I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13892

Buying Lottery Tickets for Foreign Workers: Search Cost Externalities Induced by H-1B Policy

Rishi Sharma
Chad Sparber

NOVEMBER 2020

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13892

Buying Lottery Tickets for Foreign Workers: Search Cost Externalities Induced by H-1B Policy

Rishi Sharma
Colgate University

Chad Sparber
Colgate University and IZA

NOVEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Buying Lottery Tickets for Foreign Workers: Search Cost Externalities Induced by H-1B Policy*

The H-1B program allows firms in the United States to temporarily hire high-skilled foreign citizens. H-1B workers are highly concentrated among a small number of firms. We develop a theoretical model demonstrating that this phenomenon is an artifact of policy design: When the government restricts foreign labor inflows and allocates H- 1B status by random lottery, it creates a negative externality by incentivizing firms to search for more workers than can actually be hired. Some firms rationally move toward specializing in hiring foreign labor and contracting out those workers' services to third- party sites. This outsourcing behavior further exacerbates total search costs and lottery externalities, resulting in an annual economic loss in the hundreds of millions of dollars.

JEL Classification: J61, J68, F22
Keywords: skilled workers, H-1B, outsourcing

Corresponding author:
Chad Sparber
Department of Economics
Colgate University
12 Oak Drive
Hamilton, NY 13346
USA
E-mail: csparber@colgate.edu

* All views contained herein are the authors' own.

1 Introduction

The H-1B program provides an important method for college-educated foreign citizens to secure temporary legal employment in the United States. Though available for workers in a number of occupations, the program is particularly popular among Science, Technology, Engineering, and Mathematics (STEM) workers. For example, United States Citizenship and Immigration Services (USCIS, 2020a) reports that computer-related workers alone accounted for 66.1% of H-1Bs issued in fiscal year (FY) 2019. Many economists credit the program (or foreign STEM workers more generally) for generating a substantial share of recent U.S. innovation, technology, productivity, wage, and GDP growth.[1] Such evidence led the National Academies of Sciences, Engineering, and Medicine (2017) to argue in its summary on the economic effects of immigration that "the prospects for long-run economic growth in the United States would be considerably dimmed without the contributions of high-skilled immigrants."

Less sanguine views of skilled immigration and the H-1B program generally focus on two concerns. The first is the possibility that infusions of skilled workers might reduce wages paid to subsets of native-born labor.[2] The second – as articulated by Costa and Hira (2020) – is that "Companies with an outsourcing business model rely on the H-1B program to build and expand a business model based on outsourcing jobs. In this arrangement, rather than being employed directly by the company that hired them, the H-1B workers ultimately work for third-party clients, either on- or off-site." The authors note that just 30 firms accounted for 27% of H-1B issuances in fiscal year 2019, half of which used an outsourcing model.

Our paper develops a theoretical model arguing that the US government's response to this first concern has exacerbated costs associated with the second.

In an effort to reduce potential labor market competition with skilled immigrants, the US limits the number of new H-1B entrants (at most firms) to just 85,000 per year. This cap (or quota) is far exceeded by interest in the H-1B program. At its peak to date, USCIS received more than 236,000 petitions for new H-1B status in the first week of application eligibility for fiscal year 2017. The government has responded to this excess demand by allocating new H-1B status through a random lottery. Our paper's theoretical model demonstrates that this has generated two presumably unintended costs. First, it has incentivized firms to search for (and extend offers to) far more workers than can actually be hired through the program.

[1] See Hunt and Gauthier-Loiselle (2010), Kerr and Lincoln (2010), Hunt (2011), Moser et al. (2014), Kerr, Kerr, and Lincoln (2015), Peri, Shih, and Sparber (2015), Bernstein et al. (2018), and Gunadi (2019).

[2] See Borjas and Doran (2012), Bound et al. (2015), Ma (2020), and Turner (2020). This is related to a worry that strong employer/employee links within the H-1B program might inhibit labor mobility and therefore provide a recipe for labor exploitation. Depew et al. (2017) and Hunt and Xie (2019), however, argue that H-1B workers are quite mobile, thus reducing the possibility for exploitation of this type.

Any offer extended by one firm creates a negative externality on competing firms by reducing the chances that they will win the lottery and secure permission to hire the individuals they would like to employ. Second, some firms have rationally responded to the quota and lottery by specializing in outsourcing services that contract out labor to other firms. This further increases the number of petitions for H-1B status and, therefore, the size of the negative job search externality. Altogether, this externality costs US firms hundreds of millions of dollars per year, perhaps exceeding $1 billion during the late 2010s when firms submitted an extraordinarily high number of H-1B petitions on behalf of prospective employees.

Our analysis complements three recent empirical contributions to the literature. First, Mayda et al. (2018) provide empirical evidence showing that US government restrictions on the number of new H-1Bs available to prospective workers each year caused H-1B employment to become increasingly concentrated among a small group of employers. After documenting similar trends, we build a theory that illustrates that this behavior is indeed a result of H-1B policy design. Second, Glennon (2020) finds that firms respond to H-1B restrictions by moving production to foreign affiliates. Our theory instead models domestic outsourcing behavior such that firms specializing in hiring foreign workers then send their employees to third-party worksites within the country. Third, while past papers including Kerr et al. (2010), Peri et al. (2015a), Sparber (2018), and Mayda et al. (2020) have argued that H-1B restrictions and the H-1B lottery harm aggregate productivity, wages, the selection of foreign employment, and firm outcomes, our paper adds to these costs by recognizing that the lottery allocation mechanism also generates substantial losses associated with job search externalities.

Firms in our model generate output according to a standard Constant Elasticity of Supply (CES) production function and face monopolistically competitive demand for their products. A single key assumption drives our results: While most "regular" firms experience increasing marginal costs associated with labor searches, a select group of "outsourcing" firms experiences constant marginal costs, possibly because they have a large pool of potential H-1B workers already employed at overseas headquarters or affiliates.[3] An important implication is that the changes in the H-1B quota do not generate proportional changes in the probability of winning the lottery since the quota also affects the number of petitions that firms submit. That is, a 1% decrease in the cap reduces the probability of winning by more than 1%, implying that the elasticity of the winning probability with respect to the cap is greater than 1. We provide a closed form solution demonstrating that this elasticity is positively related to

[3]Note, for example, that in FY 2019 more than half of H-1Bs granted for initial employment, and nearly three quarters of H-1Bs overall, were allocated to workers born in India, while Indian-based companies are frequently cited as outsourcing firms.

the share of firm expenditures spent hiring H-1B workers in the economy and the premium paid to outsourced workers. We can also write it as a function of the win probability and the proportion of H-1B petitions filed in excess of actual H-1B demand. In any case, the higher this elasticity, the more the quota leads to increased H-1B concentration and higher search costs.

This paper provides a timely contribution to the economics of immigration literature. In February 2018, USCIS (2018) issued a policy memorandum that made it more difficult for employers to hire H-1B workers who they intended to place at third-party sites. This policy required "detailed statements from end-user clients, copies of client contracts, and work itineraries to demonstrate that non-speculative assignments were available throughout the entire requested period" (National Law Review 2020b). Perhaps as a consequence, H-1B denial rates rose from 13% in FY 2017 to 24% the following year. Increases in denial rates between FY 2015 and 2020 were much larger among Costa and Hira's (2020) outsourcing firms such as Cognizant (52 percentage points), Infosys (57 pp.), and Tech Mahindra (32 pp.) than at non-outsourcing firms using the H-1B program heavily such as Amazon (15 pp.), Google (13 pp.), Microsoft (13 pp.), and Facebook (8 pp.).[4] H-1B employers won a lawsuit challenging these rules in the US District Court for the District of Columbia in May 2020. As a result, USCIS rescinded the memorandum and issued less restrictive guidelines.[5] Given the volatility of rules governing H-1B workers at third-party sites, it is important to understand why such outsourcing has become prevalent, and what the economic consequences of the practice might be.

2 Stylized Facts about the H-1B Program

The H-1B program provides an important channel for college-educated foreign citizens to work in the United States. USCIS grants individuals H-1B status for up to a three year initial period that can then be renewed for a total of six years. The government has always imposed a statutory limit on the number of new H-1Bs issued per fiscal year. That limit equaled 195,000 from FY 2001-2003 but it was effectively non-binding since it was never reached. The cap dropped to 65,000 for FY 2004 and has held at 85,000 since FY 2006 when the government added 20,000 H-1Bs for people who have obtained a masters degree or more education from a US institution.[6] The quota has been binding throughout this period. Importantly, all employees of colleges, universities, and non-profit institutions are exempt

[4]See Anderson (2020) and National Foundation for American Policy (2020).
[5]See National Law Review (2020a) and USCIS (2020b).
[6]This was a result of the 2004 H-1B Visa Reform Act, but it did not become operational until FY 2006.

from the cap, as are workers applying to renew their H-1B status.

USCIS grants H-1B status to individuals, not firms, but strong employer/employee links in the program can lead to confusion on this point. A firm wanting to hire an H-1B worker begins by filing a Labor Condition Application (LCA) with the US Department of Labor attesting that it will abide by specific work conditions (e.g., that it will pay H-1B workers a wage at or above the wage earned by similar U.S. workers or the prevailing wage). This form also contains basic information about the job including the wage, date of intended employment, occupation title, and work location. There is no limit to the number of LCAs a firm can file, a single LCA can be used for multiple open job positions, and there is no LCA filing fee. The LCA serves as a *de facto* government permission slip for a firm to include foreign workers in its job search. It provides a crude measure of a firm's intent to hire foreign labor with the caveat that the number of job openings with LCA approval will far exceed the number of job offers that firms eventually extend to foreign workers.[7]

A firm can extend a job offer to a qualified foreign-born worker if it has obtained an approved LCA. It must then file an I-129 petition applying for H-1B status on the individual's behalf. Unlike LCAs, I-129 forms are linked with specific individuals. Thus, these forms not only include information about the employer, but the prospective employee as well. Also unlike LCAs, I-129 submissions require filing fees that begin at $460 and might include additional fraud detection fees that vary across firms. The Consolidated Appropriations Act of 2016 (Public Law 114-113) created an additional $4000 fee for firms employing 50 or more employees if more than 50% of those employees are on H-1B status.

In principle, USCIS grants H-1B status to individuals on a first-come first-served basis. Numerical caps pertain to fiscal years that begin on October 1, but USCIS accepts I-129 petitions beginning on April 1.[8] In practice, however, the H-1B program is vastly over-subscribed. The number of cap-subject I-129 petitions submitted during the first week of April for FYs 2008, 2009, and 2014 through today exceeded the total H-1B cap for the entire year. USCIS allocated all cap-subject new H-1Bs with a random lottery in those years. Figure 1 illustrates that USCIS received roughly 89,000 cap-subject petitions when H-1B limits exceeded the number of applications in FYs 2002 and 2003. In FYs 2008 and 2009, that number reached 123,000 and 163,000, respectively. The number peaked at over 236,000 for FY 2017, and hovered around 200,000 in FYs 2018-2020.

Unfortunately, Figure 1 provides a good measure of firms' actual demand for new H-1B workers only for FY 2002 and 2003. This is due to two data limitations. First – and most

[7]See Sparber (2019) for more discussion.

[8]Starting with fiscal year 2021, the Department of Homeland Security (DHS) requires firms filing cap-subject petitions to first electronically register with USCIS during March. USCIS (2020c) reports that it received nearly 275,000 unique registrations before April 1, 2020.

obvious – firms with an interest in hiring an H-1B worker after national limits have been exceeded will never petition USCIS for H-1B approval. USCIS cannot measure this demand. However, Mayda et al. (2018) estimate that in the absence of the reduced H-1B cap, demand would have been 16-26% higher than the 85,000 cap in FYs 2004-2007, and 22-33% higher in FYs 2008-2009. This would imply a maximum of new H-1B demand between 103,700 and 113,050 in each of these years, and yet firms applied for 163,000 new H-1B workers in FY 2009. This points to the second data limitation using I-129 petitions as a measure of H-1B demand. The theoretical model in Section 3 will show that petitions rise simply because firms compete against each other for a labor source that is constrained by artificial limits. That is, the H-1B cap creates a welfare-reducing, negative externality, job search cost upon firms such that the number of jobs offered exceeds the number of employees actually demanded.

We have access to data on approved individual I-129 petitions acquired through a Freedom of Information Act (FOIA) request to USCIS. This allows us to construct several descriptive figures illustrating another important change to firm behavior: Approved, cap-subject, H-1B workers are increasingly concentrated among fewer firms. Consider Figure 2. The left panel displays Lorenz curves for the concentration of new H-1B issuances to employees of cap-subject firms. When H-1B issuances were unconstrained by the H-1B cap in FYs 2002 and 2003, 78% unique firm names petitioned for a single H-1B worker, which accounted for 41% of all new cap-subject H-1Bs issued during that period. 90% of firms filed two or fewer petitions, accounting for just over half (53%) of new issuances. Contrast this outcome with FYs 2008 and 2009 when the cap had decreased to 85,000 and all new H-1Bs were allocated by random lottery. Unique firm names petitioning for one or two cap-subject H-1B workers accounted for 83% of the sample but just over a quarter of new issuances (26.5%). New H-1B employment was far more concentrated in a fewer set of firms during these lottery years.

One limitation of this exercise is that a single firm in the dataset can be represented by different firm names due to typos, misspellings, or alternative naming conventions. For example, "IBM", "IBM CORP", and "IBM GOLBAL SVCS IGS INDIA PVT LTD" (sic) are just three of 87 unique firm names containing the letters "IBM". It is possible that firms took greater care to ensure consistent name spellings in later years, thus leading to a spurious change in the measured concentration of new H-1B hires between these time periods. We believe the right panel of Figure 2 provides suggestive evidence against this possibility, however. This panel displays Lorenz curves for colleges, universities, and non-profit research institutions – employers exempt from H-1B limits throughout this time period. Unlike with cap-subject firms, we see no discernible change in new H-1B employment concentration.[9]

[9]Given the timing of H-1B petitions, we record all I-129 petitions beginning in April and moving forward

Figure 3 provides a different look at the same phenomenon. The left panel illustrates that the share of new H-1B issuances allocated to employees of firms hiring five or fewer total H-1B workers was above 60% in FYs 2002 and 2003 for cap-subject firms. That share declined precipitously when H-1B limits fell. The share among cap-exempt firms, however, remained fairly static at 30%. In contrast, the right panel displays the shares of new H-1Bs awarded to workers at firms hiring 250 or more total H-1B employees in a given year. This H-1B intensity was never common among cap-exempt employers, but it raised dramatically from about 9% in FY 2002 and 2003 to over 35% in 2011 for cap-subject firms.

Although our FOIA I-129 dataset is limited in its years of coverage, LCA filings are publicly-available and can provide longer time series evidence. As noted above, LCAs might provide a measure of intent to hire foreign labor. Unfortunately, the data is noisier. Specified work start dates are merely projections at best and might not correspond to actual projected start dates at all.[10] LCAs provide no indication of whether the employer is cap exempt. Moreover, LCAs are not linked to employees, so they will not indicate whether the worker will be seeking new (cap-subject) or continued (cap-exempt) H-1B status.

Despite these limitations, the documented trends in Figure 4 mirror those of Figure 3. We proxy for cap-exempt firms as those with “UNIV”, “COLLEGE”, or “INSTITUTE” in their names, except for those with “UNIVERSAL”. We graph trends across calendar years of stated work start dates. Among cap-subject firms, we see a sustained decline in the proportion of LCAs filed by firms submitting five or fewer total LCAs per year, stabilizing at about 6.5% to 8.5% from 2012 through 2019. The figure hovers around 10% for cap-exempt employers throughout the period.

The right panel shows that since 2010, roughly 60% or more of LCAs filed by cap-subject firms have originated with firms seeking 250 or more H-1B workers. This compares to a figure of just 40% in years when statutory H-1B limits were high. The share among cap-exempt firms declines over this period, however.

Our paper contends that these trends – unique among cap-subject firms – are not random. Rather, they reflect firms’ rational responses to restrictive limits on H-1B hiring. The following section develops a theory to model this behavior and its cost implications.

as belong to H-1B caps for the fiscal year beginning in the following October. Mayda et al. (2018) follow this strategy and find similar behaviors in H-1B concentration.

[10]Firms need to have approved LCAs in March if they intend to file an I-129 petition in April, but they can only submit an LCA six months in advance of the stated work start date. Peri et al. (2015b) document pre-dating behavior in which firms file LCAs with stated start dates in September for work that is more accurately scheduled to begin in October.

3 Theoretical Model

In this section, we build a theoretical model to understand how restrictive H-1B caps affect firm behavior. Before turning to the mathematics, it is helpful to understand the intuition that will drive the main results.

First, imagine a lottery contest. Suppose the government decreases the number of winners and therefore the probability of any single entry from winning. A competitor with a strong interest in winning might rationally respond by increasing its number of entries. This further reduces the probability of any single entry from winning and therefore imposes an externality on other competitors. If all competitors behave this way then it will create a multiplier effect that generates and amplifies large costs. In the context of our paper, the competitors are firms that increase their number of lottery entries by offering an increasing number of jobs to foreign-born workers, many of whom will lose the H-1B lottery.

Second, entering the lottery is not a costless proposition. Firms wanting H-1B workers will need to conduct job searches, make job offers, complete government forms, and submit processing fees. Firms are heterogenous – some will find the procedure more cumbersome than others. Those finding the process to be particularly onerous might not participate at all.

Third, firms that fail to hire their desired number of H-1B workers directly are not completely shut off from the foreign labor market. They can indirectly hire such a worker by contracting with firms specializing in H-1B employment to provide the desired services. That is, they can meet their labor needs by contracting with outsourcing firms as Costa and Hira (2020) define the term.[11]

With that intuition established, we begin the formal modelling by considering a firm, indexed by i, that faces a CES demand for its product and can produce x_i units using a combination of native and foreign labor. In stage 1, the firm chooses the number of H-1B applications, F_{iA}, so as to maximize expected profits. In stage 2, the lottery outcome is realized and the firm learns of the number of successful bids, $F_{iS} \leq F_{iA}$. The firm then chooses F_i – the number of immigrant workers – and N_i – the number of native workers to use. Thus, the firm – regardless of whether it specializes in providing outsourcing services or not – chooses to interact with other firms to either rent out or contract for immigrant workers in the second stage. We can solve the firm's problem backwards starting from stage 2.

[11]Our paper builds a theoretical model of outsourcing labor to third party sites within the United States. Complementary but distinct empirical analysis in Glennon (2020) finds that restrictive H-1B policies cause firms to offshore jobs to their foreign affiliates.

3.1 Stage 2

3.1.1 Setup

After the lottery outcome is realized, the firm learns its number of successful H-1B bids F_{iS}. It can hire up to this number of workers at immigrant labor wage w^F, which is the wage actually received by H-1B workers. It can also choose to either rent out or contract for immigrant labor at a wage rate of w^O, which in equilibrium will be greater than w^F.

3.1.2 Demand and Revenue

The firm is monopolistic and faces a CES demand curve derived from an underlying quasi-linear preference structure provided in Appendix A. Demand for good x_i is:

$$x_i = p_i^{-\sigma} \alpha_i \frac{E}{P^{1-\sigma}}, \tag{1}$$

where E is the total expenditure on all varieties and is constant; P is the price index for all the varieties of the good produced in this sector, and is constant from the firm's perspective;[12] α_i is a taste parameter; and σ is the elasticity of substitution across products.

The inverse demand function for the firm is:

$$p_i = \left(\frac{1}{\alpha_i} \frac{E}{P^{1-\sigma}} \right)^{1/\sigma} x_i^{-1/\sigma} \tag{2}$$

The firm's revenue generated by sales of its good (i.e., revenue excluding any potential income from renting out excess labor) is:

$$R_i = \left(\alpha_i \frac{E}{P^{1-\sigma}} \right)^{-1/\sigma} x_i^{\frac{\sigma-1}{\sigma}} \tag{3}$$

3.1.3 Production Function and Cost

Firms produce good x_i according to a CES production function that combines native (N) and foreign (F) labor with an elasticity of substitution of ϵ, augmented by a firm-specific productivity parameter (A_i):

$$x_i = A_i \left(N_i^{\frac{\epsilon-1}{\epsilon}} + F_i^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{4}$$

Substituting (4) into (3), production revenue is:

[12] The constant E setup is useful for expositional simplicity but could be relaxed without affecting the thrust of our analysis.

$$R_i = \left(\frac{1}{\alpha_i}\frac{P^{1-\sigma}}{E}\right)^{-1/\sigma}\left[A_i\left(N_i^{\frac{\epsilon-1}{\epsilon}} + F_i^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}}\right]^{\frac{\sigma-1}{\sigma}} \tag{5}$$

The cost of production accounting for the net cost of outsourcing is:

$$C_i = w^N N_i + w^F F_i + \left(w^O - w^F\right)(F_i - F_{iS}) \tag{6}$$

Note that $F_i - F_{iS}$ is the difference between the actual number of foreign workers the firm uses in the production process and the number of successful H-1B bids it has received. When positive, it therefore represents "imported" labor. The firm contracts with third party providers to supply labor services at wage rate w^O. These third party providers are likely to be outsourcing specialists, but they may also be regular firms that simply won the right to employ more workers than it intended to hire. When negative in value, C_i includes the revenues firms receive from renting out their H-1B workers to firms desiring more labor.

3.1.4 Profit and Factor Expenditure Functions

The firm's profits in the second stage, excluding the search costs incurred in the first stage (described later) are:

$$\Pi_i = R_i - C_i \tag{7}$$

Given the realized number of successful H-1B bids, F_{iS}, the firm chooses the optimal native (N_i) and foreign (F_i) employment to maximize (7). This gives us two first-order wage conditions:

$$\frac{\partial R_i}{\partial x_i}\frac{\partial x_i}{\partial N_i} = w^N \tag{8}$$

$$\frac{\partial R_i}{\partial x_i}\frac{\partial x_i}{\partial F_i} = w^O \tag{9}$$

We see from (9) that the marginal revenue product of immigrant labor is equal to the outsourcing wage at the optimum. This may be surprising given that the firm will generally use at least some labor from its own successful bids, in which case it would pay w^F instead of w^O. However, since H-1B workers can always be "exported" to other firms, the opportunity cost of using a worker in production is still given by w^O.

The production function in (4) and wage conditions in (8) and (9) combine to deliver a standard marginal rate of technical substitution condition:

$$\frac{F_i}{N_i} = \left(\frac{w^N}{w^O}\right)^{\epsilon} \tag{10}$$

We next insert (10) into (8) and (9) to solve for the optimal expenditure on native and foreign-born labor:

$$w^N N_i = \frac{\left(w^N\right)^{1-\epsilon}}{\left(w^O\right)^{1-\epsilon} + \left(w^N\right)^{1-\epsilon}} \frac{\sigma - 1}{\sigma} R_i \tag{11}$$

$$w^O F_i = \frac{\left(w^O\right)^{1-\epsilon}}{\left(w^O\right)^{1-\epsilon} + \left(w^N\right)^{1-\epsilon}} \frac{\sigma - 1}{\sigma} R_i \tag{12}$$

These are standard CES factor expenditure functions, except with the opportunity cost of hiring foreign-born labor, w^O, being the relevant wage for immigrants.

3.1.5 Profits

We can plug expenditure functions (11) and (12) into the cost expression (6) to get:

$$C_i = \frac{\sigma - 1}{\sigma} R_i - \left(w^O - w^F\right) F_{iS} \tag{13}$$

The firm's profits are then:

$$\Pi_i = \frac{1}{\sigma} R_i + \left(w^O - w^F\right) F_{iS} \tag{14}$$

The above expression is similar to the standard CES profit maximizing choice except that the second term captures the benefit of having more successful H-1B bids.

We take two steps to calculate the optimal level of output (x_i) in terms of the parameters of the model. First we substitute the cost expressions (11) and (12) into the production function (4). Then we insert the revenue function from (3). The resulting output functions become:

$$x_i = \frac{\sigma - 1}{\sigma} A_i \left[\left(w^N\right)^{1-\epsilon} + \left(w^O\right)^{1-\epsilon}\right]^{\frac{1}{\epsilon - 1}} R_i \tag{15}$$

$$x_i = \left(\frac{\sigma - 1}{\sigma} A_i\right)^{\sigma} \left(\frac{1}{\alpha_i} \frac{P^{1-\sigma}}{E}\right) \left[\left(w^N\right)^{1-\epsilon} + \left(w^O\right)^{1-\epsilon}\right]^{\frac{\sigma}{\epsilon - 1}} \tag{16}$$

Equation (16) gives us the firm's optimal choice as a function of wages as well as demand and productivity parameters from the firm and industry. This finally allows us to write profits, given by (14), as a function of parameters and factor prices:

$$\Pi_i = K_i \left[\left(w^N\right)^{1-\epsilon} + \left(w^O\right)^{1-\epsilon} \right]^{\frac{\sigma-1}{\epsilon-1}} + \left(w^O - w^F\right) F_{iS}, \tag{17}$$

where $K_i \equiv \left(\frac{\sigma-1}{\sigma} A_i\right)^{\sigma-1} \frac{1}{\sigma} \frac{1}{\alpha_i} \frac{P^{1-\sigma}}{E}$ is a constant from the standpoint of the firm. We will now be able to use (17) in the firm's first stage problem.

3.2 Stage 1 Problem

3.2.1 Probability Distribution and Density Functions

In Stage 1, the firm chooses the optimal number of H-1B applications, F_{iA}, so as to maximize expected profits, taking as given the optimal choices it will make in Stage 2. In order to write down the firm's expected profits, we need to specify the stochastic process that determines the successful number of bids, F_{iS}. This poses a potential problem because while we can follow standard practice in ignoring any integer constraint on F_{iA}, we would still somehow have to account for the fact that the distribution of F_{iS} will be binomial with a constant probability ρ of winning the lottery. We deal with this challenge by using a continuous version of the binomial distribution (Ilienko, 2013), the cumulative density function (CDF) of which is:

$$G\left(F_{iS}; F_{iA}\right) = I_{1-\rho}\left(F_{iA} - F_{iS}, 1 + F_{iS}\right), \tag{18}$$

where $I_{1-\rho}\left(.\right)$ is the regularized incomplete beta function. The associated probability density function is:

$$g\left(F_{iS}; F_{iA}\right) = \frac{dI_{1-\rho}\left(F_{iA} - F_{iS}, 1 + F_{iS}\right)}{dF_{iS}}, \tag{19}$$

which exists but does not have a simple closed form. This continuous binomial distribution is a generalization of the regular binomial distribution that is also defined for non-integer values of F_{iA} and F_{iS} through the use of gamma functions in place of factorial-based expressions.

3.2.2 Expected Profits

Given the probability distribution functions, we can write the firm's expected profits as:

$$E\left[\pi\right] = \int_0^{F_{iA}} \Pi_i\left(F_{iS}\right) dG\left(F_{iS}\right) - cF_{iA} - c_i \frac{1}{\lambda} F_{iA}^{\lambda} \tag{20}$$

The firm's cost of filing an H-1B petition is represented by cF_{iA}. Although the government varies these fees across firms as described above, we hold them constant for simplicity. In contrast, we allow the search costs associated with finding F_{iA} potential H-1B workers $\left(c_i \frac{1}{\lambda} F_{iA}^{\lambda}\right)$ to vary, with c_i serving as a firm-specific search cost parameter. We assume that $\lambda > 1$ for most employers so that these regular firms face an increasing marginal cost of finding H-1B workers.[13] Conversely – and to be discussed more thoroughly in the next subsection – we assume that a select group of outsourcing firms face constant marginal search costs ($\lambda = 1$), which will motivate their decision to specialize in providing outsourcing services.

By substituting the second-stage maximum profit condition (17) into the first-stage expected profit function (20), we get:

$$E[\pi] = K_i \left[\left(w^N\right)^{1-\epsilon} + \left(w^O\right)^{1-\epsilon} \right]^{\frac{\sigma-1}{\epsilon-1}} + \left(w^O - w^F\right) \rho F_{iA} - cF_{iA} - c_i \frac{1}{\lambda} F_{iA}^{\lambda}, \tag{21}$$

where ρF_{iA} is the mean of the binomial distribution. The first order condition for the expected profit maximization problem is then:

$$\left(w^O - w^F\right) \rho - c - c_i F_{iA}^{\lambda-1} = 0 \tag{22}$$

The marginal benefit of a bid is the probability of success, ρ, multiplied by the premium for a successful bid, $\left(w^O - w^F\right)$. Solving for a firm's optimal number of H-1B petitions we get:

$$F_{iA} = \left[\frac{\left(w^O - w^F\right) \rho - c}{c_i} \right]^{\frac{1}{\lambda-1}} \tag{23}$$

Intuitively, the optimal number of H-1B applications is increasing in the expected premium, $\left(w^O - w^F\right) \rho$, and decreasing in both the application cost c and the search cost parameter c_i. Perhaps surprisingly, the optimal number of bids is entirely independent of the firm's actual needs, which would depend on costs and profits. This is because the firm's ability to rent out or contract for labor as needed fixes the benefit of a successful bid to equal $w^O - w^F$ at the firm's optimum.

3.2.3 Outsourcing Firms

In principle, any firm can rent out its labor to other firms. However, we distinguish between regular firms and outsourcing firms that specialize in providing labor to third party sites. Regular firms have increasing marginal search costs (driven by $\lambda > 1$) as described above. Outsourcing firms, on the other hand, face a constant marginal search cost ($\lambda = 1$) equal to c_X. As noted in Section 2, outsourcing specialists in reality are likely to have a large pool of potential H-1B workers already on their payroll overseas and would not face the same

[13] We choose this specific iso-elastic functional form for these costs for simplicity and ease of interpretation. Qualitatively, any increasing marginal cost function would give us similar results.

challenges as domestic firms when seeking to find more workers. With this assumption, the profit maximizing first order condition in (22) will occur at a corner solution such that:

$$\left(w^{O}-w^{F}\right)\rho-c_{X}-c=0 \tag{24}$$

3.3 Equilibrium

3.3.1 Equilibrium Concept

In the above analysis, we assume the firm faces uncertainty about the number of successful bids it receives (F_{iS}) but not about other values such as the probability of success (ρ) and the wage it will have to pay to H-1B workers, whether employed directly or through its decisions to contract for or rent out labor (w^O). This requires a rational expectations equilibrium concept, where firms' choices regarding the number of submitted H-1B petitions are optimal given the realized wages and ρ in the second stage. We also assume that there is no aggregate uncertainty. Ultimately, the number of petitions filed, the probability of success, and wages paid are endogenous variables determined by the size of the H-1B quota and the model's cost and elasticity parameters.

3.3.2 Equilibrium Conditions

The total supply of immigrant labor is fixed by the H-1B quota, Ω. The supply of native workers is also assumed to be constant and is equal to $\overline{N}$. We can write the outsourcing wage in terms of the native wage by combining the optimal expenditure functions in (11) and (12), and then summing over all firms. This delivers a marginal rate of technical substitution relation that is the aggregate analogue to the firm-level equivalent expressed in (10):

$$\frac{\Omega}{\overline{N}}=\left(\frac{w^{N}}{w^{O}}\right)^{\epsilon} \tag{25}$$

We can also sum over firms' foreign labor expenditures in (12) and simplify to get:

$$w^{O}\Omega=\frac{\left(w^{O}\right)^{1-\epsilon}}{\left(w^{O}\right)^{1-\epsilon}+\left(w^{N}\right)^{1-\epsilon}}\frac{\sigma-1}{\sigma}E \tag{26}$$

The model is closed by setting the successful number of bids equal to the actual H-1B quota:

$$\rho\left(\sum_{i}F_{iA}+\sum_{j}F_{jA}^{O}\right)=\Omega \tag{27}$$

Here, we separate the number of H-1B petitions (F_{iA}) submitted by regular firms i and applications (F^O_{jA}) filed by outsourcing firms j. We can now use equations (23)-(27) to determine the endogenous variables F_{iA}, $\sum F^O_{jA}$, w^O, w^N, and ρ. Note that the distribution of H-1B petitions across outsourcing firms is indeterminate because these firms make no economic profits and are therefore willing to provide any number of workers in equilibrium.

3.3.3 Solving for the Equilibrium

To solve for the equilibrium values of our endogenous variables, we first combine the optimal H-1B petition conditions in (23) and (24) to solve for the number of bids submitted by a regular firm F_{iA} in (28). Intuitively, a firm submits fewer H-1B applications as its own marginal costs of extending a job offer rise, but more applications if outsourcing specialists' marginal search costs are high.

$$F_{iA} = \left(\frac{c_X}{c_i}\right)^{\frac{1}{\lambda-1}} \tag{28}$$

Next we can combine (24), (25), and (26) to determine the outsourcing wage as a function of the H-1B quota, native employment, and the elasticity parameters of the model:

$$w^O = \frac{\sigma-1}{\sigma} E\left[\Omega + \left(\overline{N}\right)^{\frac{\epsilon-1}{\epsilon}} (\Omega)^{\frac{1}{\epsilon}}\right]^{-1} \tag{29}$$

From (29), w^O is increasing in the labor expenditure of the H-1B sector, which is given by $\frac{\sigma-1}{\sigma}E$. It is decreasing in the supply of H-1B workers, Ω. It is also decreasing in the supply of native labor, $\overline{N}$, as long as native and immigrant workers are more substitutable than what would be implied by Cobb-Douglas production (i.e., if $\epsilon > 1$).

The corner solution for the optimal number of bids submitted by outsourcing firms in (24) delivers the equilibrium probability of a successful bid, ρ, in (30). Success probabilities rise in proportion to outsourcing firms' marginal search costs and H-1B filing fees. The more difficult it is to find foreign workers, the fewer applications will be submitted, and hence the probability of winning will be larger. If the outsourcing premium ($w^O - w^F$) is large, then firms will be incentivized to submit more H-1B applications, thus reducing the probability of winning any one individual bid.

$$\rho = \frac{c_X + c}{w^O - w^F} \tag{30}$$

Finally, we can determine the total number of H-1B applications filed by outsourcing firms ($\sum F^O_{iA}$) by using the simple supply and demand framework in (27). Consistent with the discussion above, this sum total in (31) increases if the outsourcing premium is high. It decreases if marginal search costs, application fees, or the sum of regular firms' filings are high. The role of the quota is our central interest, which we examine in the next section.

$$\sum_j F_{jA}^O = \frac{\Omega}{c_X + c}\left(w^O - w^F\right) - \sum_i F_{iA} \tag{31}$$

4 Results

4.1 The H-1B Quota, H-1B Concentration, and Search Externalities

The behavior identified in (31) allows us to understand how government decisions to tighten the H-1B quota affect the number of H-1B petitions filed by outsourcing firms. Since $\sum_i F_{iA}$ is determined by (28) and is independent of the quota, this also tells us whether the bids and foreign employment become more or less concentrated among outsourcing firms. Before illustrating this, it is useful to differentiate the outsourcing wage in (29) with respect to the quota (Ω). The resulting equation (32) intuitively implies that increasing (i.e. loosening) the quota would decrease the outsourcing wage, w^O.

$$\frac{dw^O}{d\Omega} = -\frac{\sigma - 1}{\sigma} E \frac{1 + \frac{1}{\epsilon}\left(\overline{N}/\Omega\right)^{\frac{\epsilon-1}{\epsilon}}}{\left[\Omega + \left(\overline{N}\right)^{\frac{\epsilon-1}{\epsilon}}(\Omega)^{\frac{1}{\epsilon}}\right]^2} < 0 \tag{32}$$

With this derivative in mind, we can next differentiate the aggregate number of outsourcers' bids in (31) and multiply by -1 to get:

$$-\frac{d\sum_j F_{jA}^O}{d\Omega} = -\frac{1}{\rho} - \frac{\Omega}{c_X + c}\frac{dw^O}{d\Omega} \tag{33}$$

We see from this expression that in principle, a tightening of the H-1B quota could either increase or decrease the number of bids by outsourcing firms and therefore could either increase or decrease the concentration of bids amongst these firms. This is because two competing effects are at play. The first is the direct effect of the quota, Ω; the second is the effect through the probability of securing a successful bid, ρ.

Consider the first channel captured by $-\frac{1}{\rho}$. There is a mechanical relationship such that the win probability of success equals the quota divided by the total number of bids. Holding the probability constant, any decline in the quota would be offset by a decline in H-1B applications equal to $-\frac{1}{\rho}$. In economics terms, a lower quota means that firms are restricted to hiring fewer H-1B workers. Thus, there is a reduced need to rely on the services of the outsourcing firms. Conversely, if Ω was large, then at a constant ρ the industry would hire a larger number of workers and would rely on the outsourcing firms to avoid the otherwise increasing search costs.

The second channel, represented by $-\frac{\Omega}{c_X+c}\frac{dw^O}{d\Omega}$, recognizes that the probability of success will not remain constant. Thus, there is an indirect effect that operates through this probability. Since a lower Ω will reduce ρ, it means that to hire a given number of H-1B workers in the aggregate, the number of bids has to be larger and firms will need to search for more workers. These increasing search costs faced by regular firms therefore increase the demand for the services of the outsourcing firms.

This probability channel could also be interpreted as a type of lottery externality. To see how, we can consider a rather artificial but nonetheless useful counterfactual. Suppose that instead of individual firms choosing their optimal values of F_{iA} in a decentralized manner, a social planner chooses the bids for each firm. In this scenario, the social planner is still subject to the lottery system but makes the decisions on the behalf of firms. It should be apparent that the social planner would have no reason to make more aggregate bids than there are H-1B quota slots. Any petition filed in excess of Ω entails social costs in the form of search and bid costs that offer no corresponding social benefit. The externality here arises from the fact that when a firm makes a bid, it internalizes the potential benefit to itself but not the cost to other firms in the form of a reduced probability of success. Hence, any probability of success less than one reflects an inefficient use of aggregate resources. This is why the second channel in (33), which captures the effect of a change in ρ, also captures this externality.

As discussed above, equation (33) suggests that there is no particular reason why one of these channels must dominate the other *a priori*. In order to get a better sense of whether an H-1B quota would increase or decrease the number of applications submitted by outsourcing firms (i.e., the sign of equation (33)), we can write down an alternative expression by differentiating the supply and demand condition in (27) with respect to Ω:

$$-\frac{d\sum_j F_{jA}^O}{d\Omega} = \frac{1}{\rho}\left(\frac{d\rho}{d\Omega}\frac{\Omega}{\rho} - 1\right) \tag{34}$$

The right hand side of this equation includes a term representing the elasticity of the win probability with respect to the H-1B quota, $\frac{d\rho}{d\Omega}\frac{\Omega}{\rho}$. This elasticity, which we henceforth represent with ψ, is the key parameter that drives the main implications of the H-1B quota and lottery allocation in our model. The left hand side of (34) is positive so long as $\psi > 1$. In economics terms, the random allocation of H-1B status implies that a lower H-1B cap will cause an increasing concentration of H-1B employment among outsourcing firms when the elasticity of the win probability with respect to the quota is greater than one. The intuition of this result is straightforward. An elasticity of one would imply that the total number of bids remains constant when the quota is decreased so that the probability of success also decreases proportionately. In reality, and as we will show numerically in the next section, the total number of bids likely *increases* due to the quota. Interpreting this through the lens of the model, it implies a greater use of outsourcing firms and therefore a greater concentration in bids.

4.2 Search Cost Implications of Outsourcing Firms

Our analysis above explained how the H-1B lottery can lead to a greater concentration of H-1B petitions and employment among outsourcing firms, and that this process ultimately reflects a negative externality that leads to an overprovision of H-1B bids. It did not, however, answer the question of whether outsourcing firms are improving or worsening the overall situation.

One way to think about this issue is to consider possible policy actions and responses. The government could impose restrictions, for example, that would raise the search costs for outsourcing firms, c_X. This higher value for c_X would reduce regular firms' use of the services provides by outsourcing firms. This service has value since it allows firms in the H-1B sector to effectively find immigrant labor at a lower cost than would be otherwise possible, but excess bids generate negative externalities. The net effect of such a restriction is therefore not obvious, *a priori*. By focusing on a continuous metric such as c_X to evaluate the effect of such restrictions, we can evaluate externality implications using simple differentiation. The total costs of bidding, C, including both the cost of the application and the search costs, are:

$$C = \sum_i \left[cF_{iA} + c_i \frac{1}{\lambda} F_{iA}^{\lambda} \right] + \sum_j \left(c + c_X \right) F_{jA}^{O} \tag{35}$$

The first sum in (35) captures the costs incurred by regular firms and the second reflects those incurred by outsourcing firms. We then develop the implications of raising the marginal search costs of outsourcing firms in (36) by turning to the total application behavior described in (28) and (31), aggregate costs outlined in (35), and then differentiating. The key implication clearly outlined here is that an increase in the search costs of the outsourcing firms *reduces* aggregate H-1B search costs.

$$\frac{dC}{dc_X} = -\sum_i F_{iA} < 0 \tag{36}$$

The intuition for this result is straightforward. We can think of the total application cost as being equal to the total number of bids – both by regular and outsourcing firms – multiplied by the average cost per H-1B application. Consider the effect of an increase in c_X that increases $c_X + c$ by 1%. From the outsourcing firms' corner solution in (24), we see that this would lead to a 1% increase in the probability of a successful bid. Therefore, from the supply and demand condition in (27), this leads to a 1% decrease in the total number of bids as well.

In terms of average costs, a 1% increase in $c_X + c$ translates into a 1% increase in the (constant) average cost for outsourcing firms. However, it translates into a less than 1% increase in the cost of regular firms: the 1% increase in $c_X + c$ increases the marginal application cost in equilibrium for regular firms by 1% but it does not correspondingly increase the costs on all the applications up to the marginal ones. Hence, taken together, we have a less than 1% increase in the average application costs of all firms and a 1% decrease in the total number of applications. As a result, the total costs decrease when c_X increases.

These results altogether imply that a larger outsourcing sector (i.e., a lower search cost for the outsourcing sector) worsens the negative externality generated by the lottery system by encouraging a higher number of bids. While the outsourcing firms could in principle be beneficial because they allow firms to find a larger number of workers at a lower search cost, this potential benefit is not realized in the presence of the behavioral response induced by the lottery.

5 Results: Numerical Exercises

5.1 Win Elasticities

The elasticity of the win probability with respect to the quota, $\psi = \frac{d\rho}{d\Omega}\frac{\Omega}{\rho}$, is the key parameter for determining many of the model's economic predictions. Its size is an empirical question that cannot be answered through regression analysis, but we can gain insight into this parameter by re-expressing it as a ratio of log differentials, $\psi = \frac{d\ln(\rho)}{d\ln(\Omega)}$, and comparing two different eras: One with and one without a lottery.

One advantage of this approach is that it simplifies the elasticity. The log win probability differential reduces to $\ln(\rho) - \ln(1)$, where ρ now represents the win rate under the lottery scenario. One disadvantage of this approach is that non-lottery outcomes (that is, values of $\rho = 1$) exist for a range of Ω values. Thus, although we observe an H-1B cap of $\Omega = 85,000$ during the lottery era that far exceeds the high, non-binding, H-1B cap of $\Omega = 195,000$ from a period of H-1B certainty, it would be inappropriate to simply calculate the log quota differential as $\ln(85,000) - \ln(195,000)$.

As a solution to this limitation, we re-express the elasticity as a function of the change in log H-1B filings, not as a function of the quota. To do so, first let $\overline{F} = \sum_i F_{iA} + \sum_j F^O_{jA}$ represent the aggregate number of cap-subject H-1B petitions submitted. Then consider a simple partial differentiation of the log-linearized expression of the mechanical relationship between the probability of winning, the size of the quota, and the aggregate number of applications in (27):

$$d\ln(\Omega) = d\ln(\rho) + d\ln(\overline{F}) \tag{37}$$

Altogether, these assumptions allow us to rewrite our elasticity of interest as a function of the win probability when the quota is binding and the change in log H-1B petitions filed between the two periods:

$$\psi = \frac{d\ln(\rho)}{d\ln(\Omega)} = \frac{\ln(\rho)}{\ln(\rho) + d\ln(\overline{F})} \tag{38}$$

Unfortunately, measurement of $\overline{F}$ presents two of its own challenges. First, the number of desired H-1B filings is a latent variable (as described in Section 2). Second, $\overline{F}$ changes over time due to both quota-induced supply shocks and demand shocks. We are only interested in the former. That is, $d\ln(\overline{F})$ conceptually represents the percentage-change in H-1B petitions attributable to the lottery (a supply constraint), not from the actual demand for H-1B labor.

We can overcome these challanges. Suppose that real demand for H-1B workers, D, can be measured and that $\phi = \frac{\overline{F}-D}{\overline{F}}$ represents excess H-1B filings as a share of total filings – i.e., the percentage of H-1B petitions submitted solely in an attempt to win the H-1B lottery. We consider three approaches to measuring $d\ln(\overline{F})$. All three reduce our key elasticity (ψ) to a function of just two parameters: the win probability during the lottery year (ρ) and excess H-1B filings as a share of total filings (ϕ).

5.1.1 Win Elasticities: Method 1

Our first method for calculating the percentage-change in lottery-induced filings assumes that in the absence of a lottery, the number of filings would equal the actual demand for H-1B workers, D. We know from other studies, however, that the number of actual submissions filed, $\overline{F}$, far exceeds the level of actual demand. If we attribute all of these excess filing to the lottery, then $d\ln(\overline{F}) = \ln(\overline{F}) - \ln(D)$. Since $\phi = \frac{\overline{F}-D}{\overline{F}}$, we can rearrange so that $D = \overline{F} * (1-\phi)$. Substituting this into the $d\ln(\overline{F})$ term and simplifying yields:

$$\psi = \frac{\ln(\rho)}{\ln(\rho) - \ln(1-\phi)} \tag{39}$$

This expression allows us to compute win probability elasticities for values of ρ and ϕ suggested by the data. Moreover, it allows us to graph values of ψ as a function of ρ and ϕ without appealing to claims of knowing what actual demand for H-1B workers equals.

5.1.2 Win Elasticities: Method 2

The second approach to estimating ψ assumes that in the absence of a lottery, the number of filings would equal the size of the quota. In the presence of a lottery, however, excess filings beyond measurable demand exist. One must still remove any demand-driven increase in filings that have occurred over time. Thus, this approach compares excess filings beyond the quota $(\overline{F} - D + \Omega)$ filed in lottery years relative to the quota itself.

One way to conceptualize the counterfactual that this exercise captures is that it represents a pollyannaish view of the lottery: It compares excess filings above the quota to a world in which firms never file any H-1B petitions above the quota because they know they will never be able hire these workers. This would be a world in which firms exhibit no strategic behavior after the quota is reached; they make no attempt to submit additional H-1B petitions in order to secure a better chance of hiring the number of workers they actually desire. Firms do not even bother demanding any real workers beyond the quota since they know they cannot be hired. Another way to consider this counterfactual is to recognize that it measures the rise in total filings (removing contributions from trends in demand) relative to the maximum number that could be filed with a certainty of winning the lottery (i.e., such that $\rho = 1$). In any case, we substitute the expression $d\ln(\overline{F}) = \ln\left(\overline{F} - D + \Omega\right) - \ln(\Omega)$ into (38), simplify the elasticity, and express it as a function ρ and ϕ:

$$\psi = \frac{\ln(\rho)}{\ln(\rho + \phi)} \tag{40}$$

5.1.3 Win Elasticities: Method 3

The final approach to measuring demand-induced changes to H-1B filings results in a win probability elasticity much like the first. Recognizing that we need to calculate a change in $\overline{F}$ at a constant level of demand, we can write this (in an elasticity form) as:

$$\frac{d\overline{F}\big|_{D=\overline{D}}}{\overline{F}} = \left(\frac{d\overline{F} - dD}{\overline{F}}\right) \tag{41}$$

Next, we can approximate the left hand side with a log differential. We can also break apart the right hand side by approximating differentials with discrete changes between the lottery and non-lottery filing (F_0) and demand (D_0) outcomes.

$$d\ln\left(\overline{F}\right) = \left[\frac{\left(\overline{F} - F_0\right) - \left(D - D_0\right)}{\overline{F}}\right] \tag{42}$$

Finally, by recognizing that filings and demand are equivalent in a world without a lottery, this differential simplifies to:

$$d\ln\left(\overline{F}\right) = \left(\frac{\overline{F} - D}{\overline{F}}\right) = \phi \tag{43}$$

Thus, our third approach to approximating lottery-induced changes in H-1B petitions results in a win probability elasticity with respect to the lottery of:

$$\psi = \frac{\ln\left(\rho\right)}{\ln\left(\rho\right) + \phi} \tag{44}$$

5.1.4 Win Elasticities Predicted by Data

The three methods above used to approximate the elasticity of the win probability with respect to the quota are by no means exhaustive. For example, one might prefer to use a discrete measure of $\%\Delta\rho = \rho - 1$ as opposed to the log differential $d\ln\left(\rho\right) = \ln\left(\rho\right)$. Similarly, $\left(\frac{\overline{F}-D}{D}\right) = \left(\frac{\phi}{1-\phi}\right)$ represents yet another possible (discrete) way to measure the growth in filings above demand. However, we will see that all three methods in this paper deliver strikingly similar implications. Note, for example, that since $\phi \approx -\ln\left(1-\phi\right)$ when ϕ is small, the methods used to approximate the elasticities in (39) and (44) result in similar values when firms submit few H-1B applications in excess of real demand. Furthermore the elasticity in (40) is equivalent to that in (39) when H-1B demand is equal to the quota.

We calculate specific elasticities and graphically display ψ estimates for a range of ρ and ϕ values using equations (39), (40), and (44). Table 1 provides calculations suggested by the data for three years. The first two – 2008 and 2009 – were first fiscal years in which all H-1Bs were allocated by lottery. Though only 85,000 new H-1Bs were available in those years, USCIS received more than 120,000 petitions during the first week of the application period for 2008 and roughly 163,000 for the following year. Also recall that Mayda et al. (2018) report that H-1B demand was 22%-33% higher than the 85,000 cap allowed in those years. The implied real demand for H-1B workers was therefore much lower than the number H-1B petitions filed. This alone illustrates that firms are conducting H-1B searches above and beyond their actual demand for foreign labor and that $\phi > 0$.

Table 1 also includes calculations for 2017 – the peak year for H-1B submissions. This necessitates an important caveat: 2017 is beyond the sample period included in the Mayda et al. (2018) analysis, so even a 33% increase in real demand could be a lower-bound estimate for that year. Thus, our resulting estimate of excess filings might be larger than reality for 2017.

The columns of Table 1 are distinguished by the year of observation and hence the number of H-1B filings and the probability of winning. The first three use the upper-bound Mayda et al. (2018) demand estimate of 33% above the quota for 2008, 2009, and 2017; Columns (4)-(6) use the lower-bound 22% figure. We use the three methods for computing the change in log H-1B filings described above. The top panel of Table 1 computes filings relative to estimated real H-1B demand (Method 1). The middle panel measures filing differences relative to the 85,000 H-1B quota (Method 2). The bottom measures excess relative to total filings (Method 3).

The table reports a range of estimates for the elasticity of H-1B win probability with respect to the quota. At a minimum, we estimate a value of 1.29 in 2008 when excess H-1B filings were comparatively modest and we assume that real demand for foreign labor was high. The elasticity is always smallest when measuring excess filings relative to total filings (bottom panel), which delivers quantitatively similar results to using Method 1's computation of excess filings relative to demand (top panel). Elasticities are highest when calculating filings relative to the quota (middle panel). Not surprising, the largest elasticity estimates arise in 2017 when the number of H-1B petitions was exceedingly high. Most importantly, however, all values are comfortably above one, implying that H-1B hiring restrictions, coupled with random lottery allocation, generate negative job search externalities and cause H-1B employment to be concentrated in among outsourcing firms.

Next, we turn to algebraic and graphical analysis for understanding the interplay between ϕ, ρ, and ψ. This is somewhat difficult to do in the sense that the parameter values are endogenously determined by the model – this is not a typical comparative statics exercise in which one simply changes an exogenous variable in order to examine the consequences of a shock. Nonetheless, we believe the mathematics and graphs are informative both about how the endogenous variables are related to each other and are helpful in identifying elasticity values consistent with real world data.

It is straightforward to show that the elasticity ψ is increasing in both ϕ and ρ, conditional upon $\phi \in [0, 1-\rho]$ and $\rho \in [0, 1]$, where the former range restriction arises from an assumption that real H-1B demand in lottery years equals the size of the quota at a minimum. Let us consider the intuition behind these results in turn. The first is almost tautological. It implies that if excess filings are low, then a change to the quota will lead to an almost proportional change in the win probability. High values of ϕ, in contrast, arise when firms file a large number of H-1B petitions in excess of demand, thus implying that the win probability is highly responsive to quota restrictions. The importance of this result becomes more apparent when recognizing that the elasticity of the win probability with respect to the quota is guaranteed to be greater than one so long as any firm submits more H-1B petitions in excess of their actual demand for H-1B workers. Moreover, any elasticity above one is indicative of a negative externality induced by the quota and lottery. Even a small proportion of H-1B petitions filed in excess of actual H-1B demand would be enough to generate such costs.

The reasoning behind the positive relationship between ρ and ψ is somewhat more complicated. The result implies that the responsiveness of the win probability to a quota shock is high if the win probability itself is also high. Behaviorally, this means that when firms are confident that the can win the lottery, they respond to quota restrictions by trying to game the system, so to speak, by conducting an excess number of searches and extending multiple offers to try and secure their desired number of workers. Collectively, this causes the actual win probability to drop precipitously. For a more mechanical understanding, first note that $1 - \rho$ is the maximum possible value of ϕ. It follows that the larger value that ρ takes, the closer any fixed value of ϕ is relative to its maximum and the more important firms' behavioral responses become. That is, the phenomenon of excess H-1B petition filings is more relevant to the responsiveness of the win probability to the quota when the win probability itself is large.

Figure 5 plots the elasticity estimates as a function of the percentage of excess H-1B applications that arise from the Methodologies 1-3 (moving from the left panel through the right) described above. Each graph displays contour plots for different values of ρ increasing in increments of 0.05 from a minimum of $\rho = 0.30$ (light colored) to a maximum of $\rho = 0.75$ (dark). Each panel also includes point estimates implied by observed win probabilities and the number of excess H-1B filings implied by the given methodology. The red line corresponds to the assumption that real H-1B demand is 22% above the quota; the blue line assumes demand is 33% higher than the quota. Graphs share the same vertical axis scale and are displayed side-by-side for comparison purposes.

Not surprisingly, all three methods demonstrate that the win elasticity with respect to the quota is high when firms submit a large number of petitions in an effort to win the lottery (i.e., when ϕ is low). Methods 1 and 3 produce quantitatively similar estimates, though Method 3 has the advantage of producing estimates that are the least sensitive to differences in ϕ and ρ values. Using real world data, it implies a minimum elasticity estimate of 1.29 (when assuming that demand is 33% higher than the quota) and a maximum of about 2.25. Even at the lowest values, elasticity estimates are large enough to imply that the quota and lottery generated negative search cost externalities. Note that elasticity estimates are always highest when using Method 2. Differences are particularly apparent with excess petition filings are high. Despite these disparities, however, we will see below that all three methods produces very similar estimates for search cost externalities resulting from the quota and lottery.

5.2 Concentration Implications

Assuming that regular firms and outsourcing specialists face the same win probability, then we can let $\theta = \frac{\sum_j F_{jA}^O}{\overline{F}}$ represent the share of H-1B approvals awarded to outsourcing firms. By combining the response in outsourcers' filings identified in (34) with the identity in (37), we can determine how the H-1B quota restriction affects the proportion of H-1B approvals granted to H-1B outsourcing specialists:

$$-\frac{d\theta}{d\ln(\Omega)} = (\psi - 1)(1 - \theta) \tag{45}$$

This expression states that H-1B concentration among outsourcing firms increases linearly with the win elasticity ψ and is positive so long as $\psi > 1$, which – as we have seen above – is likely. Furthermore, the marginal change in this share decreases as the share itself approaches one.

To get a sense of the magnitude of these predictions, we first need to consider possible values of θ from the data. There is no one definition for what constitutes an H-1B outsourcing firm or the proportion of H-1Bs such firms hire. Individuals and organizations generally opposed to the H-1B program argue that at least a third of cap-subject H-1B workers are hired by outsourcing firms.[14] Thus, $\theta = 0.33$ serves as a useful lower bound. We also examine an upper bound of $\theta = 0.60$ to reflect the percentage of firms that pay H-1B workers low wages according to Costa and Hira (2020).[15]

Table 1 presents our estimates for the change in the outsourcing employment share for 2008, 2009, and 2017 at these values of θ. Focusing on the first three columns (which assume that aggregate H-1B demand was 33% higher than the quota in lottery years) and the top panel of estimates (which uses Method 1 – described above to measure the change in H-1B petitions as excess H-1B filings relative to demand), we estimate that the lottery caused the proportion of H-1Bs awarded to outsourcing firms to rise by 0.12 to 0.21 percentage points in 2008, depending upon the assumed starting value of θ. This effect is much larger in 2009 and 2017 when the probability of winning the lottery was much lower and the elasticity of the win

[14] North (2020) argues that "outsourcing firms... control of 36 percent of the H-1B visas." Using earlier data and analysis by Ron Hira, Park (2015) similarly reports that "13 outsourcing companies took nearly one-third of all H-1B visas in 2014."

[15] Namely, the authors find that "Sixty percent of H-1B positions certified by the U.S. Department of Labor are assigned wage levels well below the local median wage for the occupation" and that "three-fifths of all H-1B jobs were certified at the two lowest prevailing wage levels in 2019."

probability with respect to the quota was much larger. Our third method (bottom panel) for calculating the change in log H-1B filings produces the lowest elasticity estimates. It therefore also produces the most muted response in the share of H-1Bs awarded to outsourcing firms.

Another source of insight into these magnitudes comes from our own data used to produce the Lorenz Curve for cap-subject employers in Figure 2. The top 1% of H-1B employers accounted for 24.3% of new H-1B employment in 2002 and 2003 when the H-1B cap was high and not binding, and 36.7% of H-1Bs in the 2008 and 2009 lottery years. If we think of these firms as outsourcers, then this implies a 12.4 percentage point increase in θ. If we broaden our definition to include the top 5% of H-1B employers, then θ rose by nearly 20 percentage points (from 0.380 to 0.577) between these two periods. Thus, these descriptive statistics fall within the range of estimates predicted by our theoretical model when we assume lower values for the win elasticity (e.g., $\psi = 1.30$) that are nonetheless consistent with observed behavior during the 2008 lottery year. Only at larger elasticities do predicted changes in H-1B concentration grow somewhat higher than what we see in the data.

5.3 Search Costs of Outsourcing Firms

Another approach to gaining insight into the factors linked to greater concentration involves re-writing the elasticity of the probability with respect to Ω in terms of shares and basic elasticities. Using (32), we get:

$$\frac{d\rho}{d\Omega}\frac{\Omega}{\rho} = \frac{w^O\Omega}{w^O\Omega + w^N\overline{N}}\frac{w^O/w^F}{w^O/w^F - 1}\left[1 + \frac{1}{\epsilon}\left(w^O/w^N\right)^{\epsilon-1}\right] \tag{46}$$

Next, if we assume that native and foreign labor is relatively substitutable[16] so that ϵ is large and $\frac{1}{\epsilon}\left(w^O/w^N\right)^{\epsilon-1} \approx 0$, we can simplify this to:

$$\frac{d\rho}{d\Omega}\frac{\Omega}{\rho} \approx \frac{w^O\Omega}{w^O\Omega + w^N\overline{N}}\frac{w^O/w^F}{w^O/w^F - 1} \tag{47}$$

The first term is the share of wages paid to H-1B workers in the sectors of the economy that employ them. The higher this value, the larger the probability elasticity. As H-1B workers become more important to the production process, firms become more responsive to changes in the H-1B quota and are more likely to respond by aggressively changing their number of excess H-1B filings. The second term is a transformation of the outsourcing premium w^O/w^F. The probability elasticity decreases as reliance upon outside labor becomes more expensive and this wage gap rises. At its limit, the second term in (47) approaches one

[16] The smallest estimate of ϵ in the literature that we are aware of comes from Manacorda, Manning, and Wadsworth (2012), which produces a minimum value of $\epsilon = 4.6$ and a baseline value of $\epsilon = 7.8$. Ottaviano and Peri (2012) prefer a value of $\epsilon = 20$.

so that the win probability elasticity with respect to the quota is governed solely by H-1B workers' share of income.

If we could measure these values, we could then back out an estimate for our key probability elasticity. Unfortunately, however, those values – like the measures described above – are unknowable. For example, we cannot observe the wages of H-1B workers in representative datasets, and given the small size of the H-1B program relative to the US economy, we believe it would be a mistake to insert a literal measure of total H-1B expenditures as a share of the national wage bill into equation (47). Similarly, the ratio $\frac{w^O}{w^F}$ is conceptual and reflects that gap between what employers earn from renting out their workers and what they actually pay their H-1B workers. This is unobservable.

Despite these limitations, we can turn to proxies in the data for insight. First, we note the USCIS (2020a) evidence that H-1B workers are highly concentrated among computer-related occupations. Further, Mayda et al. (2018) report that 91% of new H-1B recipients are under the age of 40. Motivated by these stylized facts, we turn to the 2008 and 2009 American Community Surveys (ACSs) and calculate that foreign-born labor accounted for 40% of wages paid to college-educated employees under age 40 in computer-related occupations. We then substitute this value for our first term in (47).

Second, let us suppose that the conceptual wage gap discussed above is similar to the gap in wages paid by regular firms and those paid by low-wage H-1B employers. Consider different behaviors in the H-1B petitioning process. Wage offers exhibit a well-populated mode at \$60,000 per year. This reflects legislation passed in the H-1B Visa Reform Act of 2004 that exempts H-1B dependent employers from additional legal obligations and government scrutiny if they only employ workers at a wage of \$60,000 or higher. We allow this value to serve as a *de facto* minimum wage and substitute it for w^F. For w^0, we instead use the average wage offer to new H-1B workers receiving more than \$60,000 in fiscal years 2008 and 2009. After removing extreme outliers, this amounts to \$83,500 and implies an outsourcing premium of $w^O/w^F \approx 1.39$.[17]

After substituting this into the second term of (47), we get an estimate of the probability elasticity with respect to the quota, ψ, of 1.43. This is consistent with the lower range of our elasticity estimates in Table 1 calculated by exploiting differences in firm H-1B filing behavior over time. Most importantly, it still delivers an elasticity greater than one. The condition required for increased rather than decreased concentration as a result of the H-1B lottery is easily satisfied through this type of quantification as well.

[17] Note also that in his February 2016 testimony to the US Senate, Hira (2016) argued that firms heavily relying upon H-1B workers generate profit margins of 20-25%, figures broadly consistent with the premium we find here.

Finally, we can conduct one further back-of-the-envelope calculation to estimate the implied bid and search costs for an outsourcing firm. Specifically, if we assume a lottery success rate of 0.50 (fairly consistent with the lottery in 2009) and a $w^O - w^F$ wage gap of \$23,500, equation (24) implies that $c_X + c = \$11,750$. If we take a literal interpretation of application cost and we assume a \$4,500 cost near the upper range of the H-1B filing fees discussed in Section 2, it implies that an outsourcing firm's marginal cost of searching for and extending an offer to a worker, c_X, equals roughly \$7,250.

5.4 Search Cost Externalities

Subsection 4.1 noted that the sum of regular firms' H-1B petitions is independent of the quota and that (34) describes how the quota affects H-1B concentration among outsourcing firms. By combining these insights with the equilibrium condition in (30), we see that the H-1B quota affects the total costs of H-1B searches according to (48):

$$-\frac{dC}{d\ln(\Omega)} = \left(w^O - w^F\right)\Omega\left(\psi - 1\right) \tag{48}$$

We have discussed above how to calculate proxies for many of the terms in this equation. For example, data suggests a wage gap value of $\left(w^O - w^F\right) = \$23,500$ and quota $\Omega = 85,000$. Total costs are then a function of the win probability elasticity with respect to the quota, ψ. This in turn depends upon the method used to compute ψ outlined in Section 5.1. Interestingly, the value of win probability itself plays no role in identifying these costs when using Methods 1 and 3. Not surprisingly, all three find that excess filings increase search externalities.

Consider Figure 6. Panels display estimated search costs in millions of dollars as a function of the percentage of excess filings (ϕ). Panels are distinguished by their use of Methods 1 through 3 (from left to right) for calculating the elasticity ψ. Graphs also include point estimates for search costs implied by observed data for each fiscal year in which a lottery allocated all cap-subject H-1Bs. Table 1 additionally provides a subset of these calculations. Recall that Mayda et al. (2008) estimate that actual H-1B demand was 22% to 33% above the quota for fiscal years 2008 and 2009. Red dots correspond to estimates using the 22% figure; blue dots assume the 33% value. Note that since data beyond 2009 were not included in the Mayda et al. (2008) analysis, the point values in 2014 and beyond represent out-of-sample estimates.

Interestingly, results are remarkably similar across the three methods for estimating ψ. The higher that real demand is above the quota, the smaller that search cost externalities will be. We will focus on these more conservative cost estimates (in blue). They imply that even when the excess filings were a modest 8.4% of all H-1B petitions submitted for fiscal year 2008, firms' strategic behavior resulted in at least a \$169 million search cost externality.[18] Costs were substantially higher in 2009 when nearly 31% of firms' H-1B submissions were filed in excess of demand: The externality estimate ranges between \$612 million and \$923 million for that year. If real H-1B demand remained at a constant 33% above the quota for years in the late 2010s, then estimated annual search costs exceed \$1 billion in that period. We believe that estimates of this magnitude are quite plausible. In 2017 when losses were maximized, for example, our estimates imply a \$1.042 billion to \$1.791 billion externality (depending upon the method used to calculate ψ). These calculations amount to a \$4,400 to \$7,575 cost per H-1B petition submitted. Given that firms must actually search for these workers and file paperwork with the government on their behalf, the estimate seems quite reasonable.

Importantly, these costs arise purely because firms search for an excess number of eligible H-1B workers to meet their target hiring goal. They are not costs associated with losing workers that firms actually wish to hire. This can be seen, for example, by the absence of an externality cost at a value of $\phi = 0$. If all H-1B petitions submitted by firms are for workers those firms intend to hire, then externalities associated with attempts to win the lottery do not exist. The data, however, strongly suggests that firms engage in strategic behavior, resulting in large costs to society.

6 Conclusion

The H-1B program allows highly-educated foreign-born individuals to temporarily work in the United States. The government has imposed a limit to the number of new H-1Bs that can be distributed each year. Demand for the program is so high that the government has decided to allocate H-1B status through a random lottery. H-1B petitions peaked at over 236,000 in 2017 when only about a third of petitioners won the lottery.

Past research has estimated several losses induced by the H-1B quota and lottery distribution. By restricting the skilled labor force, the quota reduces US productivity and wages paid to American-born workers. Firm outcomes such as sales and profits have declined as well. H-1B employment is increasingly concentrated among a smaller set of firms. Since firms receive permission to hire only those who win the lottery, lottery allocation prevents employers from selecting the foreign workers whom they most desire. Some firms have responded by moving operations overseas.

[18]Since currency figures are derived from nominal wage premia from 2008 and 2009, it would be reasonable to interpret these outcomes as reflecting losses in real 2008 or 2009 dollars.

This paper adds to the existing evidence by modeling heretofore undocumented costs induced by the quota and lottery: Firms waste resources searching for workers who cannot legally be hired. Indeed, firms search for far more workers (and submit H-1B petitions on their behalf) than they truly intend to hire. They do this knowingly in an attempt to better their chances of securing enough workers who will win the lottery and satisfy actual H-1B demand. Moreover, the lottery has given rise to outsourcing firms that specialize in hiring H-1B labor for the purpose of renting their employees to firms seeking the specialized services that H-1B workers can provide. This exacerbates negative search cost externalities. We estimate that this externality cost American firms $169 million in search costs in 2008 and more than $600 million in 2009. Costs are even higher when the probability of winning the lottery is lower (and excess H-1B petitions are higher), which has likely been the case in the late 2010s.

References

[1] Anderson, Stuart (2020). "USCIS Rescinds Policy Memos That Increased H-1B Visa Denials." *Forbes*. June 22, 2020. Retrieved at https://www.forbes.com/sites/stuartanderson/2020/06/22/uscis-rescinds-policy-memos-that-increased-h-1b-visa-denials/#669df4d827eb

[2] Bernstein, Shai, Rebecca Diamond, Timothy James McQuade, and Beatriz Pousada (2012). "The Contribution of High-Skilled Immigrants to Innovation in the United States." Stanford Graduate School of Business Working Paper No. 3748.

[3] Borjas, George J. and Kirk B. Doran, (2012). "The Collapse of the Soviet Union and the Productivity of American Mathematicians." *The Quarterly Journal of Economics*, Vol. 127(3): 1143-1203.

[4] Bound, John, Breno Braga, Joseph M. Golden, and Gaurav Khanna (2015). "Recruitment of Foreigners in the Market for Computer Scientists in the United States." *Journal of Labor Economics.* Vol. 33(S1): S187-223.

[5] Costa, Daniel and Ron Hira (2020). "H-1B Visas and Prevailing Wage Levels." Economic Policy Institute Report. May 4, 2020. Retrieved at https://files.epi.org/pdf/186895.pdf

[6] DHS (2011). "Registration Requirement for Petitioners Seeking to File H-1B Petitions on Behalf of Aliens Subject to the Numerical Limitations." *Federal Register*, Vol. 76(42): 11686-11699. March 3, 2011.

[7] DHS (2018). "Registration Requirement for Petitioners Seeking To File H–1B Petitions on Behalf of Cap-Subject Aliens." *Federal Register*, Vol. 83(232): 62406-62447. December 3, 2018.

[8] Depew, Briggs, Peter Norlander, and Todd A. Sorensen (2017). "Inter-firm Mobility and Return Migration Patterns of Skilled Guest Workers." *Journal of Population Economics*. Vol 30: 681-721.

[9] Glennon, Britta (2020). "How do Restrictions on High-Skilled Immigration Affect Offshoring? Evidence from the H-1B Program." Mimeo. University of Pennsylvania. http://dx.doi.org/10.2139/ssrn.3547655

[10] Gunadi, Christian (2019). "An Inquiry on the Impact of Highly-Skilled STEM Immigration on the U.S. Economy." *Labour Economics*. Vol. 61.

[11] Hira, Ronil (2016). Testimony on the impact of high-skilled immigration on US workers. US Senate Subcommittee on Immigration and the National Interest . February 25, 2016

[12] Hunt, Jennifer. (2011). "Which Immigrants are Most Innovative and Entrepreneurial? Distinctions by Entry Visa." *Journal of Labor Economics*. Vol. 29(3): 417-457.

[13] Hunt, Jennifer and Marjolaine Gauthier-Loiselle. (2010). "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics*. Vol. 2(2): 31-56.

[14] Hunt, Jennifer and Bin Xie (2019). "How Restricted is the Job Mobility of Skilled Temporary Work Visa Holders?" *Journal of Policy Analysis and Management*. Vol. 38(1): 41-64.

[15] Ilienko, Andrii. (2013). "Continuous Counterparts of Poisson and Binomial Distributions and their Properties." *arXiv preprint arXiv:1303.5990.*

[16] Kerr, Sari Pekkala, William Kerr, and William F. Lincoln (2015). "Skilled Immigration and the Employment Structures of US Firms." *Journal of Labor Economics*. Vol. 33(S1): S147-S186.

[17] Kerr, William and William F. Lincoln. (2010) "The Supply Side of Innovation: H-1B permit Reforms and U.S. Ethnic Invention." *Journal of Labor Economics* 28(3): 2010, pp. 473-508.

[18] Ma, Jie. (2020). "High Skilled Immigration and the Market for Skilled Labor: The Role of Occupational Choice." *Labour Economics*. Vol. 63(C).

[19] Manacorda, Marco, Alan Manning, and Jonathan Wadsworth (2012). "The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain." *Journal of the European Economic Association*, Vol. 10(1): 120-51.

[20] Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber (2018). "The Effect of the H-1B Quota on Employment and Selection of Foreign-Born Labor." *European Economic Review*, Vol. 108(C): 105-128.

[21] Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber (2020). "Coping with H-1B Shortages: Firm Performance and Mitigation Strategies. "NBER Working Paper 27730.

[22] Moser, Petra, Alessandra Voena, and Fabian Waldinger. (2014). "German Jewish Émigrés and US Invention." *American Economic Review*, Vol. 104(10): 3222-55.

[23] National Academies of Sciences, Engineering, and Medicine 2017. The Economic and Fiscal Consequences of Immigration. Washington, DC: The National Academies Press. https://doi.org/10.17226/23550.

[24] National Foundation for American Policy (2020). "H-1B Denial Rates and Numerical Limits as Indicators of Current Restrictions." NFAP Policy Brief, May 2020. Retrieved at https://nfap.com/wp-content/uploads/2020/05/H-1B-Denial-Rates-and-Numerical-Limits-as-Indicators-of-Current-Restrictions.NFAP-Policy-Brief.May-2020-1.pdf

[25] National Law Review (2020a). "Rescission of USCIS Policy Memoranda in Favor of H-1B Petitioners in the IT Third Party Placement Industry." Vol. X (188). June 22, 2020.

[26] National Law Review (2020b). "Recent USCIS Settlement Offers Substantial Relief to H-1B Employers." Vol. X (188). July 6, 2020.

[27] North, David (2020). "The Government Announces Stricter H-1B Provisions for Outsourcing Firms."Press Release. Center for Immigration Studies. October 6, 2020.

[28] Ottaviano, Gianmarco I. P. and Giovanni Peri (2012). "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association*, Vol. 10(1): 152–97.

[29] Office of the Federal Register, (2011). *Federal Register* Vol. 76, No. 42: 11667-11936.

[30] Park, Haeyoun, (2015). "How Outsourcing Companies Are Gaming the Visa System." *New York Times*, November 10, 2015.

[31] Peri Giovanni, Kevin Shih and Chad Sparber, (2015a). "STEM Workers, H-1B permits and Productivity in US Cities." *Journal of Labor Economics*, Vol. 33, no. S1 Part 2 (2015): S225-S255.

[32] Peri Giovanni, Kevin Shih and Chad Sparber, (2015b). "Foreign and Native Skilled Workers: What Can We Learn from H-1B Lotteries?" NBER Working Paper 21175.

[33] Sparber, Chad (2018). "Choosing Skilled Foreign-Born Workers: Evaluating Alternative Methods for Allocating H-1B Work Permits." *Industrial Relations: A Journal of Economy and Society*, Vol. 57 (1): 3-34.

[34] Sparber, Chad (2019). "Substitution between Groups of Highly-Educated, Foreign-Born, H-1B Workers." *Labour Economics*, Vol. 61.

[35] Turner, Patrick, (2020). "High-Skilled Immigration and the Labor Market: Evidence from the H-1B Visa Program." Mimeo. University of Notre Dame.

[36] USCIS (2007). "USCIS Runs Random Selection Process for H-1Bs." USCIS Update issued April 13, 2007.

[37] USCIS (2018a). "Contracts and Itineraries Requirements for H-1B Petitions Involving Third-Party Worksites." Policy Memorandum PM-602-0157. Issued February 22, 2018.

[38] USCIS (2018b). "USCIS Completes the H-1B Cap Random Selection Process for FY 2019." Retrieved at https://www.uscis.gov/news/alerts/uscis-completes-h-1b-cap-random-selection-process-fy-2019

[39] USCIS (2019). "USCIS Completes the H-1B Cap Random Selection Process for FY 2020 and Reaches the Advanced Degree Exemption Cap." Retrieved at https://www.uscis.gov/news/alerts/uscis-completes-h-1b-cap-random-selection-process-fy-2020-and-reaches-advanced-degree-exemption-cap

[40] USCIS (2020a). "Characteristics of H-1B Specialty Occupation Workers; Fiscal Year 2019 Annual Report to Congress." Retrieved at https://www.uscis.gov/sites/default/files/reports-studies/Characteristics_of_Specialty_Occupation_Workers_H-1B_Fiscal_Year_2019.pdf

[41] USCIS (2020b). "Rescission of Policy Memoranda." Policy Memorandum PM-602-0114. Issued June 17, 2020.

[42] USCIS (2020c). "FY 2021 H-1B Cap Petitions May bBe Filed as of April 1." Release Date: April 1, 2020. Retrieved at https://www.uscis.gov/news/news-releases/fy-2021-h-1b-cap-petitions-may-be-filed-april-1.

A Partial Equilibrium Setup

This appendix provides more detail as to the underlying assumptions on preferences that would rationalize the partial equilibrium model in the main text. We assume that the representative consumer has the following quasi-linear utility over the aggregate H-1B sector good, X, and a numeraire good that represents all other goods, m:

$$U = m + \gamma \times log\left(X\right), \tag{49}$$

where γ is constant and where:

$$X = \left[\sum_{i} \left(\alpha_i\right)^{\frac{1}{\sigma}} \left(x_i\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{50}$$

The household is subject to a budget constraint:

$$m + PX = I + \Pi + w_N \overline{N} + w_F \Omega, \tag{51}$$

where I is total income excluding the H-1B sector and is exogenous given the partial equilibrium setting; Π are the total profits of the H-1B sector; and P is the price index for the aggregate H-1B sector output:

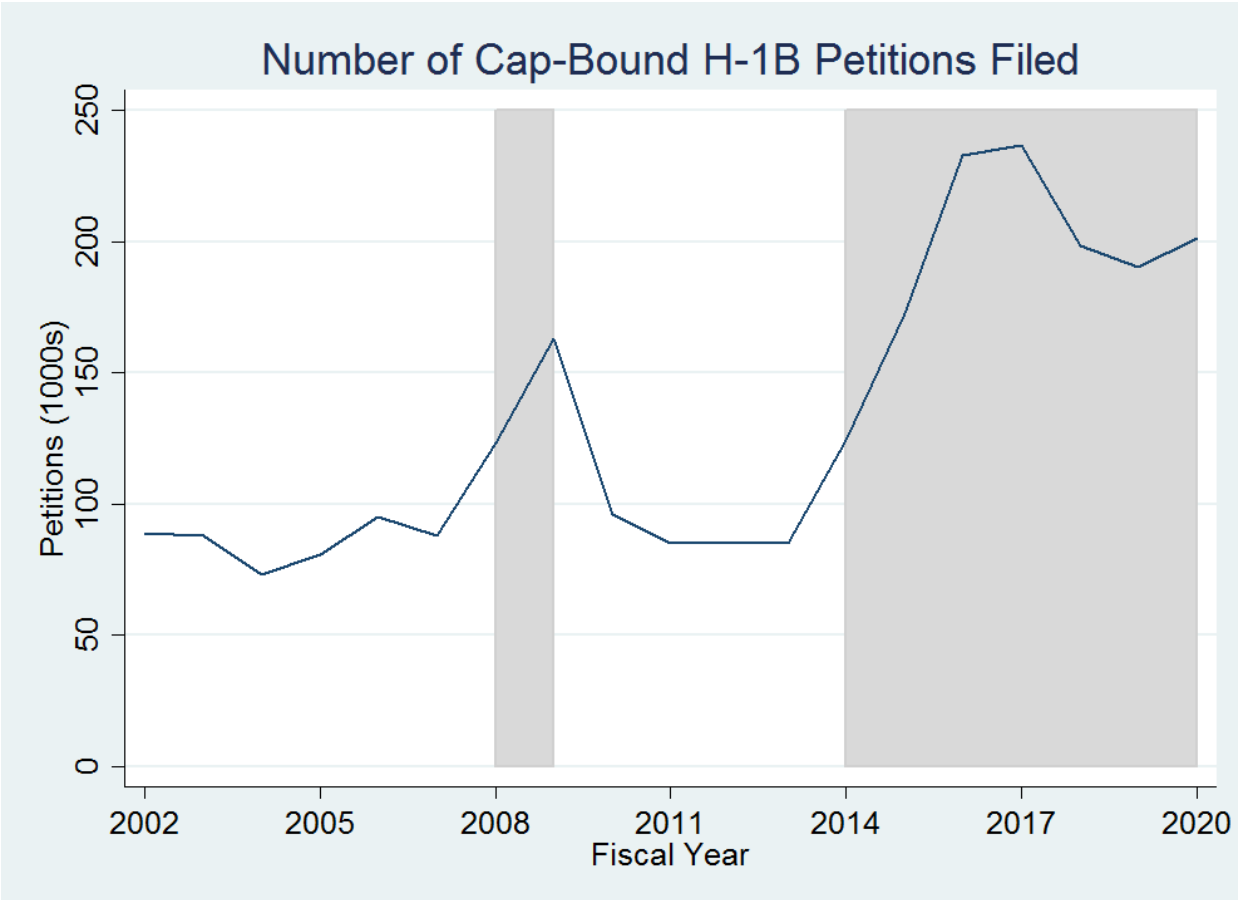
$$P = \left[\sum_{i} \alpha_i \left(p_i\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{52}$$

Note that while we include both native and immigrant H-1B wages in (51) – taking the representative household to include immigrant workers – the analysis would be basically identical if we only focused on native households. Maximizing (49) subject to (31) gives us:

$$PX = \gamma \tag{53}$$

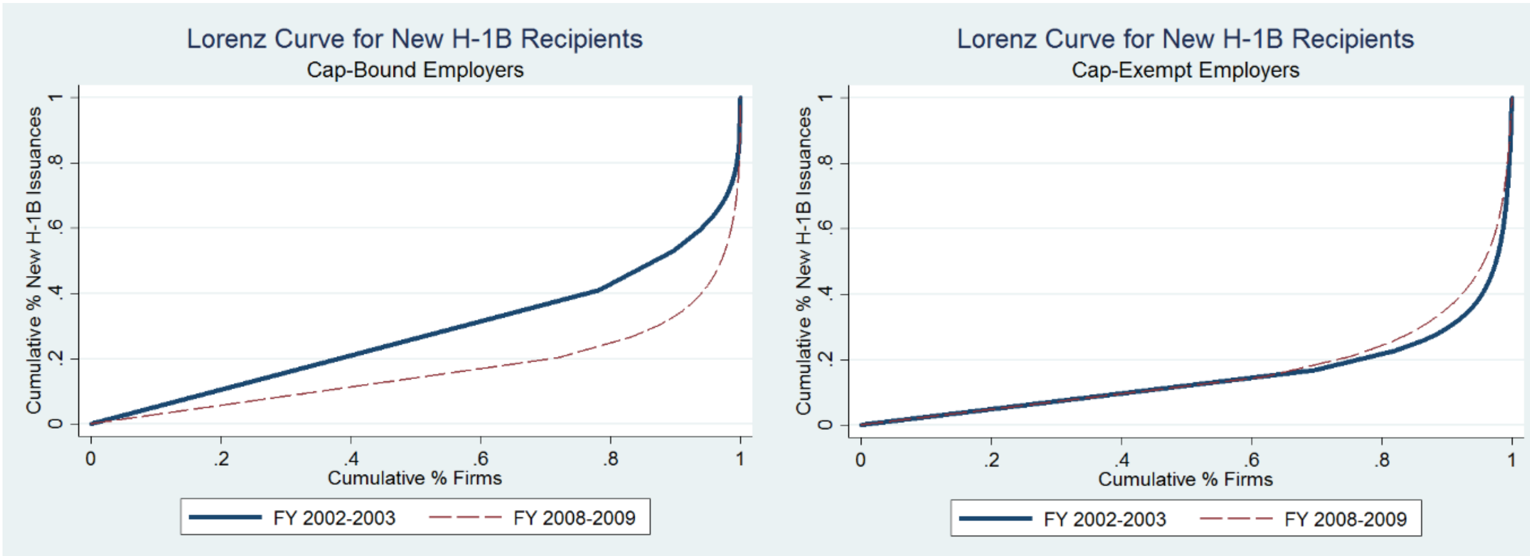
This rationalizes our constant E assumption since $PX = E$. The constant E here depends on the assumption of log substitutability. This constant E setup is helpful for expositional clarity and simplicity, but could easily be relaxed without changing the thrust of our overall analysis. Finally, we could find the optimal choice of each variety x_i within (50), and this would give us the standard CES demand in (1).

Figure 1



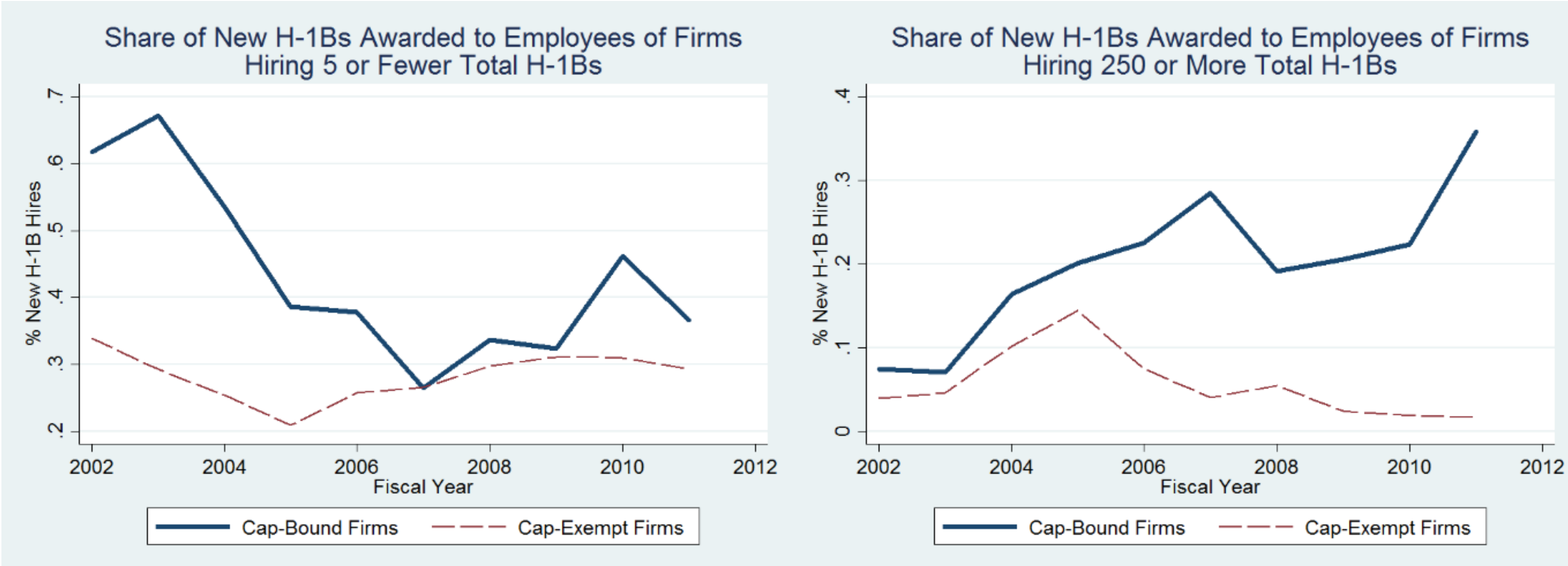
This figure displays the number of I-129 petitions filed for cap-bound H-1B workers in fiscal years 2002-2020. Years in which USCIS allocated all cap-bound H-1Bs by random lottery are shaded. Sources include USCIS (2007, 2018b, 2019), DHS (2011, 2018), and Office of the Federal Register (2011). Figures in FY 2011-2013 reflect the caps in those years that were reached several months after USCIS started accepting petitions.

Figure 2



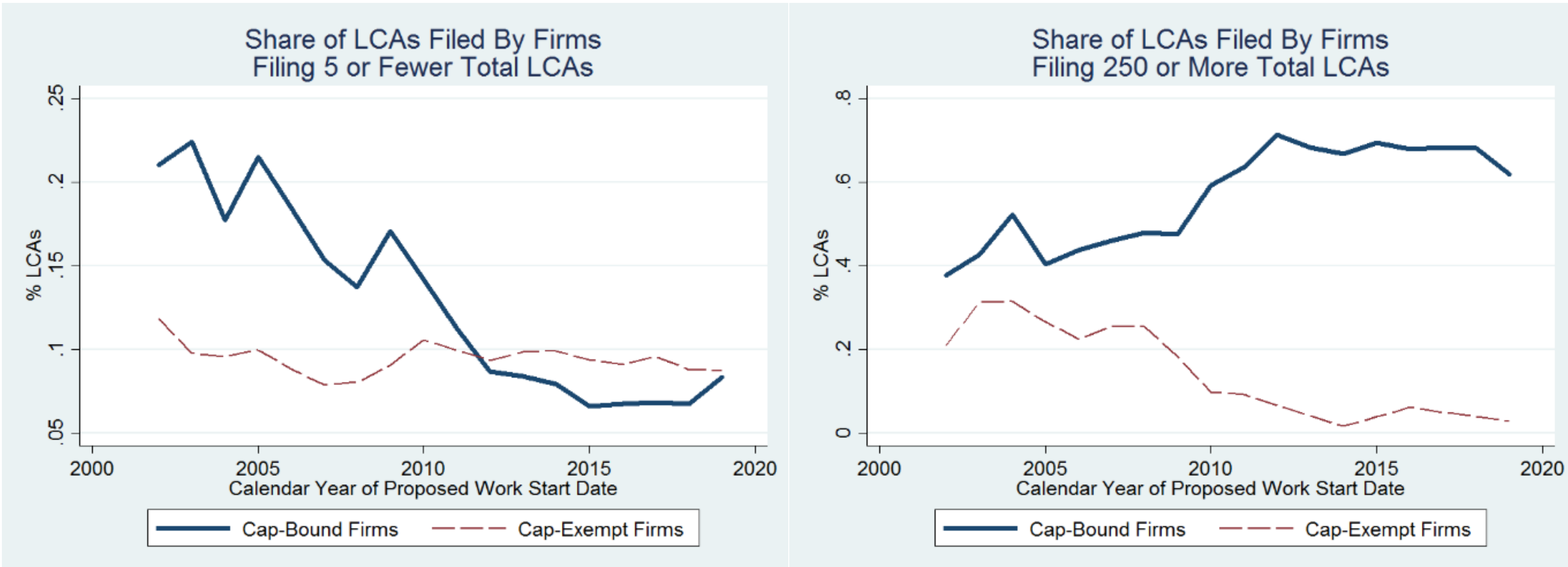
This figure displays the Lorenz Curves illustrating the concentration of new H-1B hires at cap-bound (left panel) and cap-exempt (right panel) firms based upon I-129 petitions.

Figure 3



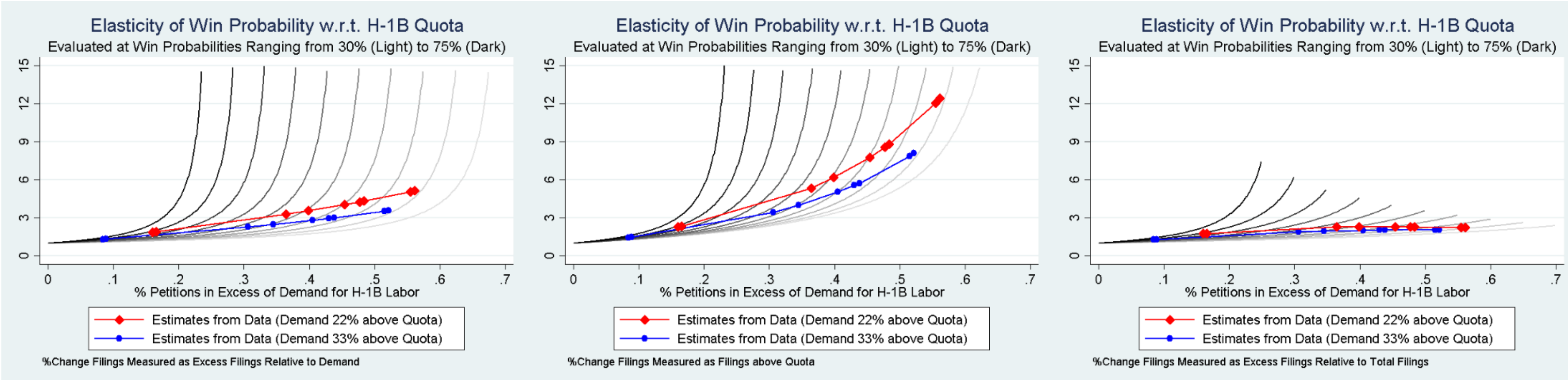
This figure displays the percentage of new H-1B workers hired at firms employing five or fewer total H-1B workers (left panel) and 250 or more total H-1B workers (right panel) for a given fiscal year. Data is based upon I-129 petitions.

Figure 4



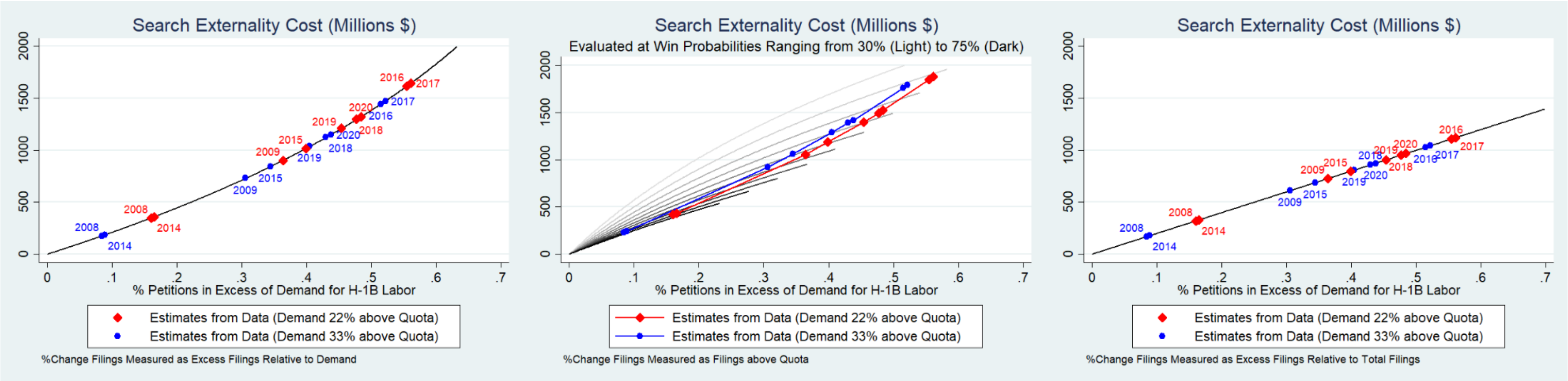
This figure displays the percentage of new H-1B workers hired at firms employing five or fewer total H-1B workers (left panel) and 250 or more total H-1B workers (right panel) for a given fiscal year. Data is based upon LCA filings.

Figure 5



This figure displays estimates for the elasticity of the H-1B win probability with respect to the quota (ψ) as a function of the win probability (ρ) and petitions filed in excess of demand for H-1B labor (ϕ). Plausible values of ϕ range from zero to $1 - \rho$. Panels differ in the methods used to compute the percentage change in filings due to the quota.

Figure 6



This figure displays estimates for the search cost externality generated by the H-1B quota as a function of the H-1B win probability (ρ, middle panel only) and petitions filed in excess of demand for H-1B labor (ϕ). Plausible values of ϕ range from zero to $1 - \rho$. Panels differ in the methods used to compute the percentage change in filings due to the quota.

Table 1

Year of Observation	2008	2009	2017	2008	2009	2017
Applications	123480	163000	236444	123480	163000	236444
Win Probability (ρ)	0.69	0.52	0.36	0.69	0.52	0.36
Demand Shock	0.33	0.33	0.33	0.22	0.22	0.22
Real Demand	113050	113050	113050	103700	103700	103700
Excess Filings	10430	49950	123394	19780	59300	132744
% Excess Filings (ϕ)	0.084	0.306	0.522	0.160	0.364	0.561
	Δln(Filings) Measured as Excess Filings Relative to Demand					
Δln(Filings)	0.09	0.37	0.74	0.17	0.45	0.82
Δln(Ω)	-0.29	-0.29	-0.29	-0.20	-0.20	-0.20
Elasticity (ψ)	1.31	2.28	3.59	1.88	3.27	5.14
$\Delta\theta$, at θ=1/3	0.21	0.86	1.72	0.59	1.52	2.76
$\Delta\theta$, at θ=3/5	0.12	0.51	1.03	0.35	0.91	1.66
Search Cost Externality	176	731	1474	349	903	1646
	Δln(Filings) Measured as Filings above Quota					
Δln(Filings)	0.12	0.46	0.90	0.21	0.53	0.94
Δln(Ω)	-0.26	-0.19	-0.13	-0.16	-0.12	-0.08
Elasticity (ψ)	1.45	3.45	8.10	2.27	5.34	12.42
$\Delta\theta$, at θ=1/3	0.30	1.63	4.73	0.85	2.90	7.61
$\Delta\theta$, at θ=3/5	0.18	0.98	2.84	0.51	1.74	4.57
Search Cost Externality	231	923	1791	418	1057	1879
	Δln(Filings) Measured as Excess Filings Relative to Total					
Δln(Filings)	0.08	0.31	0.52	0.16	0.36	0.56
Δln(Ω)	-0.29	-0.34	-0.50	-0.21	-0.29	-0.46
Elasticity (ψ)	1.29	1.89	2.04	1.75	2.27	2.22
$\Delta\theta$, at θ=1/3	0.19	0.59	0.69	0.50	0.84	0.81
$\Delta\theta$, at θ=3/5	0.12	0.36	0.42	0.30	0.51	0.49
Search Cost Externality	169	612	1042	320	727	1121