Worker Heterogeneity and Immigration Policy Inefficiencies

Caitlin Hegarty Williams College

Mishita Mehra University of Richmond

December 1, 2024

PRELIMINARY AND INCOMPLETE

Abstract

Many U.S. firms seek to employ foreign high-skilled workers when they cannot meet their skill demands with domestic labor alone. The H-1B visa program, which facilitates these hires, faces an annual cap on the number of visas granted and allocates them via randomization. This creates an additional friction in the labor market when firms face applicants of heterogeneous quality, as they cannot always hire the workers with the highest productivity. This paper introduces a general equilibrium search model to quantify these inefficiencies and evaluate alternative policy designs. The model incorporates realistic features of the H-1B visa program, with spillovers between domestic and foreign high-skilled labor and low-skilled labor. Counterfactual policy analysis reveals that increasing the visa cap and modifying the allocation of the the existing visa cap are both policies that would both improve aggregate productivity and increase domestic employment.

JEL: J61, F22, E24

Keywords: High-skilled immigration, immigration policy, search and matching, worker heterogeneity

1. Introduction

In high-skilled industries, acquisition of talented workers is crucial for firms' success. Given the importance of worker skill, firms conduct extensive screening to choose the most promising candidates. In the US, foreign skilled workers are often a vital source of highly-specialized skills in this sector, and yet firms are limited in their ability to hire these workers due to an entry quota and hiring costs. In the absence of policy frictions, firms would interview the most promising candidates of all nationalities and then accept those with the highest expected benefit to the firm. Under the restrictions of the H-1B visa program, firms may choose to hire foreign skilled workers who have high expected productivity, but they will only be allowed to hire the workers if they receive a successful visa petition.

This process leads to two sources of inefficiency. First, the uncertainty around foreign workers' visa petition and the costliness of filing it imply that the marginal domestic hire would have a lower expected productivity than the marginal foreign hire. In the absence of these frictions, the marginal productivities should be equal. Second, the visa process grants petitions on a worker-level basis rather than at the firm level. This means that the firm's most-preferred candidate has the same probability of being accepted as the firm's least-preferred candidate. In this paper, we quantify these inefficiencies using a heterogeneous-worker model.

We start with a canonical random search model and introduce (i) heterogeneity in match quality (productivity) among high-skilled workers, (ii) differences in hiring costs between foreign and domestic workers, (iii) a binding cap on the number of foreign visas granted, and (iv) complementarities in production between high-skilled and low-skilled labor. The first ingredient is necessary for generating a meaningful hiring decision with an endogenous hiring threshold that varies with aggregate conditions and the degree of frictions in the economy. The second ingredient mimics the tangible costs that firms pay to hire foreign workers, such as application fees and legal fees. The third ingredient captures the realistic uncertainty around whether a foreign worker's visa application will be accepted. The final ingredient allows us to capture spillover effects between different sectors of the labor force.

Our model produces several predictions. First, due to the application cost and uncertainty around application success, firms will have a higher productivity threshold for entering the lottery for a foreign worker than they do for hiring a domestic worker. This means the average foreign hire is more productive than the average domestic hire. Second, the degree of heterogeneity in match quality for high-skilled workers affects the composition of the workforce. More heterogeneity leads firms to be more selective in hiring, increasing average productivity. This negatively affects domestic high-skilled workers, because firms are more willing to enter the H-1B lottery when the productivity gain is greater, leading to substitution between domestic and foreign high-skilled workers. It positively affects low-skilled workers, because higher average productivity among the high-skilled sector increases demand for low-skilled labor due to complementarities in production.

We calibrate the model to match moments from the data. Then, we use the model as a laboratory to explore the policy implications of modifying the visa program by either increasing the number of visas or reallocating the existing cap on visas. We find that increasing the cap on visas has positive effects on aggregate productivity and positive effects on employment for both domestic high-skilled and low-skilled labor. Because foreign workers are on average more productive due to hiring frictions, increasing the cap by a small amount increases the average productivity of the high-skilled workforce as a whole due to composition shifting towards more foreign labor. This leads the firm to become slightly more selective about hiring domestic high-skilled labor, but also benefits domestic high-skilled workers because more foreign hiring makes the pool of high-skilled labor less congested, leading to more matches for domestic high-skilled workers. Low-skilled workers benefit as well due to the increase in average productivity.

We find that reallocating visas to the most productive workers has massive effects on productivity in general equilibrium. In our preliminary calibration, reallocating visas increases aggregate productivity by 12%, with an 80% increase in the average productivity of foreign workers. These large effects are coming from two sources. First, there are the direct effects of granting the visas to more productive hires rather than randomly allocating them. In general equilibrium, firms also internalize this rule and will only submit applications for

matches that exceed the minimum productivity criteria in equilibrium. Thus, they adjust their hiring strategy to post more vacancies to seek out the highest quality matches and no longer submit inefficient applications. This drives up the equilibrium productivity cutoff, driving the very large increase in average foreign worker productivity. Domestic high-skilled workers benefit marginally from the increase in vacancies and employment is essentially unchanged or slightly increased. Low-skilled workers benefit much more from the big increase in average productivity, with employment increasing by 0.5 percentage point.

Our study adds to the extensive list of papers studying the general equilibrium impacts of U.S. skilled immigration, for instance, Bound *et al.* (2015); Waugh (2017); Bound *et al.* (2018); and Mehra & Shen (2022). To the best of our knowledge, this is the first quantitative general equilibrium model to assess the inefficiencies due to US skilled immigration policies in the presence of productivity heterogeneity. Sharma & Sparber (2024) provides a theoretical analysis of the negative externality created by the random allocation of visas, which is present in our model as well.

We also contribute to the set of papers employing multi-worker search models with worker heterogeneity, including Elsby *et al.* (2015), Baydur (2017), and Hegarty (2024). Our paper adds a new layer to these models by introducing a second type of labor with complementarities in the production function. We also adapt the model to the H-1B setting by introducing realistic frictions in hiring domestic or foreign high-skilled labor.

2. Institutional Background

The H-1B visa program has been the main method of entry into the US workforce for foreign college-educated professionals since 1990. The H-1B visa is temporary as it is issued for only three years (and can be renewed for another three), but it is a dual intent visa since it can lead to permanent residency if the employer is willing to sponsor the worker for a green card.

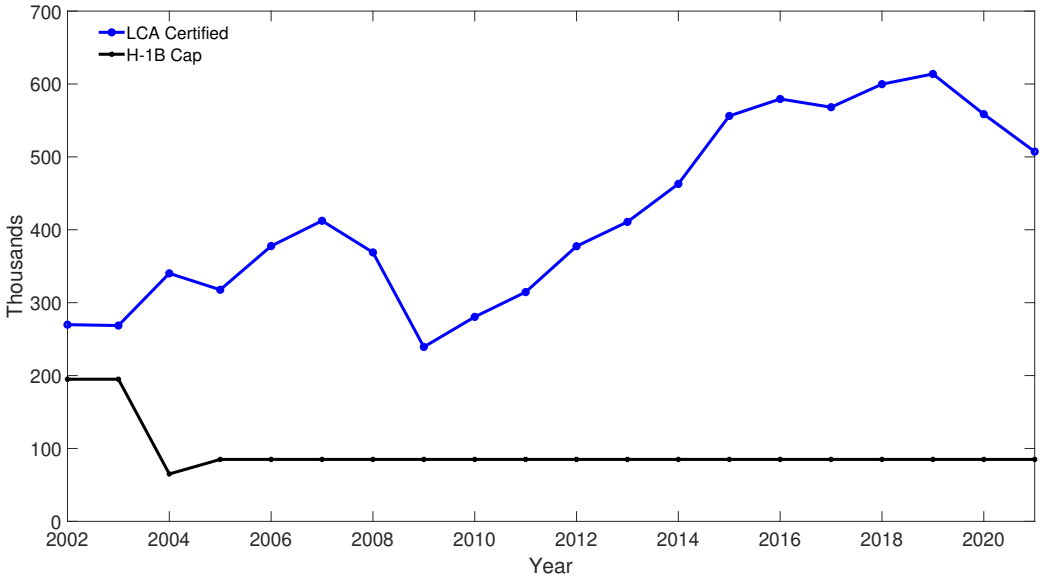
The H-1B program has been subject to an annual quota on new visa issuances. The initial visa cap was 65,000, which was subsequently increased to 115,000 in 1999 and 2000 after the cap was met in 1997. The cap was further increased to 195,000 from 2001 to 2003.

In 2001, cap exemptions were introduced for employees at higher education, nonprofit, and government research organizations. In 2004, the cap was reduced back to 65,000, but 20,000 additional visas were allocated for workers who had obtained a master's degree or higher from a US institution. The cap applies only to new H-1B visa issuances for for-profit firms.

To obtain an H-1B visa, there are several steps to be followed, and firms play a crucial role in this process. The first step requires the firm that wants to hire a foreign worker to file an LCA with the Department of Labor, which are processed relatively quickly. In the application, the firm specifies the nature of the worker's occupation and attests that it will pay them the greater of the actual compensation paid to other employees in the same job or the prevailing compensation for that occupation. The rationale given for this attestation is to help protect domestic worker wages. The LCA request also certifies that the employer must hire a foreign worker because a US citizen is not qualified, available, or willing to work in that job position. LCA forms can request one or more foreign workers for a particular occupation, and thus they signal firm vacancies or demand in specific occupations for foreign workers. An LCA is submitted for every H-1B request, whether new or a renewal, and each LCA can contain multiple H-1B workers. Figure 1 highlights the increasing gap between certified LCAs and the H1-B cap in recent years.

Once the Department of Labor approves/certifies the LCA, it is sent to the USCIS along with the I-129 form and the visa fees. The I-129 form is the formal H1-B petition which specifies the particular worker that a firm wants to hire and their qualifications. Firms have from April 1 until the beginning of the next fiscal year to file petitions for H-1B visa applications. Crucially, potential employees can only apply for an H-1B visa if they have a job offer from an employer with LCA approval. By law, the employer cannot file more than one I-129 for the same prospective employee. Most of the filing and legal fees are also borne by the employer. If the number of H-1B visa petitions that fall within the nonexempt category exceeds the cap, the USCIS randomly selects visas for processing via a lottery system until the 65,000 cap is reached. The total number of petitions filed does not indicate the true demand because the government stops collecting H-1B petitions once it has determined that the cap has been reached for a given year.

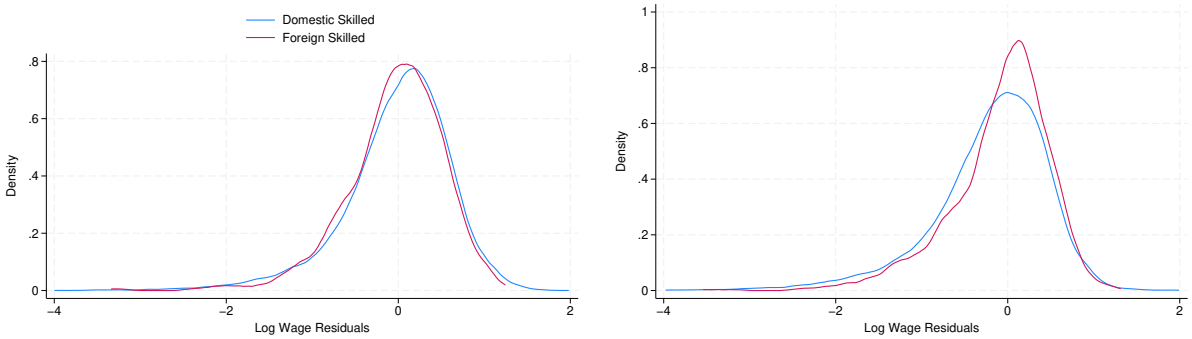
Figure 1: Demand for Foreign Workers (Labor Condition Applications) vs. H-1B Cap



In recent years, the Department of Homeland Security has slightly amended its regulations regarding the process by which the USCIS selects H-1B petitions for the filing of the H-1B cap-subject petitions. In 2020, the USCIS implemented a preregistration process that begins on March 1 for potential employees who want to file an H-1B petition. If the USCIS receives enough registrations by March 18 (based on historic projections), they will randomly select registrations. An H-1B cap-subject petition may only be filed by a petitioner whose registration was selected for an extensive analysis of the H-1B rule changes over the years). In very recent years there have also been increasing cases of fraud and multiple registrations submitted for the same employee. However, such cases are subject to legal actions.

Heterogeneity in the data To provide some motivation for the role of productivity heterogeneity, we show that there is less dispersion among the residual wages of potential H-1B workers. We regress log wages from the CPS on workers' gender, age, age-squared, education, industry, and state. Figure 2 shows the density of the residuals among domestic skilled (college-educated) workers relative to foreign. We limit the sample of foreign workers to non-citizens born in India or China because these workers account for more than 85% of

Figure 2: Heterogeneity in the data



(a) All occupations **(b)** H-1B dominant occupations

H-1B visas. Panel (a) shows the results for workers in all occupations. Panel (b) restricts the sample to the top two occupations for H-1B workers: computer-related and architecture, engineering, and surveying.

The figures show that potential H-1B workers have less dispersion in residual wages than domestic skilled workers. In our model, this will be consistent with firms hiring a more narrow set of foreign workers. Note also that these regressions do not include firm fixed effects and therefore should be interpreted as a combination of firm and worker wage residuals. We are choosing to model heterogeneity through firm-worker match quality, which is comparable to this interpretation.

3. Model

We start with a heterogeneous worker model with random search and a representative firm. The model could easily be adapted to incorporate firm heterogeneity as well. Firms produce using labor from low-skilled and high-skilled sectors. We assume that all low-skilled workers are domestic and homogeneous in their productivity.[1] High-skilled workers may be either domestic or foreign. Each time a high-skilled worker meets a firm, they receive a productivity draw that remains fixed throughout the duration of the match if it forms. We assume the

[1] Alternatively, all low-skilled workers are domestic or otherwise integrated into the domestic workforce through policies that are outside the scope of this paper.

masses of workers are fixed, with no switching between sectors and no changes in foreign migration.[2]

Matching process Firms post vacancies at cost $c_v(j)$ to attract matches in a labor market that is segmented by sector j with no on-the-job search. The market for high-skilled workers is pooled across foreign and domestic applicants. The probability that a vacancy attracts a worker depends on market tightness $\theta_j = V_j/U_j$, with $q(\theta_j)$ the probability that a vacancy attracts a match, and $\theta_j q(\theta_j)$ the probability that an unemployed worker matches to a vacancy. There is a unit mass of skilled domestic workers and a mass μ of potential foreign skilled workers. Let π be the mass of low-skilled workers. Domestic and foreign high-skilled workers are sampled proportional to their shares in the market. Let $u(d, H)$ and $u(f, H)$ be the number of domestic and foreign unemployed skilled workers. Then the probability that a high-skilled vacancy matches to a foreign worker is $\frac{u(f,H)}{u(d,H)+u(f,H)} q(\theta_H)$.

Because low-skilled workers are homogeneous, the firm chooses the number of low-skilled vacancies and hires all matches. For high-skilled workers, heterogeneous match quality means that the firm may screen matches on the basis of productivity. When a high-skilled worker meets a firm, they draw match productivity p from distribution $F(p)$. After observing this productivity, the firm decides whether or not to hire the worker. If the firm chooses to hire a domestic worker, they pay a hiring cost $c_h(d)$ and the worker is hired with certainty. If the firm chooses to hire a foreign worker, the firm faces two additional frictions. First, the firm must pay a steeper hiring cost, $c_h(f) > c_h(d)$, meant to capture the cost of filing a visa application. Second, the firm faces uncertainty about whether the visa application will be successful. The firm is allowed to hire the candidate with probability

$$x = \min\left\{\frac{\bar{N}}{N_F}, 1\right\}, \tag{1}$$

where $\bar{N}$ is the exogenous cap on the number of visas granted and N_F is the total number of applications for foreign hires submitted by firms in the economy.

[2]If the mass of foreign workers wishing to apply to jobs in the U.S. is sufficiently large, this assumption should not change our counterfactual results substantially.

If the firm successfully hires a skilled worker with match quality p, the worker contributes p units of effective labor and is paid a wage that depends on their productivity and may depend on the firm's state variables, such as size.

Firm's problem The firm starts the period with incumbent low-skilled workers and a distribution of incumbent high-skilled workers for each level of productivity, p, and worker nationality i. Let n_L be the number of low-skilled workers, $\mathbf{n_H} = \{n_H(i)\}$ be the total number of incumbent high-skilled workers of each nationality i and $\mathbf{G} = \{G(p,i)\}$ be the distributions of productivity p among incumbent high-skilled workers. The firm also observes a set of aggregate state variables, including market tightness, θ, aggregate productivity, a, and the number of domestic and foreign unemployed workers, $u_H(i)$. To ease notation, we use a subscript t to capture these aggregate states.

The firm chooses vacancies, v_j, and hiring rules, $h(p,i) \in \{0,1\}$, for potential hires with match quality p. These choices implicitly pin down the composition of employment at the production stage. To simplify the firm's problem, we let $p^*(i)$ be a cut-off such that firms hire all high-skilled workers of nationality i with a productivity $p^*(i)$ or higher. The firm's

value function can be written as follows,

$$J_t(n_L, \mathbf{n_H}, \mathbf{G}) = \max_{v_j, p^*(i)} -v_L c_{vL} - v_H c_{vH} - \sum_{i \in \{d,f\}} \underbrace{v_H q(\theta_{tH}) \frac{u_{tH}(i)}{u_{tH}} (1 - F(p^*(i)))}_{\text{number of hires, } i} c_h(i)$$

$$+ a_t {n'_L}^{\alpha_L} (\bar{p}' n'_H)^{\alpha_H} - n'_L w_{tL}(n'_L, \mathbf{n_H}', \mathbf{G}') - \sum_i \int_p w_{tH}(p, n'_L, \mathbf{n_H}', \mathbf{G}') n'_H(i) dG'(p, i)$$

$$+ \beta \mathbb{E} J_{t+1}(n'_L, \mathbf{n'_H}, \mathbf{G}') \tag{2}$$

s.t.

$$\bar{p}' n'_H = \underbrace{n'_H(d) \int_p p dG'(p, d) + n'_H(f) \int_p p dG'(p, f)}_{\text{productivity-adjusted employment}} \tag{3}$$

$$n'_L = (1 - \delta_L) n_L + v_L q(\theta_{tL}) \tag{4}$$

$$G'(p, i) = \frac{(1 - \delta_H) n_H(i)}{n'_H(i)} G(p, i) + \frac{(n'_H(i) - (1 - \delta_H) n_H(i))}{n'_H(i)} \frac{F(p) - F(p^*(i))}{1 - F(p^*(i))} \tag{5}$$

$$n'_H(d) = \underbrace{(1 - \delta_H) n_H(d)}_{\text{retained incumbents, } d} + \underbrace{v_H q(\theta_{tH}) \frac{u_{tH}(d)}{u_{tH}} (1 - F(p^*(d)))}_{\text{number of hires, } d} \tag{6}$$

$$n'_H(f) = \underbrace{(1 - \delta_H) n_H(f)}_{\text{retained incumbents, } f} + \underbrace{x_t v_H q(\theta_{tH}) \frac{u_{tH}(f)}{u_{tH}} (1 - F(p^*(f)))}_{\text{number of hires, } f} \tag{7}$$

where the first line captures the costs of posting vacancies and the cost of hiring all skilled workers with productivity above the hiring threshold, $p^*(i)$. $F(p)$ is the cumulative distribution of productivity for new matches. The second line captures production with decreasing returns to scale technology and complementarities between low- and high-skilled workers, accounting for productivity of high-skilled employment, as defined in equation (3). Equation (4) shows the law of motion for low-skilled employment, where workers separate at an exogenous rate δ_L. Because firms bargain with all workers, the wages of workers in each sector may depend on employment in the other sector. The distribution of productivity among skilled workers after hiring, $G'(p, i)$, is derived in equation (5) as a weighted average of the distribution of incumbent workers from the previous period, $G(p, i)$, and the distribution of new hires, which is a censored distribution of the match quality pool that depends on hiring

threshold $p^*(i)$.

Equations (6) and (7) represent the laws of motion for the number of skilled workers with nationality i. An exogenous share δ_H of incumbent workers separate at the start of the period. The key difference between the laws of motion for domestic and foreign employment is that foreign hires are granted with probability x_t. Thus, even if the firm sets the exact same threshold for both groups and groups face equal unemployment shares and productivity distributions, the firm will ultimately hire fewer foreign workers if the visa cap is binding, $x_t < 1$.

We assume firms are owned by workers with linear utility, so firms discount the future at the workers' discount rate β.

High-skilled worker's problem The value to a skilled worker of nationality i employed in a match with productivity p at a firm with size $n_L, \mathbf{n_H}$ and productivity distribution $\mathbf{G}$ is

$$\begin{aligned} W_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G}, i) = w_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G}) \\ + \beta \mathbb{E}_t \left[U_{t+1H}(i) + (1 - \delta_H) \left(W_{t+1H}(p, n'_L, \mathbf{n_H}', \mathbf{G}', i) - U_{t+1H}(i) \right) \right], \end{aligned} \tag{8}$$

where $U_{tH}(i)$ is the value of unemployment. This value can be written as follows,

$$U_{tH}(i) = b_H + \Omega_{tH}(i) + \beta \mathbb{E}_t U_{t+1H}(i), \tag{9}$$

where $\Omega_{tH}(i)$ represents the value the worker gets from being able to search for a new job in the following period. We define $\Omega_{tH}(i)$ as follows,

$$\Omega_{tH}(d) \equiv \beta \mathbb{E}_t \left[\theta_{t+1H} q(\theta_{t+1H}) \int_{p \geq p'^*(d)} \left(W_{t+1H}(p, n'_L, \mathbf{n_H}', \mathbf{G}', d) - U_{t+1H}(d) \right) dF(p) \right] \tag{10}$$

$$\Omega_{tH}(f) \equiv \beta E_t \left[\theta_{t+1H} q(\theta_{t+1H}) x_{t+1} \int_{p \geq p'^*(f)} \left(W_{t+1H}(p, n'_L, \mathbf{n_H}', \mathbf{G}', f) - U_{t+1H}(f) \right) dF(p) \right], \tag{11}$$

where the term inside the integral captures the expected value of a match next period, conditional on the firm's optimal threshold choice. The only difference in the value functions of domestic and foreign workers operates through this endogenous value of search, $\Omega_t(i)$, which will differ if the visa cap is binding, $x_{t+1} < 1$, and/or firms have different hiring thresholds, $p'^*(d) \neq p'^*(f)$.

Workers will accept an offer if $W_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G}, i) \geq U_{tH}(i)$.

Low-skilled worker's problem The low-skilled worker's problem is simpler because there is no heterogeneity in matches, though the value of employment still depends on the firm attributes. We can define value functions below

$$W_{tL}(n_L, \mathbf{n_H}, \mathbf{G}) = w_{tL}(n_L, \mathbf{n_H}, \mathbf{G}) + \beta \mathbb{E}_t \left[U_{t+1L} + (1 - \delta_L) \left(W_{t+1L}(n'_L, \mathbf{n_H}', \mathbf{G}') - U_{t+1L} \right) \right] \tag{12}$$

$$U_{tL} = b_L + \Omega_{tL} + \beta \mathbb{E}_t U_{t+1L} \tag{13}$$

$$\Omega_{tL} \equiv \beta \mathbb{E}_t \left[\theta_{t+1L} q(\theta_{t+1L}) (W_{t+1L}(n'_L, \mathbf{n'_H}, \mathbf{G}') - U_{t+1L}) \right], \tag{14}$$

where Ω_{tL} is defined as the value of search next period, to be symmetric with the high-skilled worker's problem.

Wage bargaining We assume that wages are determined via Stole & Zwiebel (1996) bargaining, where firms bargain separately with each sector, taking the labor from the other sector as given. For high-skilled workers, we assume that firms always assume that they are bargaining with a domestic worker. We believe this is a fair assumption because the firm is required to pay workers the maximum of the within-firm occupation wage and the prevailing market wage for that occupation. If there were another firm in the market that hired only domestic workers, this firm would underpay its foreign workers relative to the other firm if it does not treat them as domestic workers. We explain how the wage equation would differ if firms are allowed to pay foreign workers lower wages, which may happen in practice even though it is technically illegal.

Let n_H be the total number of high-skilled workers and G be the productivity distribution of high-skilled workers. A domestic high-skilled worker with productivity p enjoys a surplus of

$$W_{tH}(p, n_L, n_H, G, d) - U_{tH}(d) = w_{tH}(p, n_L, n_H, G) - (b_H + \Omega_{tH}(d)) + \beta(1-\delta_H)(W_{t+1H}(p, n'_L, n'_H, G', d) - U_{t+1H}(d)), \quad (15)$$

where $\Omega_{tH}(d)$ captures that if the worker walks away from the firm at bargaining, they will be eligible to search again next period, and we assume this value is the value anticipated by a domestic worker, since the firm assumes they are always bargaining with a domestic worker.

A low-skilled worker has a surplus of

$$W_{tL}(n_L, n_H, G) - U_{tL} = w_{tL}(n_L, n_H, G) - (b_L + \Omega_{tL}) + \beta(1-\delta_L)(W_{t+1L}(n'_L, n'_H, G') - U_{t+1L}), \quad (16)$$

with Ω_{tL} similarly capturing the value of future search.

The firm's total surplus at the time of bargaining is

$$D_t(n_L, n_H, G) = a_t n_L^{\alpha_L} \left(\int_p p n_H dG(p) \right)^{\alpha_H} - \int_p n_H w_{tH}(p, n_L, n_H, G) dG(p) - n_L w_{tL}(n_L, n_H, G) + \beta \mathbb{E}_t J_{t+1}(n'_L, n'_H, G'). \quad (17)$$

Let $n_H(p) = n_H \times g(p)$. The marginal surplus the firm enjoys from employing a high-skilled worker with productivity p is

$$\frac{\partial D_t(n_L, n_H, G)}{\partial n_H(p)} = p\alpha_H a_t n_L^{\alpha_L} \left(\int_p p n_H dG(p) \right)^{\alpha_H - 1} - w_{tH}(p, n_L, n_H, G) - \int_{\tilde{p}} n_H \frac{\partial w_{tH}(\tilde{p}, n_L, n_H, G)}{\partial n_H(p)} dG(\tilde{p}) - \frac{\partial w_{tL}(n_L, n_H, G)}{\partial n_H(p)} n_L + \beta(1-\delta_H)\mathbb{E}_t \frac{\partial D_{t+1}(n'_L, n'_H, G')}{\partial n_H(p)'}, \quad (18)$$

and the marginal surplus from employing a low-skilled worker is

$$\frac{\partial D_t(n_L, n_H, G)}{\partial n_L} = p\alpha_L a_t n_L^{\alpha_L - 1} \left(\int_p p n_H dG(p) \right)^{\alpha_H} - \int_{\tilde{p}} n_H \frac{\partial w_{tH}(p, n_L, n_H, G)}{\partial n_L} dG(p) \quad (19)$$
$$- w_{tL}(n_L, n_H, G) - \frac{\partial w_{tL}(n_L, n_H, G)}{\partial n_L} n_L + \beta(1-\delta_L)\mathbb{E}_t \frac{\partial D_{t+1}(n'_L, n'_H, G')}{\partial n'_L}.$$

The bargaining rule states that

$$\phi \frac{\partial D_t(n_L, n_H, G)}{\partial n_H(p)} = (1-\phi)\left(W_{tH}(p, n_L, n_H, G, d) - U_{tH}(d)\right) \quad (20)$$
$$\phi \frac{\partial D_t(n_L, n_H, G)}{\partial n_L} = (1-\phi)\left(W_{tL}(n_L, n_H, G) - U_{tL}\right) \quad (21)$$

Using the bargaining rules, we get two partial differential equations, which we can solve to get the following wage functions,

$$w_{tH}(p, n_L, n_H, G) = p \frac{\phi\alpha_H}{(1-\phi) + \phi(\alpha_L + \alpha_H)} a_t n_L^{\alpha_L} \left(\int_{\tilde{p}} \tilde{p} n_H dG(\tilde{p}) \right)^{\alpha_H - 1}$$
$$+ (1-\phi)(b_H + \Omega_{tH}(d)) \quad (22)$$
$$w_{tL}(n_L, n_H, G) = \frac{\phi\alpha_L}{(1-\phi) + \phi(\alpha_L + \alpha_H)} a_t n_L^{\alpha_L - 1} \left(\int_p p n_H dG(p) \right)^{\alpha_H}$$
$$+ (1-\phi)(b_L + \Omega_{tL}) \quad (23)$$

Given the wage equation solutions, we can rewrite the arguments to be $w_{tH}(p, n_L, n_H, \bar{p})$ and $w_{tL}(n_L, n_H, \bar{p})$, where $\bar{p} \equiv \int_p p dG(p)$. This implies than any individual high-skilled worker's wage depends on their own idiosyncratic productivity and the average productivity of all high-skilled workers.

Evolution of aggregate states We show the aggregation rules and laws of motion for aggregate state variables. The total number of vacancies in each sector is given by the exogenous mass of firms, ι, multiplied by the representative firm's optimal vacancy choice,

v_j^*

$$V_{tj} = \iota v_{tj}^*(n_L, \mathbf{n_H}, \mathbf{G}). \tag{24}$$

The masses of unemployed workers in the next period ar given by

$$u_{t+1H}(d) = \iota n_H(d)\delta_H + u_{tH}(d)\Big(1 - \theta_{tH} q(\theta_{tH})(1 - F(p^*(d)))\Big) \tag{25}$$

$$u_{t+1H}(f) = \iota n_H(f)\delta_H + u_{tH}(f)\Big(1 - \theta_{tH} q(\theta_{tH}) x_t (1 - F(p^*(f)))\Big). \tag{26}$$

$$u_{t+1L} = \iota n_L \delta_L + u_{tL}\Big(1 - \theta_{tL} q(\theta_{tL})\Big). \tag{27}$$

The probability that a visa application is successful is given by,

$$x_t = \frac{\bar{N}}{u_{tH}(f)\theta_{tH} q(\theta_{tH})(1 - F(p^*(f)))}. \tag{28}$$

Equilibrium Given the exogenous masses of firms and foreign workers, ι and μ, and an exogenous cap on the number of visas granted, $\bar{N}$, a stationary recursive competitive equilibrium for this economy is a list of functions: (i) value functions for firms $J_t(n_L, \mathbf{n_H}, \mathbf{G})$, (ii) decision rules for vacancies and hiring, $v_{tj}(n_L, \mathbf{n_H}, \mathbf{G})$ and $h_t(p, n_L, \mathbf{n_H}, \mathbf{G}, i)$, (iii) value function for workers, $W_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G}, i)$, $W_{tL}(n_L, \mathbf{n_H}, \mathbf{G})$, $U_{tH}(i)$ and U_{tL}, (iv) wage function $w_t(p, n_L, \mathbf{n_H}, \mathbf{G})$; and market tightness, θ_j, visa acceptance rate, x, worker outside options, $\mathbf{\Omega_H(i)}, \mathbf{\Omega_L}$, and a stationary distribution of productivity in employment and workers in unemployment $\mathbf{u_H(i)}, \mathbf{u_L}$.

1. *Firm optimization*: Given $\mathbf{u_H(i)}$, x, θ_H, θ_L, $w_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G})$, and $w_{tL}(n_L, \mathbf{n_H}, \mathbf{G})$, $v_{tH}(n_L, \mathbf{n_H}, \mathbf{G})$, $v_{tL}(n_L, \mathbf{n_H}, \mathbf{G})$, $h_t(p, n_L, \mathbf{n_H}, \mathbf{G}, i)$ solve the firm's problem.

2. *Worker optimization*: Given $w_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G})$, $w_{tL}(n_L, \mathbf{n_H}, \mathbf{G})$ and $\mathbf{\Omega_H(i)}, \mathbf{\Omega_L}$, value functions $W_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G}, i)$, $W_{tL}(n_L, \mathbf{n_H}, \mathbf{G})$, $U_{tH}(i)$ and U_{tL} solve the worker's problem. Given θ_H, x, $h_t(p, \mathbf{n}, \mathbf{G}, i)$, $\mathbf{\Omega_H(i)}$ is consistent with equations 10 and 11. Given θ_L, Ω_L is consistent with equation 14.

3. *Wage bargaining*: Wage functions $w_{tH}(p, n_L, \mathbf{n_H}, \mathbf{G})$ and $w_{tL}(n_L, \mathbf{n_H}, \mathbf{G})$ satisfy the bargaining rule.

4. *Consistency*: Given θ, $\mathbf{u_H(i)}$, $h_t(p, n_L, \mathbf{n_H}, \mathbf{G}, f)$, x is consistent with the number of visa applications submitted. Given $\mathbf{u_H(i)}$ and $v_{tj}(n_L, \mathbf{n_H}, \mathbf{G})$, $\{\theta_\mathbf{j}\}$ is consistent with the aggregate number of vacancies. Stationary distributions $\mathbf{u_H(i)}, \mathbf{u_L}$ are consistent with the laws of motion given by equations 25, 26, and 27.

Model solution The solution to the wage bargaining problem allows us to simplify the firm's problem such that the firm only needs to keep track of the average productivity of high-skilled workers, rather than the full distribution of productivity. Let $\bar{\mathbf{p}} = \{\bar{p}(d), \bar{p}(f)\}$ be the average productivity of incumbent skilled workers of each nationality. We can rewrite the firm's value function as follows,

$$\begin{aligned} J_t(n_L, \mathbf{n_H}, \bar{\mathbf{p}}) = \max_{v_j, p^*(i), n'_L, \mathbf{n'_H}} & -v_H c_{vH} - v_L c_{vL} - \sum_{i \in \{d,f\}} \underbrace{v_H q(\theta_{tH}) \frac{u_{tH}(i)}{u_{tH}} (1 - F(p^*(i)))}_{\text{number of hires, } i} c_h(i) \\ & + \left(\frac{1-\phi}{1-\phi+\phi(\alpha_H+\alpha_L)} \right) a_t (n'_L)^{\alpha_L} \Big(\bar{p}'(d) n'_H(d) + \bar{p}'(f) n'_H(f) \Big)^{\alpha_H} \\ & - (1-\phi)(b_H + \Omega_{tH}(d)) \Big(n'_H(d) + n'_H(f) \Big) - (1-\phi)(b_L + \Omega_{tL}(d)) n'_L \\ & + \beta \mathbb{E} J_{t+1}(n_L, \mathbf{n'_H}, \bar{\mathbf{p}}') \end{aligned} \tag{29}$$

s.t. (4), (6), (7) and

$$\bar{p}'(d) n'(d) = (1-\delta) n(d) \bar{p}(d) + v q(\theta_t) \frac{u(d)}{u(d)+u(f)} \int_{p \geq p^*(d)} p dF(p) \tag{30}$$

$$\bar{p}'(f) n'(f) = (1-\delta) n(f) \bar{p}(f) + v q(\theta_t) x_t \frac{u(f)}{u(d)+u(f)} \int_{p \geq p^*(f)} p dF(p) \tag{31}$$

Let $\lambda_{tH}(i)$ be the Lagrange multiplier on the constraint for the law of motion for effective high-skilled employment $\bar{p}'(i)n'(i)$, as defined in equations 30 and 31, $\mu_{tH}(i)$ be the Lagrange multiplier on the constraint for the law of motion for high-skilled employment $n'(i)$, as defined in equations 6 and 7, and λ_{tL} be the Lagrange multiplier on the constraint for the law of motion for low-skilled employment, as defined in equation 4. Taking first order conditions

with respect to $n'(i)$, we get the shadow value of one additional worker with nationality i and average productivity $\bar{p}(i)$, $\bar{p}(i)\lambda_{tH}(i) + \mu_{tH}(i)$, can be expressed as

$$\bar{p}(i)\lambda_{tH}(i) + \mu_{tH}(i) = \bar{p}(i)\alpha_H \left(\frac{1-\phi}{1-\phi+\phi(\alpha_H+\alpha_L)}\right) a_t (n'_L)^{\alpha_L} (\bar{p}(d)' n_H(d)' + \bar{p}(f)' n_H(f)')^{\alpha_H - 1}$$
$$- (1-\phi)(b_H + \Omega_{tH}(d)) + \beta(1-\delta_H)\mathbb{E}_t \left[\bar{p}(i)\lambda_{t+1H}(i) + \mu_{t+1H}(i)\right], \tag{32}$$

which can be interpreted as the worker's marginal product minus wages, plus the expected discounted value of continuing to employ the worker in the next period. The term multiplying the marginal product of labor captures spillovers that employing more high-skilled workers has on both high-skilled and low-skilled wages. The shadow value can be decomposed into a term that depends on the productivity of the worker, $\lambda_{tH}(i)$, and a term that is constant across workers, $\mu_{tH}(i)$. This reflects that if a firm hires a worker with productivity $p' > p$, they get a higher marginal product of labor and have to pay a higher wage, but they have to compensate both workers equally for their outside option of returning to unemployment.

The shadow value of a low-skilled worker is

$$\lambda_{tL} = \alpha_L \left(\frac{1-\phi}{1-\phi+\phi(\alpha_H+\alpha_L)}\right) a_t (n'_L)^{\alpha_L - 1} (\bar{p}(d)' n_H(d)' + \bar{p}(f)' n_H(f)')^{\alpha_H}$$
$$- (1-\phi)(b_L + \Omega_{tL}) + \beta(1-\delta_L)\mathbb{E}_t \lambda_{t+1L}, \tag{33}$$

Note the similarities between the shadow values of the low- and high-skilled effective workers. The key difference is that the outside options for the low-skilled worker do not need a component scaled by average productivity because effective low-skilled labor is equal to the number of low-skilled workers.

Taking first order conditions with respect to $p^*(i)$, we get

$$c_h(d) = \lambda_t(d) p^*(d) + \mu_{tH}(d) \tag{34}$$
$$\frac{c_h(f)}{x_t} = \lambda_t(f) p^*(f) + \mu_{tH}(f). \tag{35}$$

Given a set of matches, these conditions state that the firm should hire matches until the

marginal cost of hiring is equal to the marginal benefit, or the shadow value of the marginal hire. The condition for the foreign worker threshold accounts for the lower probability that firms are actually able to employ the worker, x_t. Under the assumption that firms only bargain wages with domestic workers, $\lambda_t(d) = \lambda_t(f)$ and $\mu_{tH}(d) = \mu_{tH}(f)$. Equations 34 and 35 thus show that the marginal foreign hire will have a higher productivity than the marginal domestic hire, given $c_h(d) < c_h(f)$ and $x_t < 1$.

Finally, from the first order conditions with respect to vacancy posting we get

$$\frac{c_{vH}}{q(\theta_{tH})} = \frac{u_{tH}(d)}{u_{tH}}\left(\lambda_{tH}(d)\int_{p\geq p^*(d)} pdF(p) + (\mu_t(d) - c_h(d))(1 - F(p^*(d)))\right) \tag{36}$$
$$+ \frac{u_{tH}(f)}{u_{tH}}\left(x_t\lambda_{tH}(f)\int_{p\geq p^*(f)} pdF(p) + (x_t\mu_t(f) - c_h(f))(1 - F(p^*(f)))\right)$$
$$\frac{c_{vL}}{q(\theta_{tL})} = \lambda_{tL} \tag{37}$$

which is equivalent to the free-entry condition in a one-worker one-firm model in each market. Firms will post vacancies until the marginal cost of posting an additional vacancy times the expected duration of the vacancy, $\frac{c_{vj}}{q(\theta_{tj})}$ is equal to the marginal benefit. For low-skilled workers, this is simply equal to the shadow value. For the high-skilled sector, the marginal benefit depends on the probability that the match is foreign or domestic, the probability that the firm would hire the marginal match, the average productivity of matches, and the likelihood of a foreign visa being granted.

4. Quantification

Calibration sketch We calibrate the parameters in two parts. First, there is a set of parameters that are fixed using estimates from other papers or normalizations. We assume a Cobb-Douglas matching function, $M(U, V) = \zeta_j U^{\psi} V^{1-\psi}$ with $\psi = 0.5$, as in Petrongolo & Pissarides (2001). We allow the scale parameter to differ for the high-skilled and low-skilled markets because the vacancies differ in their nature. In the high-skilled sector, firms hire only a fraction of their matches, whereas they hire all low-skilled matches. We choose ζ_j to

match the job-filling rate in each sector.

We set $\beta = 0.96$, and $\alpha_H + \alpha_L$=0.67, consistent with the RBC literature. The visa cap, $\bar{N}$, is set such that the probability of a successful visa is 40%. Following Baydur (2017), the flow value of unemployment for high-skilled workers, b_H, is set such that the total outside option of high-skilled workers, $b_h + \Omega_H(d) = 1$, and the mass of firms in the economy, ι, is set such that market tightness for the high-skilled sector is 0.72. We choose the vacancy cost for high-skilled workers such that firms hire 20% of matches, as in Barron *et al.* (1997) or Villena-Roldan (2012). We choose the vacancy cost for low-skilled workers such that the shadow value of the marginal high-skilled worker is equal to the shadow value of a low-skilled worker.

Table 1: Parameters (preliminary)

Parameter	Value	Meaning	Target
β	0.96	Discount rate	Real interest rate
α_H	0.47	Production returns to scale	RBC lit
α_L	0.2	Production returns to scale	Relative employment
ψ	0.5	Matching elasticity	Petrongolo & Pissarides (2001)
ζ_H	0.81	Matching function scalar	Job-filling rate
ζ_L	0.40	Matching function scalar	Job-filling rate
δ_H	0.003	Separation rate	Unemployment rate
δ_L	0.018	Separation rate	Unemployment rate
a	14	Aggregate productivity	Avg firm size
b_H	0.25	Flow unemployment value	Normalize outside option
b_L	0.67	Flow unemployment value	High/low-skilled wages
ϕ	0.10	Worker bargaining power	
$\bar{N}$		Visa cap	Prob successful visa = 0.4
μ	0.13	Mass of foreign workers	Relative employment shares
π	1.81	Mass of low-skilled workers	L Job-finding rate
ι	0.004	Mass of firms	H market tightness
c_{hd}	2.80	Domestic hiring cost	Foreign relative wage
c_{hf}	2.91	Foreign hiring cost	Cost of filing H-1B/avg wage
c_{vH}	7.67	Vacancy cost	Hired share matches
c_{vL}	1.02	Vacancy cost	Equal shadow values of marginal workers
σ	1.44	SD of productivity dist	SD of domestic wage residuals

Next, we calibrate the second set of parameters to match moments from the data. The parameters affect all moments, but we provide some intuition on identification. The returns

to scale for low-skilled workers, α_L, is disciplined by the relative employment of high-skilled and low-skilled workers. The separation rates for high-skilled and low-skilled workers are informative about the equilibrium domestic unemployment rates for each group. We focus on domestic high-skilled unemployment because we can't know how many foreign applicants are captured by surveys like the CPS versus returning to their home countries. Aggregate productivity, a, is used to target an average firm size of xx. The flow value of unemployment for low-skilled workers is used to target the relative wage gap between high-skilled and low-skilled workers. The mass of foreign workers is influential for the relative employment of domestic and foreign high-skilled workers, while the mass of low-skilled workers is informative about the job-finding rate for this group.

The gap between foreign and domestic hiring cost governs the relative selectivity of the two groups of high-skilled workers, and is therefore informative about the average wage differential between the two groups. The domestic worker hiring cost determines the scale, which we pin down by targeting the cost of filing an H-1B visa application relative to the average salary. We assume that match quality is drawn from a log-normal distribution with mean 0 and standard deviation σ. We calibrate σ by targeting the variance of domestic wage residuals among employed workers, as shown in Figure 2.

Table 2 shows the fit of the model relative to the targeted moments in our preliminary calibration.

Counterfactual analysis In this section, we consider two policy experiments. First, we increase the visa cap by 2%. Second, we maintain the existing cap, but we change how visas are allocated. Visas are granted on the basis of productivity and firms internalize the change in procedures.[3] Thus, this counterfactual includes the partial effect of reallocating visas given the firm's baseline decision rules as well as the general equilibrium effect of firms adjusting these rules in response to the policy change. For comparison, we also show a third counterfactual that eliminates the application costs for H-1B visas but keeps the lottery in tact with the existing cap.

[3]In our model, this could be achieved by allocating on the basis of wages since wages are perfectly correlated with productivity.

Table 2: Targeted moments (preliminary)

Moment	Data	Model
Employment shares (ppt)		
Foreign	10.92	10.77
High-skilled	38.00	38.76
Wage ratios		
Foreign/domestic high-skilled	1.03	1.03
High/low skilled	1.82	1.56
SD Domestic wages	0.45	0.45
Domestic unemployment rates (ppt)		
High-skilled	2.24	2.25
Low-skilled	4.39	4.42
Filing cost/avg wage (ppt)	8.90	8.70

Table 3: Counterfactual exercises (preliminary)

		% change from baseline steady state		
	Baseline	Increase cap	Rank candidates	Same cost
Productivity	10.36	1.33	11.97	-0.07
Foreign	11.47	-2.53	79.76	-0.54
Domestic	10.23	1.84	2.80	-0.01
Employment (per firm)	715	0.15	0.30	0.00
Foreign	30	2.00	0.00	0.00
Domestic high-skilled	247	0.05	0.02	0.00
Domestic low-skilled	438	0.07	0.48	0.00
Output	1994	0.76	5.57	-0.03

Table 3 reports the results for our comparison across steady states. Increasing the visa cap leads to a 1.3% increase in average productivity among high-skilled workers. The average productivity of foreign workers falls, because it becomes relatively less costly to hire them with a more lenient cap, so firms will relax their hiring standards. They will also become more selective about hiring domestic workers, because there is some substitution between the two groups. However, average productivity rises both due to the increase in selectivity for

domestic workers and the compositional shift towards more foreign workers, who still have a higher average productivity than domestic workers. Employment for foreign high-skilled workers mechanically increases by 2%. We see quantitatively small increases in employment for domestic workers. For high-skilled domestic workers, this increase comes from less congestion in the labor market. Although firms are being more selective about which domestic high-skilled workers they hire, they are matching with relatively more domestic high-skilled workers now that more of the foreign high-skilled workers have become employed and left the applicant pool. This arises due to the random search assumption, but can be interpreted as firms having to review fewer applications from foreign applicants that they are not going to hire. Domestic low-skilled workers also see a small boost in employment due to the increase in average productivity and complementarities in production.

In the second counterfactual, we allocate visas by ranking candidates. This has huge effects on productivity. Mechanically, the productivity of foreign workers should increase due to the reallocation in partial equilibrium. Productivity will increase further for both groups because it is now more lucrative for firms to post more vacancies to seek out the most productive workers, given that they will not get a visa accepted for a lower-productivity foreign worker anymore. Turning to employment, since the cap hasn't increased, foreign employment is unchanged. Domestic high-skilled employment is essentially unchanged but marginally increases. This effect is due to the change in hiring strategy, with firms posting relatively more vacancies in a way that more-than-offsets the increase in selectivity for domestic workers. Notably, we don't see this reallocation leading firms to hire fewer domestic workers. This is again a consequence of random search. If the reallocation of visas leads firms to recruit more vigorously, they will also end up recruiting more domestic applicants, more of which will be hired than foreign since they don't face a lower bound on productivity for this group. Finally, we see big gains to employment for low-skilled workers of about 0.5 percent, due to the large increase in productivity.

For comparison, we show that the H-1B visa application costs are relatively insignificant. Removing this friction while keeping the cap in place would lead to a slight decrease in productivity, because firms would be less selective about hiring. This would create spillovers

through adjustments in their hiring strategy that would decrease average productivity very slightly for domestic workers as well. The effects are quantitatively small enough that they do not affect domestic employment, though output decreases very slightly.

References

Barron, John M., Berger, Mark C., & Black, Dan A. 1997. Employer Search, Training, and Vacancy Duration. *Economic Inquiry*, **XXXV**, 167–192.

Baydur, Ismail. 2017. Worker Selection, Hiring, and Vacancies. *American Economic Journal: Macroeconomics*, **9**(1), 88–127.

Bound, John, Braga, Breno, Golden, Joseph M., & Khanna, Gaurav. 2015. Recruitment of Foreigners in the Market for Computer Scientists in the United States. *Journal of Labor Economics*, **33**(S1), S187–S223. Publisher: The University of Chicago Press.

Bound, John, Khanna, Gaurav, & Morales, Nicolas. 2018. Understanding the Economic Impact of the H-1B Program on the United States. *Pages 109–175 of:* Hanson, Gordon H., Kerr, William R., & Turner, Sarah (eds), *High-Skilled Migration to the United States and Its Economic Consequences*. University of Chicago Press.

Elsby, Michæl W.L., Michæls, Ryan, & Ratner, David. 2015. The Beveridge Curve: A Survey. *Journal of Economic Literature*, **53**(3), 571–630. Publisher: American Economic Association.

Flood, Sarah, King, Miriam, Rodgers, Renae, Ruggles, Steven, & Warren, J. Robert. 2022. *Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]*. Place: Minneapolis, MN.

Hegarty, Caitlin. 2024. Firm Heterogeneity and Racial Labor Market Disparities.

Mehra, Mishita, & Shen, Hewei. 2022. Skilled immigration, firms, and policy. *Journal of International Economics*, **139**(Nov.), 103662.

Petrongolo, Barbara, & Pissarides, Christopher A. 2001. Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, **39**(2), 390–431.

Sharma, Rishi R., & Sparber, Chad. 2024. Buying lottery tickets for foreign workers: Lost quota rents induced by H-1B policy. *Journal of International Economics*, **150**(July), 103932.

Stole, Lars A., & Zwiebel, Jeffrey. 1996. Intra-firm Bargaining under Non-binding Contracts. *The Review of Economic Studies*, **63**(3), 375–410. Publisher: Oxford Academic ISBN: 63/3/375/1552875.

Villena-Roldan, Benjamin. 2012. *Aggregate Implications of Employer Search and Recruiting Selection*. University of Chile Center for Applied Economics Working Paper 271.

Waugh, Michael E. 2017. Firm Dynamics and Immigration: The Case of High-Skilled Immigration. *Pages 205–238 of: High-Skilled Migration to the United States and Its Economic Consequences*. University of Chicago Press.