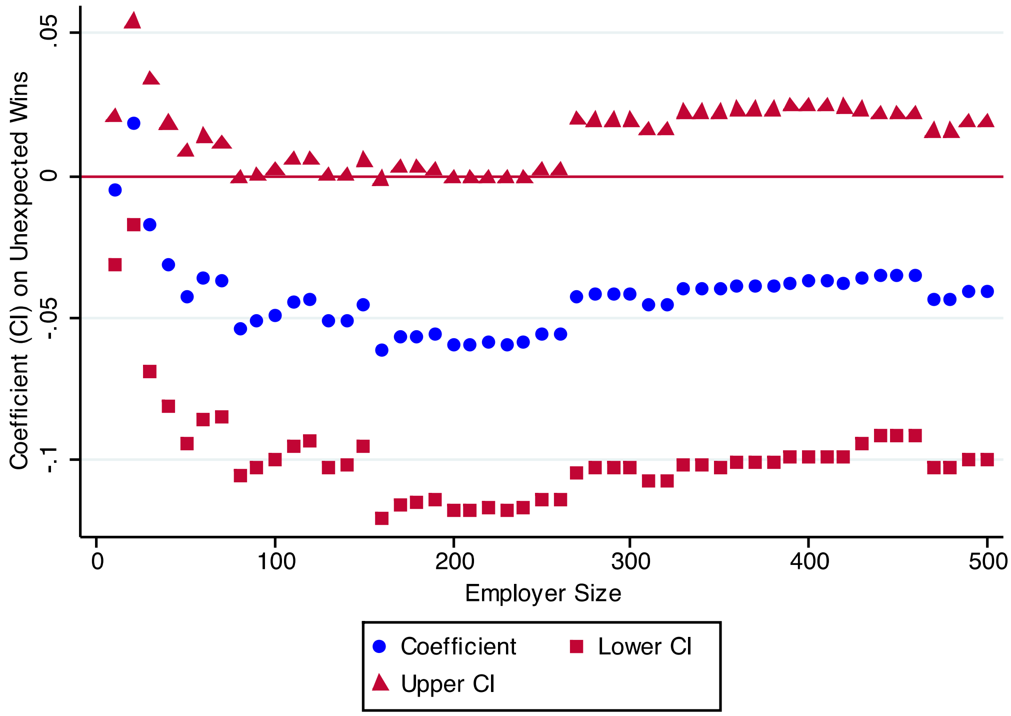
Peri, Giovanni, Kevin Shih, and Chad Sparber. "The Effects of Foreign Skilled Workers on Natives: Evidence from the H-1B Visa Lottery." UC Davis Working Paper (2014).

U.S. Customs and Immigration Services. “Change in H-1B Procedures Trims Weeks Off Final Selection Process.” Web. http://www.uscis.gov/sites/default/files/files/pressrelease/H1Bfy08CapUpdate041907.pdf (accessed October 7, 2014).

U.S. Department of Labor. Office of Foreign Labor Certification Data. Web. http://www.foreignlaborcert.doleta.gov/performancedata.cfm#stat (accessed October 7, 2014).

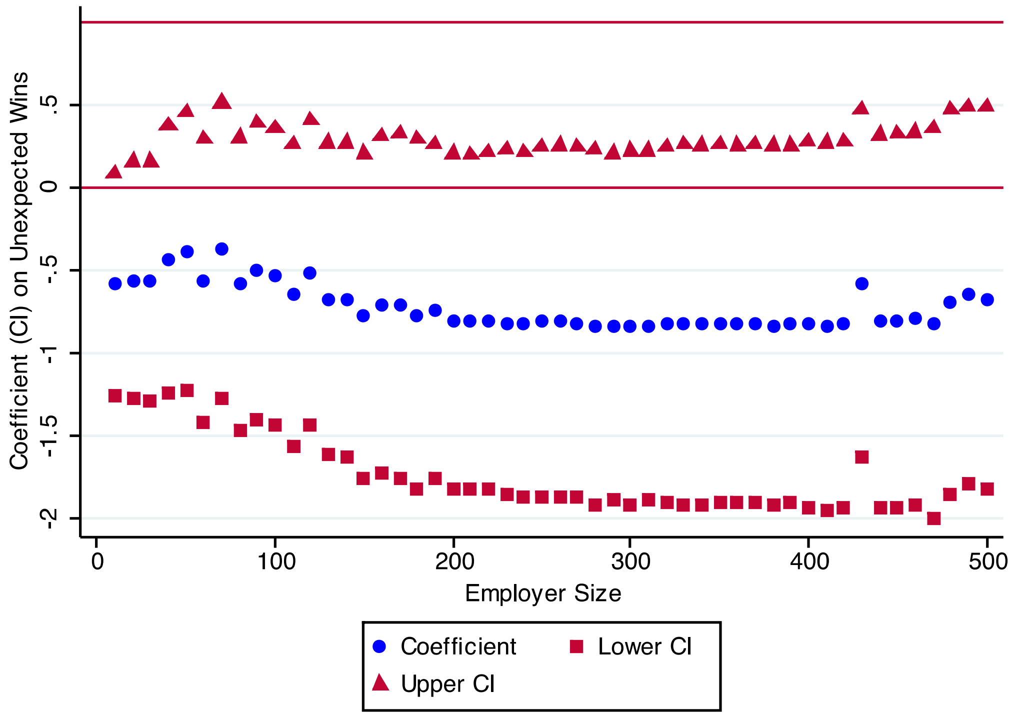
U.S. Patent and Trade Office. “Performance and Accountability Report: Fiscal Year 2012.” Washington, D.C.: U.S. Government Printing Office.

Figure 1. *Effect of H-1B Visas on Patents, by Employer Size*



Notes: The figure shows the coefficient and 95 percent confidence interval on approved H-1B visas when the dependent variable is the inverse hyperbolic sine of number of patents among employers of the indicated sizes or smaller in Year -1 (where employer size is shown on the x-axis). We show the coefficient for employers of each size range from 0-10 to 0-500, with the upper bound of the size range in increments of 10. We use the baseline specification, in which we control for lagged number of patents and expected lottery wins. After multiplying by 100, the coefficient should be interpreted as the approximate percentage increase in total firm employment associated with an unexpected H-1B visa lottery win.

Figure 2. *Effect of H-1B Visas on Employment, by Employer Size*



Notes: The figure shows the coefficient and 95 percent confidence interval on unexpected lottery wins from median regressions when the dependent variable is the number of employees in a firm in Quarters 1-4 of the first fiscal year an employee can work at the firm, among employers of the indicated size or smaller in Year -1 (where employer size is shown on the x-axis). We show the coefficient for employers of each size range from 0-10 to 0-500, with the upper bound of the size range in increments of 10. We use the specification in which we control for lagged employment and expected lottery wins.

Table 1. *Summary Statistics*

Variable	Mean (SD)	N
Fraction Patenting (all)	0.093	3,050
	(0.29)	
Fraction Patenting (≤30)	0.033	1,276
	(0.18)	
Fraction Patenting (≤10)	0.025	749
	(0.16)	
Number of Patents (all)	37.74	3,050
	(390.95)	
Number of Patents (≤30)	1.92	1,276
	(61.74)	
Number of Patents (≤10)	0.19	749
	(2.87)	
Inverse hyperbolic sine of patents (all)	0.33	3,050
	(1.28)	
Inverse hyperbolic sine of patents (≤30)	0.064	1,276
	(0.37)	
Inverse hyperbolic sine of patents (≤10)	0.048	749
	(0.34)	
Number of employees in Q1-Q4 (all)	1,877.84	9,803
	(39,721.31)	
Number of employees in Q1-Q4 (≤30)	43.09	4,909
	(1,904.34)	
Number of employees in Q1-Q4 (≤10)	9.64	2,862
	(55.63)	
Median employees in Q1-Q4 (all)	31	9,803
Median employees in Q1-Q4 (≤30)	10	4,909
Median employees in Q1-Q4 (≤10)	6	2,862
Winsorized emp. first difference in Q1-Q4 (all)	27.28	9,803
	(92.39)	
Winsorized emp. first difference in Q1-Q4 (≤30)	4.35	4,909
	(9.43)	
Winsorized emp. first difference in Q1-Q4 (≤10)	3.22	2,862
	(6.84)	
Fraction winning lottery		
2006 Regular	0.038	2,687
2006 ADE	0.17	306
2007 Regular	0.98	3,954
2007 ADE	0.55	296
Unexpected lottery wins	0.00	3,050
	(0.33)	

Notes: The source of the data is IRS and USCIS administrative data, and the Patent Network Dataverse. "All" refers to the full sample of firms that enter the lottery; "≤30" refers to those firms that have 30 or fewer employees in Year -1; "≤10" refers to those firms that have 10 or fewer employees in Year -1. Employment data are observed in Q1-Q4, the first four quarters when the H-1B worker may work at the firm (which are the same four quarters we investigate in our employment main results in Table 5). The number of patents refers to approved patents from the year of the lottery (2006 or 2007) and the subsequent six years. "N" refers to the number of observations in the sample. When we aggregate across Q1-Q4, N's refer to the number of observations; the number of firms is smaller, because we observe most firms several times over Q1-Q4 combined.

Table 2. *Validity of Randomized Design. OLS Regressions of Placebo Outcomes on Unexpected H-1B Lottery Wins*

Dependent Variable	Coefficient (SE) on Unexpected Wins
Lottery data has firm information	0.0028
	(0.0032)
Whether match to tax master file	0.0080
	(0.0079)
Whether match to quarterly employment data	-0.0031
	(0.0096)
Patents from Year -3 to Year -1 (all)	12.27
	(11.89)
Patents from Year -3 to Year -1 (≤30)	-0.30
	(0.28)
Patents from Year -3 to Year -1 (≤10)	-0.014
	(0.025)
Inverse hyperbolic sine of patents from Year -3 to Year -1 (all)	0.079
	(0.060)
Inverse hyperbolic sine of patents from Year -3 to Year -1 (≤30)	-0.036
	(0.025)
Inverse hyperbolic sine of patents from Year -3 to Year -1 (≤10)	-0.012
	(0.0087)
Patented from Year -3 to Year -1 (all)	-0.0039
	(0.021)
Patented from Year -3 to Year -1 (≤30)	-0.026
	(0.019)
Patented from Year -3 to Year -1 (≤10)	-0.0032
	(0.0097)
Employment in Year -2 (all, quantile)	0.56
	(0.62)
Employment in Year -2 (≤30, quantile)	-0.55
	(0.45)
Employment in Year -2 (≤10, quantile)	-0.31
	(0.44)
Employment in Year -2 (all, winsorized)	0.082
	(9.71)
Employment in Year -2 (≤30, winsorized)	0.56
	(0.89)
Employment in Year -2 (≤10, winsorized)	-0.091
	(0.57)
Employment in Year -1 (all, quantile)	2.91
	(4.41)
Employment in Year -1 (all, winsorized)	30.35
	(104.55)
Dummy for NAICS=54 (all)	0.007
	(0.03)
Dummy for NAICS=54 (≤30)	-0.033
	(0.043)
Dummy for NAICS=54 (≤10)	0.010
	(0.058)

Notes: The table illustrates the validity of the randomized design. In the specifications in which employment in Year -2 is the dependent variable, we control for employment in Year -1. In the specifications in which employment in Year -1 is the dependent variable, we have no controls (as we clearly cannot control for Year -1 employment in this context), and we only investigate the results in the "All" sample because selecting this sample based on Year -1 employment could lead to biased and inconsistent results. Standard errors are clustered by firm. *** refers to significance at the 1% level; ** at the 5% level, and * at the 10% level.

Table 3. *Two-Stage Least Squares Regressions of Patent Outcomes on Approved H-1B Visas, Using Unexpected H-1B Lottery Wins as the Instrument: Coefficient (CI) on Unexpected Wins*

	Inverse hyp. sine of # patents		# Patents		Patenting Dummy	
Panel A: ≤10 employees						
A) Years 0 to 6	-0.0030	-0.0052	0.0027	0.0032	-0.011	-0.016
	[-0.021, 0.015]	[-0.031, 0.021]	[-0.035, 0.040]	[-0.051, 0.057]	[-0.029, 0.0084]	[-0.042, 0.010]
B) Years 0 to 2	-0.0019	-0.0037	0.0033	0.0037	-0.010	-0.016
	[-0.017, 0.013]	[-0.026, 0.019]	[-0.025, 0.032]	[-0.038, 0.045]	[-0.029, 0.0086]	[-0.042, 0.011]
C) Years 3 to 6	0.0026	0.0038	-0.00062	-0.00058	0.000021	-0.00037
	[-0.0070, 0.012]	[-0.0097, 0.017]	[-0.014, 0.012]	[-0.019, 0.018]	[-0.0070, 0.0070]	[-0.011, 0.010]
Panel B: ≤30 employees						
D) Years 0 to 6	-0.015	-0.0017	-0.19	-0.22	0.0085	0.010
	[-0.060, 0.030]	[-0.068, 0.034]	[-0.52, 0.14]	[-0.57, 0.13]	[-0.020, 0.037]	[-0.022, 0.042]
E) Years 0 to 2	-0.012	-0.014	-0.14	-0.17	0.0086	0.010
	[-0.059, 0.035]	[-0.067, 0.040]	[-0.38, 0.09]	[-0.42, 0.086]	[-0.026, 0.043]	[-0.029, 0.049]
F) Years 3 to 6	-0.0079	-0.0091	-0.048	-0.056	0.0012	0.0014
	[-0.037, 0.021]	[-0.042, 0.024]	[-0.18, 0.083]	[-0.20, 0.090]	[-0.018, 0.021]	[-0.021, 0.024]
Panel C: All						
G) Years 0 to 6	-0.16	-0.059	26.51	10.10	-0.013	-0.0050
	[-0.87, 0.55]	[-0.13, 0.013]	[-100.78, 153.79]	[-15.10, 35.30]	[-0.11, 0.083]	[-0.035, 0.025]
H) Years 0 to 2	-0.10	-0.037	21.74	8.57	-0.015	-0.0055
	[-0.58, 0.37]	[-0.095, 0.021]	[-72.50, 116.00]	[-5.37, 22.50]	[-0.12, 0.089]	[-0.035, 0.025]
I) Years 3 to 6	-0.11	-0.040	4.76	1.54	-0.027	-0.010
	[-0.60, 0.39]	[-0.10, 0.024]	[-34.62, 44.13]	[-10.98, 14.05]	[-0.17, 0.11]	[-0.039, 0.018]
Prior patents	X	X	X	X	X	X
Prior patents, E[wins]		X		X		X

Notes: The table shows the effect of an extra H-1B visa on patent outcomes over the indicated years. The table shows coefficients and 95% confidence intervals on approved H-1B visas. The "prior patents" specifications control for the total number of patents from 2000 to Year -1. The "prior patents, E[wins]" specifications control for patents in the pre-period and expected lottery wins (equal to number of H-1B applications considered in a lottery multiplied by the probability of winning the lottery). See Table 1 for additional notes and sample sizes. Standard errors are clustered by firm. *** refers to significance at the 1% level; ** at the 5% level, and * at the 10% level.

Table 4. *Two-Stage Least Squares Regressions of Inverse Hyperbolic Sine of Patents on Approved H-1B Visas, Using Unexpected H-1B Lottery Wins as the Instrument*

Outcome	(1) Regular	(2) ADE	(3) Professional, scientific, and technical services	(4) Other industries
Panel A: ≤10 employees				
	0.011	-0.077	-0.014	0.020
	[-0.0057, 0.028]	[-0.19, 0.035]	[-0.049, 0.020]	[-0.0079, 0.048]
N	681	68	484	265
Panel B: ≤30 employees				
	-0.012	-0.031	-0.025	0.013
	[-0.074, 0.050]	[-0.12, 0.059]	[-0.089, 0.039]	[-0.039, 0.065]
N	1,136	140	837	439
Panel C: All				
	-0.022	-0.10	-0.075	-0.016
	[-0.076, 0.032]	[-0.29, 0.87]	[-0.16, 0.011]	[-0.14, 0.11]
N	2,540	510	1,721	1,329

Notes: The table shows the effect of an extra H-1B visa on the inverse hyperbolic sine of the number of patents from Years 0 to 6 (inclusive). The table shows coefficients and 95% confidence intervals on approved H-1B visas. All specifications control for patents in the pre-period and expected lottery wins, as in the baseline. The results are comparable when we investigate the patenting dummy or the number of patents as the dependent variable. See Tables 1 and 3 for additional notes and sample sizes. Standard errors are clustered by firm. *** refers to significance at the 1% level; ** at the 5% level, and * at the 10% level.

Table 5. *Effect of H-1B Visa on Employment Outcomes: Median and Two-Stage Least Squares Regressions*

	Median Regressions		Two-stage least squares	
Panel A: ≤10 employees				
A) Q1 to Q4 (n=2,862)	-0.53	-0.52	-0.54	-1.10
	[-1.18, 0.12]***	[-1.15, 0.11]***	[-1.95, 0.88]**	[-2.88, 0.68]**
B) Q1 (n=679)	-0.00	-0.031	0.072	-0.15
	[-1.28, 1.28]	[-1.64, 1.58]	[-1.24, 1.39]	[-2.15, 1.86]
C) Q2 (n=696)	-0.00	-0.41	-0.80	-1.46
	[-0.68, 0.68]***	[-1.17, 0.36]***	[-2.34, 0.75]**	[-3.29, 0.36]***
D) Q3 (n=689)	-0.78	-0.53	-0.66	-1.33
	[-1.78, 0.23]***	[-1.42, 0.36]***	[-2.40, 1.08]*	[-3.47, 0.80]**
E) Q4 (n=684)	-0.76	-0.61	-0.90	-1.72
	[-2.05, 0.51]***	[-1.79, 0.57]***	[-3.12, 1.31]*	[-4.52, 1.08]*
Panel B: ≤30 employees				
F) Q1 to Q4 (n=4,909)	-0.44	-0.36	-0.97	-1.26
	[-1.16, 0.28]***	[-1.09, 0.37]***	[-2.96, 1.01]*	[-3.25, 0.71]**
G) Q1 (n=1,121)	-0.35	-0.32	-1.05	-1.31
	[-1.41, 0.72]***	[-1.38, 0.73]**	[-3.17, 1.06]*	[-3.47, 0.85]**
H) Q2 (n=1,163)	-0.22	-0.17	-0.73	-0.95
	[-1.08, 0.65]***	[-1.11, 0.78]**	[-2.57, 1.10]*	[-2.90, 1.00]*
I) Q3 (n=1,154)	-0.95	-0.76	-1.00	-1.33
	[-2.17, 0.27]***	[-1.83, 0.31]***	[-3.23, 1.23]*	[-3.62, 0.96]**
J) Q4 (n=1,147)	-0.53	-0.53	-0.92	-1.25
	[-1.82, 0.76]***	[-1.85, 0.79]**	[-3.51, 1.67]	[-3.99, 1.49]
Panel C: All				
L) Q1 to Q4 (n=9,803)	-1.27	-1.05	-20.37	-2.41
	[-3.08, 0.55]***	[-2.67, 0.57]**	[-230.99, 190.24]	[-17.76, 12.94]
M) Q1 (n=2,131)	-1.41	-1.67	-62.10	-9.40
	[-3.40, 0.58]***	[-3.89, 0.54]**	[-768.40, 644.19]	[-22.73, 3.92]
O) Q2 (n=2,240)	-1.35	-1.00	-17.32	-2.75
	[-3.72, 1.02]*	[-3.11, 1.12]*	[-180.09, 145.44]	[-18.09, 12.58]
P) Q3 (n=2,226)	-0.055	0.25	4.76	4.43
	[-3.15, 3.03]	[-2.33, 2.83]	[-72.71, 82.24]	[-15.97, 24.83]
Q) Q4 (n=2,219)	1.36	-0.31	-13.70	0.04
	[-4.80, 2.07]	[-3.64, 3.01]	[-191.01, 163.60]	[-21.57, 21.64]
Prior employment	X	X	X	X
Prior employment, E[wins]		X		X

Notes: The table shows point estimates and 95% confidence intervals. The first two columns show median regressions of employment on unexpected lottery wins. The next two columns show two-stage least squares regressions where the dependent variable, the difference of employment from the first quarter of Year -1 to the quarter in question, has been winsorized at the 95th percentile. The 5th and 95th percentiles of the first difference in employment are -109 and 352, respectively, in the full sample; are -9 and 30, respectively, among those with 30 or fewer employees; and are -6 and 22, respectively, among those with 10 or fewer. In these regressions, the instrument is unexpected lottery wins and the endogenous variable is approved H-1B visas. The "prior employment" specifications control for employment from the first quarter of Year -1, and the "prior employment, E[wins]" specifications additionally control for the number of expected lottery wins. None of the estimates is significantly different from 0 at any conventional significance level. "n" refers to the total number of observations. See Tables 1 and 3 for other notes. *** denotes estimates that are significantly different *from 1* at the 1% level; ** at the 5% level; * at the 10% level.

Table 6. *Median Regressions of Employment on Unexpected Lottery Wins*

Outcome	(1) Regular	(2) ADE	(3) Professional, scientific, and technical services	(4) Other industries
A) ≤10 employees	-0.41	-0.0000002	-0.58	0.36
	[-1.10, 0.27]***	[-1.36, 1.36]	[-1.54, 0.39]***	[-0.50, 1.22]
	{n=2,635}	{n=227}	{n=1,889}	{n=973}
B) ≤30 employees	-0.59	0.52	-0.72	0.65
	[-1.46, 0.28]***	[-1.51, 2.55]	[-1.92, 0.48]***	[-0.36, 1.65]
	{n=4,431}	{n=478}	{n=3,269}	{n=1,640}
C) All	-1.26	1.38	-1.46	1.16
	[-3.33, 0.81]**	[-5.63, 8.39]	[-3.60, 0.67]**	[-2.74, 5.05]
	{n=8,349}	{n=1,454}	{n=5,767}	{n=4,036}

Notes: The table shows the effect of unexpected lottery wins on employment, displaying point estimates and 95% confidence intervals in square brackets for median regressions of employment on unexpected lottery wins. All specifications control for employment in the pre-period and expected lottery wins, as in the baseline. The number of observations included in each regression appears in curly brackets below the confidence interval. See Tables 4 and 5 for additional notes. *** denotes estimates that are significantly different *from 1* at the 1% level; ** at the 5% level; * at the 10% level. None of the estimates is significantly different from zero at any conventional significance level.

Appendix 1. Description of matching procedure

As described in the main text, we performed an intentionally liberal automatic matching procedure between these datasets in order to obtain all plausible matches between companies and patents. We then searched through the matches by hand in order to detect and remove all matches that appeared spurious.

The automatic matching procedure proceeded as follows. First, we assigned clearly related firm names to single categories (i.e., "Sony", "Sony Co.", "Sony Corporation", etc). Then we searched for complete string matches between the name categories in the USPTO data and the name categories in the USCIS H1-B visa lottery data, and we classified these as matches between the datasets. After all such matches were made, we then searched for complete string matches between these two sets of name categories with all spaces in the names removed and also classified these as matches. Finally, we performed a "fuzzy" match between USPTO and USCIS firm names. The fuzzy matching procedure calculated a "distance" between words in each list by determining how many characters in the words need to be edited in order to transform a word from one list into a word in the other. This is necessary to identify all matches because, for example, firm names are occasionally misspelled. Pairs of words in firm name categories were classified as non-matching if the number of characters that differed between the words was more than one for words with six or fewer characters, or when the number of characters that differed between the words was more than two for words with seven or more characters (using the word as spelled in the USCIS data to determine the number of characters in the word). Otherwise, this pair of words was classified as a possible match. If at least 75% of the pairs of words in the firm name were possible matches, then the entire firm name was classified as a possible match.

We intentionally designed this "liberal" procedure so that it is liable to classify many non-matches as matches (but not the reverse); thus, if a firm did not match at all between the two datasets according to the fuzzy match, we can be quite certain that it was not granted any US patents between 1975 and 2013. This matching procedure identified all *potential* matches between the two datasets (that satisfy the match criteria described above), *i.e.* the procedure did not only find the single best match but also found other matches. The goal of this automatic matching procedure was to generate a list of *all* potential matches, which we could then winnow by hand in the next step.

Once this automatic matching procedure was complete, all of the resulting matches were checked by hand to determine whether they appeared to be a possible match. Of the 668 companies in the USCIS lottery list which obtained at least one automatic match, we identified 208 cases in which all of that company's matches were clearly incorrect through by-hand inspection. We further identified 392 cases in which all of that company's matches were clearly correct (legitimate variations on the correct company name) through by-hand inspection. Finally, we identified 63 cases in which the matches were ambiguous; in our judgment the match is possibly correct, but we cannot be fully confident that it is correct. We assume that both unmatched companies and those that received clearly incorrect matches did not patent at all between 1975 and 2013. In the results that we report below, we exclude the 63 possible matches from the list of matched companies. In the Appendix, we show that the results are robust to assuming that the possible matches were in fact matches. The results are also robust to alternative assumptions and similar alternative matching procedures.

Appendix Table 1. *First stage regressions*

Sample	Coefficient (SE) on Unexpected Lottery Wins	First-stage F-statistic
All	0.87 (0.03)***	993.51
≤30	0.88 (0.04)***	420.25
≤10	0.86 (0.06)***	239.94

The table shows the first stage regression of the number of approved H-1Bs on the number of unexpected wins. We show the first stage regression for the baseline specification (controlling for the lagged dependent variable and the expected number of H-1B lottery wins); the first stage in other specifications is extremely similar. See other notes to Tables 1 and 3. *** denotes $p<0.01$; ** denotes $p<0.05$; * denotes $p<0.10$.

Appendix Table 2. *Two-Stage Least Squares Regressions of Patent Outcomes on Approved H-1B Visas, Using Unexpected H-1B Lottery Wins as the Instrument: Coefficient (CI) on Unexpected Wins*

	Inverse hyp. sine of # patents		# Patents		Patenting Dummy	
Panel A: ≤10 employees						
A) Years 0 to 6	-0.0017	-0.0031	0.038	0.055	-0.012	-0.018
	[-0.020, 0.016]	[-0.029, 0.023]	[-0.042, 0.12]	[-0.054, 0.16]	[-0.032, 0.0078]	[-0.045, 0.0089]
B) Years 0 to 2	0.00042	-0.000054	0.038	0.054	-0.012	-0.018
	[-0.016, 0.016]	[-0.024, 0.024]	[-0.038, 0.11]	[-0.049, 0.16]	[-0.032, 0.0076]	[-0.045, 0.0084]
C) Years 3 to 6	0.0033	0.0049	0.000043	0.00058	0.000093	-0.00018
	[-0.0066, 0.013]	[-0.0088, 0.019]	[-0.013, 0.013]	[-0.018, 0.019]	[-0.0073, 0.0075]	[-0.011, 0.011]
Panel B: ≤30 employees						
D) Years 0 to 6	-0.018	-0.020	-0.14	-0.16	0.0064	0.0077
	[-0.064, 0.028]	[-0.072, 0.031]	[-0.48, 0.21]	[-0.54, 0.22]	[-0.022, 0.035]	[-0.025, 0.041]
E) Years 0 to 2	-0.014	-0.016	-0.12	-0.14	0.0059	0.0070
	[-0.062, 0.034]	[-0.070, 0.039]	[-0.36, 0.12]	[-0.40, 0.13]	[-0.028, 0.040]	[-0.032, 0.046]
F) Years 3 to 6	-0.0087	-0.0098	-0.017	-0.020	-0.00026	-0.00018
	[-0.038, 0.021]	[-0.043, 0.024]	[-0.16, 0.13]	[-0.18, 0.14]	[-0.020, 0.019]	[-0.023, 0.022]
Panel C: All						
G) Years 0 to 6	-0.23	-0.087	8.98	3.15	-0.023	-0.0094
	[-1.21, 0.75]	[-0.17, 0.00065]*	[-74.12, 92.08]	[-24.56, 30.86]	[-0.15, 0.10]	[-0.040, 0.021]
H) Years 0 to 2	-0.16	-0.060	16.01	6.30	-0.030	-0.012
	[-0.86, 0.54]	[-0.13, 0.014]	[-58.74, 90.75]	[-7.95, 20.54]	[-0.18, 0.12]	[-0.042, 0.019]
I) Years 3 to 6	-0.17	-0.066	-7.03	-3.15	-0.039	-0.015
	[-0.92, 0.57]	[-0.14, 0.0089]*	[-51.79, 37.73]	[-18.20, 11.91]	[-0.22, 0.14]	[-0.045, 0.015]
Prior patents	X	X	X	X	X	X
Prior patents, E[wins]		X		X		X

Notes: See notes to Table 3. The table is identical to Table 3, except that in defining which firms match between the USCIS data and the Patent Dataverse, Appendix Table 2 includes those firms that are "possible" matches (whereas Table 2 excludes those firms). *** refers to significance at the 1% level; ** at the 5% level, and * at the 10% level.

Appendix Table 3. *Additional employment specifications*

	(1) Level, winsorized at 99%	(2) Inverse hyperbolic sine	(3) Inverse hyperbolic sine of difference, winsorized at 99%	(4) Inverse hyperbolic sine of level, winsorized at 99%	(5) First difference of employment, no controls
A) ≤10 employees	-1.86	-0.18	-0.18	-0.18	-0.53
	[-4.34, 0.62]**	[-0.43, 0.066]**	[-0.43, 0.067]**	[-0.42, 0.068]**	[-1.37, 0.31]***
B) ≤30 employees	-1.69	-0.16	-0.15	-0.16	-0.69
	[-4.55, 1.17]*	[-0.35, 0.035]*	[-0.34, 0.034]**	[-0.35, 0.037]**	[-1.68, 0.31]***
C) All	1.06	0.034	0.045	0.032	-1.07
	[-73.91, 76.03]	[-0.15, 0.22]	[-0.14, 0.23]	[-0.14, 0.21]	[-3.05, 0.92]**

Notes: Columns 1-4 of the table show two-stage least squares regressions of employment outcomes on approved H-1B visas, where unexpected lottery wins are the instrument for approved H-1B visas. In Column 1, the dependent variable is the difference of employment from the first quarter of Year -1 to pooled Q1, Q2, Q3, and Q4 employment, and winsorized at the 99th percentile. The 1st and 99th percentiles of the first difference in employment are -5,559 and 2,430, respectively, in the full sample; are -20 and 62, respectively, among those with 30 or fewer employees; and are -10 and 53, respectively, among those with 10 or fewer. In Column 2, the dependent variable is the inverse hyperbolic sine of the difference in employment over the same periods. In Column 3, the dependent variable is the inverse hyperbolic sine of the difference in employment over the same periods, winsorized at the 99th percentile. In Column 4, the dependent variable is the inverse hyperbolic sine of the level of employment in Q1-Q4 pooled, winsorized at the 99th percentile, and the results are nearly identical to those in Column 3. All specifications in Columns 1, 2, 3, and 4 control for prior employment and the number of expected lottery wins, as in the baseline; the results are similar with other controls. In Column 5, we run median regressions (as in Table 5) and the dependent variable is the first difference of employment (from the first quarter of calendar Year -1 to a given quarter of Year 0, and pooling this measure from Q1 to Q4), but we do not include any controls. The results of these regressions are also very similar when we include controls, such as controlling for expected wins. In all columns, we pool across Q1-Q4, as in the baseline; the results are comparable (though typically slightly less precise) when we examine each quarter separately. None of the estimates is significantly different from 0 at any conventional significance level. In the case of these inverse hyperbolic sine specifications, before testing whether a coefficient is equal to 1, we transform the coefficient from the regression (which reflects the percentage increase in employment, rather than the increase in the absolute level of employment) by multiplying it by the mean level of employment. We then test whether this transformed coefficient is equal to 1. The test results that are reported above refer to this test. *** denotes estimates that are significantly different *from 1* at the 1% level; ** at the 5% level; * at the 10% level. See Table 5 for other notes.

Appendix Table 4. *Effect of H-1B Visa on Being out of Business: Two-Stage Least Squares Regressions*

Panel A: ≤10 employees		
A) Q1 to Q4 (n=2,963)	0.024	0.033
	[-0.016, 0.063]	[-0.022, 0.088]
B) Q1 (n=716)	0.016	0.023
	[-0.020, 0.052]	[-0.030, 0.077]
C) Q2 (n=749)	0.017	0.022
	[-0.033, 0.066]	[-0.051, 0.095]
D) Q3 (n=749)	0.032	0.046
	[-0.014, 0.079]	[-0.015, 0.11]
E) Q4 (n=749)	0.029	0.041
	[-0.017, 0.076]	[-0.022, 0.10]
Panel B: ≤30 employees		
F) Q1 to Q4 (n=5,309)	0.010	0.012
	[-0.019, 0.040]	[-0.024, 0.047]
G) Q1 (n=1,211)	0.0033	0.0033
	[-0.028, 0.034]	[-0.034, 0.40]
H) Q2 (n=1,276)	0.0030	0.0029
	[-0.035, 0.041]	[-0.043, 0.049]
I) Q3 (n=1,276)	0.015	0.017
	[-0.020, 0.050]	[-0.023, 0.058]
J) Q4 (n=1,276)	0.020	0.023
	[-0.013, 0.052]	[-0.014, 0.060]
Panel C: All		
L) Q1 to Q4 (n=10,022)	0.0050	0.0024
	[-0.068, 0.078]	[-0.014, 0.019]
M) Q1 (n=2,378)	-0.032	-0.0053
	[-0.39, 0.32]	[-0.022, 0.011]
O) Q2 (n=2,548)	-0.013	-0.0024
	[-0.13, 0.11]	[-0.024, 0.019]
P) Q3 (n=2,548)	-0.015	0.0054
	[-0.10, 0.13]	[-0.014, 0.025]
Q) Q4 (n=2,548)	0.037	-0.011
	[-0.21, 0.28]	[-0.0084, 0.031]
Prior employment	X	X
Prior employment, E[wins]		X

Notes: The table shows point estimates and 95% confidence intervals from two-stage least squares (linear probability) regressions where the dependent variable is a dummy for whether the firm is "out of business." We define a firm as being "out of business" if it has either zero employees or is missing the number of employees, *and* it also has either zero total payroll or is missing total payroll. The results are similar with other definitions of being out of business. In these regressions, the instrument is unexpected lottery wins and the endogenous variable is approved H-1B visas. The "prior employment" specifications control for employment from the first quarter of Year -1, and the "prior employment, E[wins]" specifications additionally control for the number of expected lottery wins. None of the estimates is significantly different from 0 at any conventional significance level. "n" refers to the total number of observations. Note that for a given firm size threshold, the number of observations is the same in Q2, Q3, and Q4, but it is different in Q1; this is because we drop data from the FY2006 ADE lottery in Q1 because this lottery occurred after Q1 ended, as explained in the main text. See Tables 1, 3, and 5 for other notes. *** denotes estimates that are significantly different *from 1* at the 1% level; ** at the 5% level; * at the 10% level.