Skilled Immigration, Firms, and Policy*

Mishita Mehra[†]

November 13, 2017

Abstract

This paper studies the macroeconomic general equilibrium effects of skilled immigration policy changes by explicitly taking into account the role of firm demand for foreign skilled labor. To this end, I develop a two-sector dynamic stochastic general equilibrium model with monopolistically competitive firms and heterogeneous workers. Unlike most previous studies that view immigration as a supply-induced shock, the paper models skilled labor immigration as an endogenous response to an increase in firm labor demand in the receiving economy. The model is calibrated to mimic the U.S. economy with its current immigration policy: Firms face hiring costs and there is an occasionally binding cap on the foreign skilled workers that can be hired each period. The results indicate that a less restrictive skilled immigration policy via an immigration cap increase leads to heterogeneous effects on skilled and unskilled workers — unskilled domestic workers gain but skilled domestic workers lose. However, the magnitude of the welfare impacts depends on the state of the economy at the time of the cap change and also on the structure of the labor market (presence of search frictions). This paper also evaluates the welfare and efficiency gain from moving toward an alternate skilled immigration policy with a market-driven allocation of permits for hiring skilled foreign workers. Such a policy increases welfare and brings the economy's allocation closer to the social planner's first-best allocation.

*I am extremely grateful to my adviser, Fabio Ghironi, for his constant guidance and support. I am also thankful to Federico Mandelman, Oksana Leukhina, and Quan Wen for their guidance. I also thank Mu-Jeung Yang and Dominique Brabant for helpful discussions. I am grateful to financial support from the Alfred P. Sloan Foundation Pre-doctoral Fellowship on the Economics of High Skill Immigration, awarded through the NBER. I am also extremely thankful to the Federal Reserve Bank of Atlanta for their generous support and hospitality.

[†]PhD candidate, Department of Economics, University of Washington, Seattle, 98105, email: mishita@uw.edu

1 Introduction

There has been a rapid increase in the number of foreign skilled workers in the U.S. labor force. Among all foreign-born individuals, those with at least a bachelor's degree witnessed the sharpest increase (42 percent) during the 2004 - 2015 period (Figure A.1). The corresponding increase for the native born in the same skill group was 26 percent. This led to an increase in the proportion of college-educated foreign born in the U.S. labor force from 14 percent to 16 percent (Figure A.2).[1]

Firm demand for foreign skilled workers has played an important role in generating this increase. Since its inception in 1990, the H1-B visa program remains the dominant entry route of foreign skilled workers into the U.S. labor force (Figure A.3).[2] Firms play a crucial role in hiring, sponsoring, and incurring costs at each stage of the H1-B application process for a foreign worker. The first step requires a firm that wants to hire a foreign worker to file a Labor Condition Application (LCA) with the Department of Labor in which one of the items that they need to specify is the number of foreign workers they would like to hire for a particular occupation. These LCAs signal vacancies or firm demand for foreign skilled labor. However, the actual number of visas issued to foreign workers is determined by a policy-imposed cap.

The gap between firm demand, measured by the number of workers requested in the LCAs filed, and visas issued tends to grow during expansionary periods (Figure 1). Moreover, the visa cap was met in each year since 2004, in less than a week in seven of those years (Figure 2), and visas for foreign skilled workers were allocated according to a lottery process.[3] These facts indicate that there is a strong demand for foreign skilled workers that is not accommodated by the current immigration policy.

The role of firm demand of foreign skilled workers and the current allocation mechanism of foreign workers across them have implications for how an increase in skilled immigration and changes in immigration policies impact the aggregate economy. However, these implications are not fully understood in the current literature since most studies view immigration as a supply-induced shock.

[1] Over a longer horizon of two and a half decades, the foreign-born share of the total population with a bachelor's degree in the U.S. labor force increased from 10 percent in 1990 to 17 percent in 2015 (U.S. Census Bureau's 1990 and 2000 Decennial Census). Data from 2004 - 2015 is complied from the Current Population Survey (CPS). Foreign born in this survey include legally-admitted immigrants, refugees, temporary residents and temporary workers, and undocumented immigrants. However, the number of undocumented unskilled immigrants is likely to be underreported. In this study, I do not distinguish between foreign born and immigrants even though the legal definitions are different.

[2] The Appendix B discusses details on the H1-B visa program.

[3] Firms have from April 1st until the beginning of the next fiscal year to file petitions for H1-B visa applications.

Motivated by these facts, the goal of this paper is to the address the following questions: First, what are the welfare impacts of skilled immigration policy changes on skilled and unskilled domestic households in a general equilibrium framework that focuses on firm hiring of foreign workers according to U.S. skilled immigration policy? Second, can an alternate immigration policy in which the government allocates the same quota of visas according to a market-driven allocation of permits increase welfare of domestic households? Specifically, would such a policy close the gap between the decentralized economy's allocation and the efficient allocation chosen by a social planner?

To this end, I develop a two-sector dynamic stochastic general equilibrium (DSGE) model. This is the baseline model in the paper. Heterogeneous monopolistically competitive firms in the skill-intensive sector produce output by employing skilled domestic and foreign labor. The main incentive for hiring foreign labor in the baseline model is the inelastic supply of domestic skilled labor.[4] Skilled labor immigration is modeled as an endogenous response to an increase in firm labor demand in the domestic economy, subject to immigration policy restrictions that mimic current U.S. policy: Firms face hiring costs and there is a cap on the number of foreign workers that can be hired each period. This cap binds when economic conditions are such that the aggregate demand for foreign labor exceeds the policy-imposed quota. If the cap is met, the endogenously determined probability of being able to hire each foreign worker is less than one. Firms take into account these immigration policy restrictions and optimally choose to hire foreign labor until the expected discounted benefit from hiring foreign skilled workers is equal to the expected cost. Since a significant proportion of foreign skilled workers on an H1-B visa are temporary workers, the model allows for an exogenous probability of return to the country of origin. Revenue from immigration policy is collected by the government and transferred to domestic households. Unskilled labor in the economy produces a homogeneous good in the second sector.[5]

I calibrate the main parameters of the baseline model that pertain to immigration to match the U.S. economy during the 2004 - 2014 period. I then employ the calibrated model to study dynamics of economic variables in response to productivity shocks and changes in immigration policy. I calculate the welfare effect on domestic households of an increase

[4]This assumption is relaxed in the extended model with search and matching frictions.

[5]This two-sector model is qualitatively similar to a one-sector version of the model in which firms hire both skilled and unskilled workers. In the two-sector version, complementarities between skilled and unskilled workers exist through the consumption basket, while in the one-sector version, these complementarities exist through the production technology. Moreover, in the context of skilled foreign-born labor, around 73 percent of the Labor Condition Applications are requested by the relatively skill-intensive NAICS Sector 54 — Professional, Scientific, Technical services Sector (United States Department of Labor). The Professional and Business sector as a whole contributes around 12.4 percent of value added as a percentage of U.S. Gross Domestic Product (Bureau of Economic Analysis).

in the policy-imposed cap. I show that this welfare impact quantitatively depends on the realized state of the economy at the time of the cap change as this influences firm hiring of foreign labor. This experiment is motivated by the fact that the cap has been non binding during certain, especially recessionary periods. For instance, the cap was increased to 195,000 between the 2001 to 2003 period, during the start of which the economy entered a recessionary phase. The cap did not bind during this period as firms did not increase hiring by the full amount of the cap increase (Figure 2).[6] Therefore, for this experiment, I calibrate the baseline model to match U.S. data during the 2001 - 2003 period, when the cap was non binding.

I then compare the skilled immigration policy in the baseline model with an alternate policy in which the government allocates permits for hiring foreign skilled workers via a market-driven mechanism (Peri (2012)). To this end, I solve for the social planner's optimal allocation and derive the distortions and inefficiency wedges present in the baseline skilled immigration policy setup. I then compute the welfare and efficiency gains from a perfect foresight change in the current skilled immigration policy towards the alternate market-driven policy. Finally, I extend the baseline model to include search and matching frictions and non-competitive wage setting in order to study how the welfare impact of an immigration cap increase depends on the structure of the domestic labor market.

The results highlight some key insights that emphasize the importance of focusing on the role of the firm when studying skilled immigration. Under the baseline policy, unskilled domestic households gain (due to complementarities that increase unskilled wages) but skilled domestic households lose (due to substitutabilities that reduce skilled wages) from a ten percent immigration cap increase. The welfare gain (including transitional dynamics) amount to 0.0697 percent of annualized steady-state consumption for unskilled workers. For skilled domestic workers, the welfare loss amounts to 0.0502 percent of annualized steady-state consumption. An increase in the stock of foreign skilled labor also increases firm output and profits over time. However, the magnitude and impact of an increase in the cap may depend on two relatively overlooked factors.

First, the welfare changes for both skilled and unskilled domestic workers are much smaller in magnitude (one-sixth) if the cap change is implemented at a time when the economy is transitioning after a negative productivity shock. In this case, firms do not increase hiring by the full amount of the cap increase and the foreign labor stock increases by less. Therefore, it is important to take into account the state of the economy at the time of the cap change in order to evaluate the welfare implications of an immigration cap change.

Second, the structure of the labor market is important. The extension of the baseline

[6] This recession particularly hit the technology sector.

model that includes search and matching frictions (with the same immigration policy) shows that the welfare impact on domestic households differs (qualitatively and quantitatively), when compared to the baseline model. Even with perfectly substitutable domestic and foreign workers, an immigration cap increase leads to a higher employment of domestic skilled workers and an overall positive welfare gain for these workers, despite a fall in their wages. The welfare gain to unskilled households is also more than double compared to the baseline model.

The main mechanism behind this is that the cap increase encourages firms to post more vacancies, and this increases both domestic and foreign matches. In the search and matching framework, firms can be matched with either domestic or foreign workers (depending on the relative proportion of job searchers). However, under the current immigration policy, if the cap binds, firms are able to hire only a fraction of their foreign matches (given by the probability of an application being selected as in the baseline model). When the cap increases, this probability increases, and firms are able to retain more foreign matches. This increases their expected discounted benefit from posting a vacancy. More vacancies posted lead to an increase in the number of matched domestic skilled workers and hence in their domestic employment and welfare.

The main result from analyzing the alternate skilled immigration policy with a market-driven allocation of permits is that such a policy increases welfare of both skilled and unskilled domestic workers, compared to the baseline model. This is because immigration policy revenues collected by the government are about 36 percent higher. This has important implications because it shows that an alternate allocation of the same quota of foreign workers can potentially increase welfare and close a part of the inefficiency wedge between the decentralized economy's allocation and the first-best allocation chosen by a social planner. This is particularly relevant if the government wants to keep the cap on skilled foreign workers unchanged due to political resistance.

This paper has three main contributions. First, it includes a more realistic skilled immigration policy with an occasionally binding entry cap on foreign workers. The results show that this is relevant for evaluating the welfare impact of a cap change in the U.S. economy. In contrast, most studies that analyze the impact of skilled immigration policy changes view immigration as a supply-induced shock. Second, apart from studying the impact of skilled immigration policy reform via changes in the cap alone, this paper also begins to evaluate the impact of an alternate skilled immigration policy setup through a market-driven allocation of permits, which is related to the skilled immigration policy reform proposed in Peri (2012). Third, by incorporating a more realistic skilled immigration policy setup within the search and matching framework, the paper shows that even when U.S. and foreign workers are perfectly substitutable, an increase in the immigration cap can increase welfare of domestic workers. This is relevant because a key debate in the empirical literature is regarding the elasticity of substitution between native and immigrants in the same skill group and this is one of the reasons behind the lack of consensus regarding the impact of immigration on domestic workers (Ottaviano and Peri (2012), Borjas et al. (2008)). However, if labor markets are imperfectly competitive, domestic workers can gain from an immigration increase despite perfect substitution between domestic and foreign workers.

2 Related Literature

This research adds to the emerging literature that examines the implications of high skilled migration. This includes Borjas and Doran (2012), Ottaviano, Shih, and Sparber (2015), and Kerr and Lincoln (2010). The paper is also related to studies that measure the welfare gains from lowering barriers to labor mobility (Urrutia (1998); Klein and Ventura (2007, 2009); Iranzo and Peri (2009); Levchenko et. al. (2015); Ehrlich and Kim (2015)). In the context of DSGE models of international business cycles, the paper is related to Mandelman and Zlate (2012), who develop a two-country business cycle model with unskilled labor migration.

This paper is also related to empirical studies that highlight the role of firms in the context of skilled immigration (Kerr et al. (2013) and Ottaviano, Peri, and Wright (2015)). Kerr et al. (2014) stress the "need to increasingly develop a better understanding of the general equilibrium effects of skilled immigration with firms as a central element." Some recent studies have explicitly focused on the role of firms in a macroeconomic general equilibrium framework while discussing impacts of skilled immigration. Waugh (2017) studies the impact of a larger labor force (through an expansion of the H1-B visa program) on dynamics of firm entry and exit, and the effect on wages, aggregate output, and welfare. Bound et al. (2016) also use a general equilibrium model to study the effect of an increase in high-skill foreign born on domestic workers, consumers and firms, during the 1990s. My baseline model is consistent with their results — skilled immigration reduces wages of domestic skilled households, while redistributing gains to complements in production. Immigration lowers prices and raises output and profits of firms in the relevant sectors. However, skilled immigration is modeled as a labor supply shock in these studies. The explicit focus on the role of firm hiring of foreign skilled labor leads to some new insights that are relevant for evaluating skilled immigration policy changes — particularly that the welfare impacts of current immigration policy changes depend on how firms respond to the policy change, which in turn depends on the realized state of the economy.

The extended model with search and matching frictions is related to recent literature that studies the effects of immigration on the welfare of native individuals in a general equilibrium model featuring search frictions and wage bargaining (Chassamboulli and Palivos (2014), Battisti et al. (2014), Kingi (2015)). In all these, as long as immigrants have inferior outside options compared to natives, an increase in immigration raises firms' incentives to create vacancies which benefits all workers, including native skilled workers. My results are consistent with these papers — unskilled native workers gain unambiguously. Skilled native workers, on the other hand, gain in terms of employment despite wage losses. However, by focusing on a more realistic immigration policy that is relevant to the U.S., the extended version of my model with search and matching frictions is able to capture an additional channel by which immigration impacts vacancy postings and hence employment of domestic households — an increase in the immigration cap increases the probability of being able to retain a foreign match, and therefore the overall surplus from posting a vacancy.

3 Baseline Model

The baseline model features a two-sector economy that is populated by skilled and unskilled households and households with the same skill level are identical. Heterogeneous monopolistically competitive firms (as in Melitz (2003)) in sector 1 (the skill-intensive sector) produce differentiated goods using domestic and foreign skilled labor.[7] In the background, there is a foreign country that is assumed to have a large elastic supply of skilled workers that can be hired by domestic firms, subject to domestic firm demand and migration policy restrictions that mimic U.S. immigration policies - costs of hiring and a cap that occasionally binds, depending on the state of the economy.[8] Therefore, the constraint that firms face for hiring skilled labor is an outcome of immigration policy, rather than the supply of foreign skilled labor.[9]

Foreign and domestic skilled workers are perfect substitutes in the baseline model and earn the same wage under competitive labor markets. This is consistent with the overall evidence on relative earnings of foreign born as a percent of native born for workers with a bachelors degree or higher (Figure A.5). Moreover, when filing a Labor Condition Ap-

[7]Kerr et al. (2014) highlight heterogeneity in demand of foreign skilled workers across firms (Figure A.4).

[8]For the H1-B visa program, it does not matter whether the foreign-born worker is employed directly from the foreign country or from the domestic economy (for instance, after studying in the U.S.), as firms need to go through the same procedures in both cases. Although there is an additional quota for workers who obtain a master's degree or higher from a U.S. institution, I ignore this distinction in the model.

[9]This assumption is realistic due to the significant wage differences between OECD and developing countries. If hired, there is a strong incentive for a foreign skilled worker to migrate. Empirically, Clemson (2012) estimates that there is a six-fold increase in salary for skilled workers who migrate to the US.

plication, firms attest that they will pay the worker the prevailing compensation for that occupation. Given the debate surrounding the degree of substitutability between domestic and foreign workers, I show how the model can be modified to include complementarities between the workers in the appendix.[10]

I ignore emigration from the domestic economy as I treat the domestic economy as OECD countries like the U.S. and the foreign economy as developing countries (China, India, and the Philippines) and there is a very small share of migration from OECD to developing countries (OECD, 2013).[11] Here, I focus on the domestic economy and do not model foreign explicitly.

For immigrants, there is an exogenous probability of return to the country of origin, to account for the fact that a bulk of foreign skilled workers are on a temporary work visa and a fraction returns every period. Moreover, the exogenous return to the country of origin helps ensure that even in the absence of shocks, there is some demand for foreign skilled labor in every period, as is evident in the data.

Representative perfectly competitive firms in sector 2 (the unskilled sector) produce output using unskilled domestic labor. There is a government that collects revenue from immigration policy and rebates it symmetrically to domestic households. All contracts and prices are written in nominal terms, and prices and wages are flexible. Thus, the model solution will focus only on real variables.

3.1 Domestic Households

The Home economy consists of a continuum of two types of infinitely-lived domestic households that supply units of skilled and unskilled labor inelastically. The labor supply of the representative skilled household is normalized to 1, and that of the representative unskilled household is $\bar{l}_u$. Each skilled and unskilled representative household has the same preferences over a basket of goods produced at Home. The lifetime utility of skilled and unskilled households is given by:

$$\max_{C_{j,t}} E_t \sum_{\tau=t}^{\infty} \beta^{\tau-t} \Big(lnC_{j,\tau} \Big) \quad \forall j \in \{s, u\}$$

[10]The estimated degree of imperfect substitutability between native and immigrants within the same education and experience group is around 20 (Ottaviano and Peri (2012)). In the model with complementaries (Appendix C), this corresponds to $\gamma = 0.9477$, and therefore results are similar to the baseline model where $\gamma = 1$.

[11]Theoretically, this can be justified in terms of technology differences between the domestic economy and the foreign economy that ensure that wages in the home economy will always be higher than wages at foreign, which would removes the incentive to migrate to the foreign economy.

where $C_{j.t} = (\frac{c_{1,t}}{\alpha})^{\alpha}(\frac{c_{2,t}}{1-\alpha})^{1-\alpha}$ is the consumption basket of each household. $c_{1,t}$ is the basket of sector 1 goods consumed, and $c_{2,t}$ is the sector 2 good consumed. $\alpha \in (0,1]$ is the weight of sector 1 goods in consumption. The consumption-based price index is $P_t = (p_{1,t})^{\alpha}(p_{2,t})^{1-\alpha}$, where $p_{1,t}$ and $p_{2,t}$ are the price indices of sector 1 and sector 2 goods, respectively. The price indices in units of the consumption basket are $\rho_{1,t} = p_{1,t}/P_t$ and $\rho_{2,t} = p_{2,t}/P_t$. Therefore, the consumption-based price index can also be expressed as $1 = (\rho_{1,t})^{\alpha}(\rho_{2,t})^{1-\alpha}$ in units of the consumption basket. The basket of sector 1 goods is given by $c_{1,t} = \int_{\omega\in\Omega}(c_{1,t}(\omega)^{\frac{\theta-1}{\theta}}d\omega)^{\frac{\theta}{\theta-1}}$, where $\theta > 1$ is households' symmetric elasticity of substitution across sector 1 goods. Thus, the price index of sector 1 output is $p_{1,t} = \int_{\omega\in\Omega}(p_{1,t}(\omega)^{1-\theta}d\omega)^{\frac{1}{1-\theta}}$ where $p_{1,t}(\omega)$ is the price of the good ω. The demand for each good in sector 1 by household type $j \in \{s, u\}$ is given by $\alpha(\frac{p_{1,t}(\omega)}{p_{1.t}})^{-\theta}\frac{P_t}{p_{1,t}}C_{j,t}$ or $\alpha(\frac{\rho_{1,t}(\omega)}{\rho_{1.t}})^{-\theta}\frac{1}{\rho_{1,t}}C_{j,t}$.

The demand for the sector 2 good by household j is given by $(1-\alpha)\frac{P_t}{p_{2,t}}C_{j,t} = (1-\alpha)\frac{1}{\rho_{2,t}}C_{j,t}$, where $\rho_{2,t}$ is the price of sector 2 output in units of the consumption basket. The budget constraint for the domestic skilled household is $w_{s,t} + d_t + T_{s,t} = C_{s,t}$ where d_t is the profit income of sector 1 firms and $T_{s,t}$ are the transfers from the government, both in units of the consumption basket. $w_{s,t}$ is the real wage paid to skilled labor, which will be determined in the competitive labor market for skilled workers. Skilled households are the sector 1 firm owners in the baseline model. Unskilled households consume the sum of their labor income and transfers from the government i.e. $C_{u,t} = w_{u,t}l_{u,t} + T_{u,t}$, where $w_{u,t}$ is the real wage paid to unskilled labor, and is also determined competitively in a separate labor market for unskilled labor.

3.2 Production

3.2.1 Skill-Intensive Sector (Sector 1)

There are a continuum of heterogeneous monopolistically competitive firms, each producing a differentiated variety $\omega \in \Omega$. There is no endogenous firm entry or exit. The constant mass of firms is normalized to 1. Production requires skilled (Home or Foreign) labor. Aggregate labor productivity is Z_t which is exogenous and follows an AR[1] process in logs. Firms are heterogeneous as they produce with different technologies indexed by relative productivity z. Firm specific productivity z follows a Pareto distribution $G(z)$, with shape parameter k, and support on $(z_{min}, \infty]$. Output supplied by firm z in sector 1 is $y_{1,t}(z) = Z_t z l_{s,t}(z)$, where the total mass of skilled labor employed is

$$l_{s,t}(z) = l^s_{h,t}(z) + l^s_{f,t}(z)$$

where h and f denote domestic and foreign skilled labor respectively. The domestic supply of skilled labor at home is inelastic and normalized to 1. $l^s_{f,t}(z)$ is the stock of Foreign skilled labor employed at firm z. Skilled domestic and Foreign labor are assumed to be perfect substitutes.

Domestic firms face certain immigration policy restrictions when hiring foreign workers: Firms have to pay hiring costs, and there is a government-imposed cap on the number of foreign workers that can be hired each period. The sunk hiring costs can be decomposed into two components — costs due to immigration policy, and technological costs of hiring foreign workers. Firms have to pay cost g_t to the government for each foreign skilled worker they apply for, which is refunded back if the worker is not allocated to the firm (unless it is discovered that multiple H1-B petitions are submitted for the same employee).[12] This reflects actual policy in which the filing fees is refunded back to firms for the workers that are not selected in the lottery, in the event that the cap for foreign skilled workers is binding. Firms also have to incur a sunk cost, $f_{R,t}$, for all foreign workers they apply for, which reflects the regulatory component of the immigration policy cost — legal fees and other administrative costs involved in the various processes for hiring foreign skilled workers.[13] To facilitate comparison with the Social Planner's allocation, an additional cost that firms face is the technologically imposed cost of hiring skilled foreign workers, $f_{T,t}$, which is the same cost that a social planner would face for hiring a foreign worker. One way to interpret this cost would be to think of this as the cost incurred by firms on airfare or relocation of foreign workers, once they are hired and approved by immigration policy to join the firm. Therefore, these costs are only applicable to foreign workers that firms are actually able to bring to the firm after the approval of the application. All costs are in units of the consumption basket. If firm z optimally chooses to submit applications for $N_{e,t}(z)$ workers, then the total cost that the firm will incur is $f_{R,t}N_{e,t}(z) + (g_t + f_{T,t})N_{e,t}(z)q_t$, where q_t turns out to be the endogenous probability or fraction of workers that firms are allocated if the cap binds, and is described below. Higher immigration policy costs imply a more restrictive immigration policy.

The entry cap for foreign skilled workers is exogenously set at $\bar{N}_{e,t}$. Since each firm submits applications for $N_{e,t}(z)$ foreign skilled workers, the probability of each application being selected will be $q_t = min[\frac{\bar{N}_{e,t}}{(\int_{z_{min}}^{\infty} N_{e,t}(z)dG(z))}, 1]$, where $q_t < 1$ if the aggregate demand of

[12]This includes various fees firms have to pay when filing the H1-B petition for each worker, which on average amount to $3,000.

[13]Firms in multiple surveys (for instance, by the Government Accountability Office (GAO)), document a range of direct and indirect costs associated with the H-1B program, including legal and administrative costs. Firms note that apart from the filing fees paid to the Department of Homeland Security, the main cost incurred is due to the opportunity cost of the time and effort spent in the process, which is captured by the regulatory component of the sunk cost, $f_{R,t}$, in the model.

foreign skilled workers, $\int_{z_{min}}^{\infty} N_{e,t}(z)dG(z)$, exceeds the cap, and the cap endogenously binds. Each firm knows that if it submits $N_{e,t}(z)$ applications, it will get $q_t N_{e,t}(z)$ workers. Each firm is of measure 0 and takes q_t as given in its hiring decision.[14]

The timing is as follows. A fraction δ of the foreign skilled workers currently employed by domestic firms (including newly hired workers from the previous period) are separated from firms at the beginning of the period. The state Z_t of the economy is realized, wages are determined, and firms produce period-t output. Firms then maximize expected discounted profits and optimally choose the number of foreign skilled workers to hire (or submit applications for), after taking into account the immigration policy restrictions. The realized state of the economy and the corresponding firm demand for foreign workers determine whether the cap binds, and an endogenously-determined fraction q_t of the applications are approved. These are the workers that firms are able to 'bring' to the firm for production. There is a time to build lag and those workers that survive the separation shock are added to the stock of next period's skilled labor stock.

Thus, the stock of foreign skilled labor at firm z in period $t+1$ is given by:

$$l^s_{f,t+1}(z) = (1-\delta)(l^s_{f,t}(z) + q_t N_{e,t}(z)) \tag{1}$$

Expressed in units of the consumption basket, the inter-temporal profit function of firm z is given by:

$$E_t \sum_{\tau=t}^{\infty} \beta_{\tau,t} \left[\rho_{1,\tau}(z) Z_\tau z l_{s,\tau}(z) - w_{s,\tau}(l_{s,\tau}(z)_{-} f_{R,\tau} N_{e,\tau}(z) - (g_\tau + f_{T,\tau}) N_{e,\tau}(z) q_\tau \right]$$

The inter-temporal discount factor that the firm applies to its profits is $\beta_{\tau,t} \equiv \beta(u'(C_{s,\tau})/u'(C_{s,t}))$, which is the inter-temporal discount factor of domestic skilled households, who are assumed to be domestic firm owners.

Optimal Hiring of Skilled Foreign Workers:

Firms maximize inter-temporal profits subject to (1). Each period, firms hire and submit applications for skilled foreign workers such that the expected discounted profit generated

[14]While no actual risk in the realization of the share of foreign workers allocated to firms may seem restrictive, it is consistent with how U.S. firms in the economy behave. To mitigate risk, U.S. firms subcontract a large part of the H1-B hiring process to large IT management firms. These firms substantially reduce the risk of procuring H-1B visas by applying in bulk. These outsourcing firms were awarded almost 20 percent of total H-1Bs in 2016, and the workers were then allocated to U.S. employers through subcontracting.

from an additional skilled foreign worker, v_t, is equal to the expected sunk hiring cost:

$$v_t = f_{R,t}/q_t + g_t + f_{T,t}$$

where

$$v_t = \sum_{\tau=t+1}^{\infty} E_t\{\beta_{\tau,t}(1-\delta)^{\tau-t}[\rho_{1,\tau}(z)Z_\tau z - w_{s,\tau}]\}$$

v_t can be expressed as:

$$f_{R,t}/q_t + g_t + f_{T,t} = (1-\delta)E_t\{\beta(C_{s,t+1}/C_{s,t})^{-1}[\rho_{1,t+1}(z)Z_{t+1}z - w_{s,t+1} + \frac{f_{R,t+1}}{q_{t+1}} + g_{t+1} + f_{T,t+1}]\} \tag{2}$$

where $w_{s,t}$ is the real wage paid to each skilled foreign worker. Labor markets are competitive and the real marginal cost of production for firm z is given by $\psi_{1,t} = \frac{w_{s,t}}{Z_t z}$. Thus differences in productivity across firms translate into different marginal costs, and firms with higher z have a lower marginal cost of production.

Firms serve only the domestic market. Market clearing for each firm z is given by $Z_t z l_{s,t}(z) = (\frac{\rho_{1,t}(z)}{\tilde{\rho}_{1,t}})^{-\theta} Y_t^c/\tilde{\rho}_{1,t}$, where the price $\rho_{1,t}(z)$ set by the firm is a proportional markup over the marginal cost: $\rho_{1,t}(z) = \frac{\theta}{\theta-1}\psi_{1,t}(z)$. The average sector 1 price, $\tilde{\rho}_{1,t}$, and aggregate demand, Y_t^c, are given in section 2.2.4. As is standard in the Melitz (2003) model, more productive firms face a higher demand for their output due to lower prices, and hence employ more skilled labor, including skilled foreign labor. Firm profits in period t are given by $d_t(z) = \rho_{1,t}(z)y_{1.t}(z)/\theta - f_{R,t}N_{e,t}(z) - (g_t + f_{T,t})q_t N_{e,t}(z)$.

3.2.2 Unskilled Sector (Sector 2)

Sector 2 output is produced by competitive firms that have an identical technology:

$$Y_{2,t} = Z_t l_{u,t}$$

where $l_{u,t}$ is the unskilled labor employed by the representative firm. The marginal cost of production for the firm is $w_{u,t}/Z_t$. Thus the price of the representative sector 2 good in units of the consumption basket is $\rho_{2,t} = w_{u,t}/Z_t$.

3.2.3 Government

The government collects revenue from immigration policy — the total revenue collected in units of the consumption basket is $R_t = g_t q_t \int_{z_{min}}^{\infty} N_{e,t}(z)dG(z)$. The revenue depends on the aggregate applications filed by all firms, and the cap, which affects q_t. The Government's budget constraint is given by $T_t = R_t$, where $T_t = T_{s,t} + T_{u,t}$ is lumpsum transfer to domestic households. I assume that $T_{s,t} = T_{u,t} = T_t/2$.

3.2.4 Aggregate Accounting and Equilibrium

The distribution of firm productivities is given by a Pareto distribution $G(z) = 1 - (\frac{z_{min}}{z})^k$, with lower bound z_{min} and shape parameter $k > \theta - 1$. As in Melitz (2003), aggregate productivity is defined as $\tilde{z} = [\int_{z_{min}}^{\infty} z^{\theta-1} dG(z)]^{\frac{1}{\theta-1}}$. From each firm's market clearing in sector 1, we can write the aggregate stock of foreign labor as a function of firm specific productivity i.e. $l_{s,t}(z) = f(z^{\theta-1})$. Thus, we can aggregate skilled labor as $\tilde{l}_{s,t} = l_{s,t}(\tilde{z}_t) = 1 + \tilde{l}^s_{f,t}$. The aggregate sector 1 output is $Y_{1,t} = Z_t \tilde{z}(1 + \tilde{l}^s_{f,t})$. The aggregate sector 1 price index is given by $\tilde{\rho}_{1,t} = \int_{z_{min}}^{\infty} (\rho_{1,t}(z))^{1-\theta} dG(z) \big)^{\frac{1}{1-\theta}} = \rho_{1,t}(\tilde{z}_t)$.

Aggregate consumption by households is given by $C_{s,t}+C_{u,t}+C_{i,t}$ i.e. the sum of consumption by domestic skilled, unskilled, and immigrant workers residing at Home. Immigrants consume their labor income, $C_{i,t} = w_{s,t}\tilde{l}^s_{f,t}$. Domestic labor market clearing requires that the aggregate domestic labor employed is equal to the inelastic supply i.e. $\int_{z_{min}}^{\infty} l^s_{h,t}(z) dG(z) = 1$ and $l_{u,t} = \bar{l}_u$. Firm demand for skilled foreign workers is met at the prevailing skilled wages as foreign households are assumed to elastically meet all domestic firm demand.

Aggregate accounting in the economy requires that $\tilde{\rho}_{1,t} Z_t \tilde{z} \tilde{l}_{s,t} + \rho_{2,t} Z_t \bar{l}_u = C_{s,t} + C_{u,t} + C_{i,t} + f_{R,t}\tilde{N}_{e,t} + f_{T,t} q_t \tilde{N}_{e,t}$.[15]

Table 1 summarizes the key equilibrium conditions in the model. There are 17 equations in 17 endogenous variables of interest: $Y_{1,t}$, $Y_{2,t}$, $\tilde{l}^s_{f,t}$, $\tilde{N}_{e,t}$, q_t, $w_{u,t}$, $w_{s,t}$, d_t, $\tilde{\rho}_{1,t}$, $\rho_{2,t}$, $\psi_{1,t}$, $C_{u,t}$, $C_{s,t}$, Y^c_t, R_t, T_t, $C_{i,t}$. Z_t follows an exogenous AR[1] process in logs. f_R, g, f_T,and $\bar{N}_{e,t}$ are exogenously set and calibrated in Section 5.

4 The Steady State

I now turn to the consequences of skilled immigration and skilled immigration policy changes by studying how skilled immigration responds to a temporary productivity shock, as well as the transitional dynamics and the long-run effects of a permanent increase in the immigration cap. Before presenting these results, I discuss implications of some key steady-state relationships that highlight some of the main model mechanisms. The analytical solution for the steady state of the model is given in Appendix D.

[15] As in Cacciatore (2012), aggregate demand (Y^c_t) includes a component other than household consumption. However, it is in the same units as the consumption basket.

Steady-state stock of foreign born

Since the model features an occasionally binding constraint, the model is equivalent to one with two regimes.[16] The constraint is binding under one regime and slack under the other and each regime has a separate non-stochastic steady state. In the appendix, I derive the steady-state stock of foreign-born labor in the regime when the cap does not bind ($q = 1$) as follows:

$$\tilde{l}_f^s = \frac{\alpha \tilde{z}^{\frac{\alpha}{(1-\alpha)}} \bar{l}_u}{(1-\alpha)} \left[\frac{(1-\delta)\beta Z}{\theta(f_R + f_T + g)(1-(1-\delta)\beta)} \right]^{\frac{1}{(1-\alpha)}} - 1 \tag{3}$$

Equation (3) helps identify the factors that influence a larger firm demand for foreign skilled labor, and consequently a larger stock of foreign skilled workers. The model predicts that the stock of foreign skilled labor will be higher when a country has a higher aggregate labor productivity, Z; a higher aggregate firm specific productivity, $\tilde{z}$; the skill-intensive sector has a larger weight, α, in the consumption basket; the stock of domestic unskilled workers, $\bar{l}_u$, is higher (due to complementarities via the consumption basket); the hiring costs of foreign skilled workers (g, f_R, and f_T) are lower (less restrictive immigration policy); there is a lower probability of return, δ, to the country of origin; when elasticity of substitution across goods (θ) is lower (which increases firm profits from each unit produced); and when there is a lower stock of domestic skilled workers available.

Long run consumption: skilled native households

In this section I show how the stock of foreign-born workers affects firm profits and skilled wages. One of the key hurdles facing policy easing comes from critics of skilled immigration who highlight the adverse impacts of an increase in the supply of skilled immigrants on the wages and income of native skilled workers. In order to analyze this, I decompose the steady-state impact of an increase in the stock of skilled foreign labor on the various components of consumption of skilled domestic workers as below.

$$C_s = \underbrace{(\theta-1) Z \tilde{z}^{\alpha} \left(\frac{\alpha \bar{l}_u}{(1-\alpha)(1+\tilde{l}_f^s)} \right)^{1-\alpha} / \theta}_{\text{Wage income of domestic skilled households}} \tag{4}$$

$$+ \underbrace{\left(\frac{\alpha \bar{l}_u}{1-\alpha} \right)^{1-\alpha} Z \tilde{z}^{\alpha} (1+\tilde{l}_f^s)^{\alpha} / \theta - (f_R + g + f_T) \delta \tilde{l}_f^s / (1-\delta)}_{\text{Firm Profits}} + \underbrace{g \delta \tilde{l}_f^s / (2(1-\delta))}_{\text{Transfers}} \tag{5}$$

In the model, an increase in skilled immigrant workers would reduce real wages (due to

[16] Guerrieri and Iacoviello (2015)

substitutabilities between domestic and foreign skilled workers) and thus have an adverse impact on long-run real consumption of domestic skilled workers (equation 4). However, an increase in skilled foreign immigrants also increases long-run output and firm profits. Even though a larger stock of foreign skilled workers implies larger hiring costs incurred by firms which would negatively affect their profit, as long as these costs are relatively low, firm profits would rise. If some of these profits are distributed back to skilled households as dividend (profit) income, this channel would mitigate some of the adverse effects on native skilled households' long-run real income and consumption (equation 5). Moreover, a larger foreign labor stock also ensures larger revenue collected as part of immigration policy. if some of the revenue from immigration policy is transferred back to skilled domestic households, then the adverse impact on wages due to an increase in foreign skilled workers can be further mitigated.

An important implication is that in order to carry out accurate welfare analysis of immigration policy changes, it is important to empirically estimate the profit distribution of firms across households. For instance, if part of the dividend income goes to immigrant workers (as part of stock options given to them), then the negative impact of immigration on domestic skilled workers would be worse.

Steady state when cap binds

In the alternate regime, when the entry constraint binds in steady state, the aggregate stock of foreign skilled workers is given by: $\tilde{l}_f^s = (1-\delta)\bar{N}_e/\delta$ (using (1) and $q\tilde{N}_e = \bar{N}_e$). In this case, the appendix shows that steady state hiring probability is given by:

$$q = f_R\left[\frac{(1-\delta)\beta(\alpha\delta\bar{l}_u)^{1-\alpha}Z\tilde{z}}{\theta(1-(1-\delta)\beta)((1-\alpha)\tilde{z}(\delta+(1-\delta)\bar{N}_e))^{1-\alpha}} - f_T - g\right]^{-1}$$

The probability of hiring each foreign skilled worker is higher when the demand for foreign skilled workers is lower, or when the entry cap $\bar{N}_e$ is lower. A larger cap makes it easier for firms to hire and get workers that they apply for. However, any of the factors that increase foreign skilled worker demand, reduce the probability of hiring such workers. The aggregate flow of foreign labor demanded is given by $\tilde{N}_e = \bar{N}_e/q$.

5 Calibration

In order to study the dynamics numerically, I calibrate the parameters of the model under the assumption that the steady state mimics the U.S. economy during the 2004 - 2014 period. I interpret each period as a year to accommodate the annual allocation of the H-1B visa cap. I calibrate the parameters that pertain to immigration to match average annual U.S. data from the Current Population Survey (CPS), and the United States Citizenship and Immigration Service (USCIS), between 2004 to 2014. I rely on existing literature for the values of the other parameters. I set $\beta = .96$, which implies an annual real interest of 4 percent. Following Ghironi and Melitz (2004), I set the elasticity of substitution across product varieties equal to $\theta = 3.8$, the dispersion of firm productivity draws, $k = 3.4$, and normalize z_{min} to 1. The share of sector 1 goods in consumption is set at $\alpha = 0.5$ so that the results are not biased due to differential share of consumption across sectors. I set the exogenous return shock to the country of origin at $\delta = 0.1$, in order to match the annual return migration rate of 10 percent (Center for Immigration Studies, 2011).[17]

The immigration cap $\bar{N}_e$ is set to 0.0022 in order to match the average cap imposed by actual policy (85,000) as a proportion of the normalized average domestic skilled labor in the economy. I calibrate the filing fees paid to the government g to 0.0519, to match the average filing fees incurred while submitting the H1-B petition as a proportion of annual skilled wages over the same period.[18] The sunk regulatory cost f_R is set to 0.8 to target the average petitions filed during the period which would generate a steady-state application selection probability of about 0.4. The domestic unskilled labor supply is calibrated to 1.84 (given the normalization of domestic skilled labor supply to 1) to match the share of domestic workers in the U.S. with less than a bachelor's degree of 34 percent. I set the technological part of the hiring cost at 0.0833 to target one month's real wage in the data.[19] These sunk hiring costs are assumed to stay fixed in the baseline setup.

Given that the model is calibrated to a period when the cap is binding, the model economy is in a binding steady-state regime. The steady state of the model features a skill premium of 1.41, which is within reasonable estimates.

6 Transition Dynamics and Welfare Results in the Baseline Model

In this section, I solve the calibrated model numerically to study dynamics in response to a temporary positive productivity shock and changes in immigration policy. I then calculate

[17] In reality, H1-B visas are allocated for a period of three years, and can be extended for another three years. However a sizeable proportion of H1-B workers stay for longer as firms sponsor their green card application. There is no concrete estimate of this proportion. Moreover some workers may end up leaving before the visa expires. Thus I take an annual average rate of return to the country of origin.

[18] The average filing fees at the time of submitting the petitions adds up to about \$3, 000.

[19] While there is no direct estimate of this cost in the data, a month's wage is a reasonable estimate of the relocation and other expenses that are meant to capture this cost.

the welfare effects of a perfect foresight increase in the immigration policy cap. I also show how these welfare impacts vary with the state of the economy and with the structure of the labor market (Section 9). All the welfare experiments in the paper are summarized in Table 2.

6.1 Dynamic Response to a Temporary Productivity Increase

I study the responses (percent deviation from the steady state) to a temporary 1 percent increase in aggregate productivity (Z_t) in the baseline model. In order to analyze the impact of the cap on the economy's dynamic responses to a productivity shock, I compare two cases. In case 1 (the benchmark case), there is a policy-imposed entry cap on foreign skilled workers in the economy. In case 2, there is no entry cap.

Figure 3 shows that in the presence of an entry cap, the increase in firm demand for foreign skilled workers in response to the productivity increase is less than half, when compared to case without the cap. This is because firms have to pay the sunk cost f_R for hiring workers that may not eventually be able to join the firm due to the binding cap, and thus would not contribute to firms' output and profit. In other words, to hire one worker, the firm needs to pay for f_R/q_t workers. Therefore, costs associated with the current immigration policy may distort firms' incentives for hiring foreign skilled workers. An implication of this is that the costs incurred due to the burdensome current immigration policy may lead to an inaccurate signal of firm demand for foreign workers.

Figure 3 also shows that the stock of skilled labor is inelastic in the short run and rises only slowly due to the time-to-build lag. In the presence of the cap, the stock of foreign skilled labor rises by much less. As a result, the increase in output, profits, and real wages of unskilled workers (and thus their consumption) is smaller in the presence of the cap. Without the entry cap, firm profits initially fall more as more resources are spent on hiring. However, this quickly recovers as the stock of foreign skilled workers increases over time and firms are able to produce more output. Unskilled wages are higher without the cap as a larger stock of foreign workers increases demand for goods produced in sector 2, which increases demand for unskilled labor and puts an upward pressure on their wages.[20] However, the real wage of skilled labor falls by less in the presence of the entry cap due to the smaller increase in the stock of skilled labor in the domestic economy. Despite a lower decrease in wages of skilled workers, their consumption is not higher in the presence of an entry cap during most of the transition period as the increase in firm profits is also lower in this case. However, an alternate distribution of firm profits could lead to an unambiguously lower consumption

[20]This is consistent with empirical evidence in Hong and McLaren (2015).

of domestic skilled workers. Overall, the presence of the constraint dampens the economy's response to aggregate shocks.

6.2 Welfare Analysis

The above dynamic responses indicate that an increase in skilled foreign labor has different impacts on heterogeneous workers. In order to draw inferences about the impact of current migration policy changes, it is important to quantify the welfare changes across different sets of workers. I calculate welfare impacts after a 10 percent perfect foresight increase in the entry cap. The long-run welfare gain of each type of native worker from the immigration policy easing is computed as the percentage increase (Δ) in consumption that would leave the households indifferent between the initial policy and the new policy with the higher cap, when the new policy is implemented at time $t = 0$. Transitional dynamics have been included in the welfare computations. Thus, Δ solves:

$$u\left[C_j(1+\frac{\Delta}{100})\right] = (1-\beta)\sum_{t=0}^{\infty}\beta^t u(C_{j,t}) \quad \forall\{j \in s, u\}$$

Suppose the cap increases by 10 percent. Since the baseline model is calibrated to a period when the cap is 'very' binding in the sense that the gap between firm demand for foreign labor and the cap is very large, firms increase hiring by the full 10 percent. The dynamics following this cap change are given in Figure 4. Firms increase hiring of foreign skilled labor following the cap increase and as the stock of skilled workers builds up, sector 1 output, unskilled wages, and consumption rise. As before, the increase in unskilled wages and therefore consumption is due to the higher demand for sector 2 output by the larger stock of skilled foreign labor. Average firm profits in sector 1 fall initially as they have to bear costs of hiring more foreign workers (the cost of hiring foreign skilled workers remains unchanged). However, profits recover over time as output increases. Skilled wages fall due to the larger inflow of foreign workers. The net effect on real consumption of domestic skilled workers is negative, despite their profit income increasing. The probability of hiring a foreign worker is higher in the new steady state due to the higher cap.

Table 2 shows that the welfare gain (including transitional dynamics) amount to 0.0697 percent of annualized steady-state consumption for unskilled workers. For skilled domestic workers, the welfare loss amounts to 0.0502 percent of annualized steady-state consumption. Thus, there are different effects of a skilled immigration policy change on heterogeneous workers — workers most complimentary to skilled immigrants gain, while those most substitutable lose. Part of the negative impact on consumption of skilled workers is mitigated as firm profits, and transfers from the government rise over time. The transitional dynamics in Figure 4 show that most of the welfare changes are realized slowly over the longer horizon.

I next turn to analyzing how these welfare effects depend on the state of the economy. In section 9, I evaluate how these welfare effects depend on the structure of the labor market by adding search and matching frictions in the baseline model.

6.3 Welfare Impacts and the State of the Economy

An important insight from explicitly taking into account the role for firm demand in endogenous skilled immigration is that the demand of foreign labor depends on the realized state Z_t of the economy. Therefore, welfare impacts of an immigration cap change will depend on how firms adjust their hiring in response to the change, which in turn depends on the realized state of the economy. For instance, when the H1-B cap was raised by 80,000 workers in 2001, after it had been binding for the previous three years, the cap did not bind between 2001-2003 (the period for which the cap was raised), as firms did not increase hiring by the full amount of the cap increase. Since the stock of foreign workers rose by less than anticipated, evaluating the impact of the policy change as a labor supply increase (increase in 80,000 workers), would be misleading.

In the context of the model, suppose the immigration cap change is implemented at a time when the economy is transitioning after a negative productivity shock, and the cap is relatively 'less binding' in the sense that the gap between firm demand and the cap is relatively small. Following the cap change, firms will optimally hire more foreign skilled labor, yet they may not increase hiring by the full cap increase. To see this, I calibrate the baseline model to match U.S. labor and immigration data between 2001-2003. In 2001, the cap was raised from 115,000 to 195,000, a 69.5 percent increase. At that point, the economy entered a recessionary phase, and firms did not increase their hiring of foreign workers by 69.5 percent. In fact, the unused cap in 2001, 2002, and 2003, was 31,400, 115,900, and 117,000, respectively.[21] The average additional cap used over the entire period was about 20 percent. I calibrate the model such that the cap change is implemented at a time when the economy is transitioning from a negative productivity shock. The negative shock is calibrated to match the fact that on average, firms increased there hiring of foreign labor by 20 percent due to the additional quota of 69.5 percent.

Figure 5 plots the transitional dynamics in response to a 69.5 percent cap change under two cases. The first is the one described above when the economy is transitioning after a negative productivity shock and is called the 'less binding case'. In the second case, the state

[21]Using USCIS historical data.

of the economy is such that the economy is not transitioning after a negative productivity shock and the immigration cap is 'more binding' for firms. This case can be interpreted as the case when a cap change of 69.5 percent would lead to a pure skilled labor supply increase of 69.5 percent because firms increase hiring by the full amount of the cap increase. Therefore, while in the first case, firms increase their hiring of foreign workers by only about 20 percent, in the second case, firms increase the hiring of foreign workers by the full amount of the cap change. Consequently, the stock of foreign skilled labor increases by much less in the first case and the impacts on all endogenous variables are smaller in the 'less binding' case compared to the 'more binding' case. The welfare impacts on both skilled and unskilled workers are also therefore dampened (Table 2). The welfare impact in the first case is about one-sixth of that in the second case, for both workers.

Thus the welfare impact would be lower than the welfare impact that would have been evaluated by considering the policy change without taking into account firm hiring responses. The key point is that it may be important to take into account the role of firm demand for foreign skilled workers, and also the timing of the reform, when evaluating the impact of immigration policy changes.

7 Social Planner's Solution and Inefficiency Wedges

In this section, I discuss the first-best, efficient allocation chosen by a social planner. I then compare the equilibrium conditions in the baseline decentralized economy (Table 1) to those implied by the planner's solution (Table 3) as this allows us to identify the distortions in the model economy and to define the inefficiency wedges relative to the efficient allocation. The aim is to analyze whether moving toward an alternate policy with a market-driven allocation of permits (Section 8) would close some of the inefficiency wedges and bring the market economy closer to the optimal allocation.

Appendix E presents the planner's problem and the equilibrium conditions implied by the solution to the problem. The social planner maximizes welfare of domestic households and chooses the optimal entry of foreign skilled workers in the domestic labor force, taking the firm size distribution, preferences, technology, and resources available in the economy as given. The only hiring cost that is relevant to the social planner is the technological component of the overall entry cost, f_T. Therefore, in the planner's environment, $g = f_R = 0$.

Appendix F describes the major distortions and derives the inefficiency wedges in the economy. The analysis shows that the market economy features three sources of distortions relative to the planned economy. These distortions lead to two margins of inefficiency wedges that are discussed below.

Job creation margin: Comparing the entry condition of foreign skilled workers under the decentralized solution with the entry condition under the social planner's equilibrium implicitly defines the inefficiency wedge under the market economy's job creation margin for foreign skilled workers.

The inefficiency along the job creation margin results in the wedge $\Sigma_{jc,t}$ given by:

$$\Sigma_{jc,t} = E_t\left[B_{t,t+1}(1-\delta)\left(\tilde{\rho}_{1,t+1}Z_{t+1}\tilde{z}\frac{(\Upsilon_{R,t}+f_T\Upsilon_\theta)}{f_T(\Upsilon_{R,t}+f_T)} + \frac{\Upsilon_{R,t}-\Upsilon_{R,t+1}}{\Upsilon_{R,t}+f_T}\right)\right]$$

First, monopoly power leads to a lower incentive for hiring foreign skilled workers by inducing a lower marginal revenue product of a match (captured by Υ_θ). Second, the presence of immigration policy costs (f_R and g) that differ from the technological component of the hiring cost (f_T) leads to another source of distortion in the market economy. Third, the presence of a binding cap further distorts the costs that firms have to incur. The additional distortion that the cap imposes is due to the fact that when the cap binds and $q_t < 1$ in the decentralized economy, in order to hire one worker, firms need to submit $1/q_t$ applications and hence incur f_R/q_t as regulatory costs. As described in Appendix F, $\Upsilon_{R,t} = f_R + g + \frac{f_R\Upsilon_{q,t}}{1+\Upsilon_{q,t}}$, where $\Upsilon_{q,t} = 1 - q_t$ is the distortion due to the presence of the cap that lowers the probability of being able to hire a foreign worker to less than 1 . Thus, $\Upsilon_{R,t}$ decomposes the distortions in the existing immigration policy into distortions imposed due to immigration policy costs and those due to the cap. In the baseline calibration, around 60 percent of this distortion is due to the presence of a binding cap and 40 percent is due to the costs. If $f_R + g = \Upsilon_{q,t} = \Upsilon_\theta = 0$, then the job creation wedge is equal to 0.

Consumption resource constraint wedge: Sunk regulatory costs imply a diversion of resources from consumption, leading to a consumption-output efficiency wedge given by $\Sigma_{r,t} = \Upsilon_R\tilde{N}_{e,t} - f_T\tilde{N}_{e,t}\Upsilon_{q,t}$. Because of the various procedures involved in the current immigration policy, firms have to incur extra regulatory costs. However, firms have to incur lower technological hiring costs because these are applicable only to workers that firms are able to bring to the firm and the presence of the cap leads to a lower number of such workers in the decentralized economy.

In the baseline calibration, the job creation wedge is 13.1348 and the consumption resource wedge is 0.0045, both in units of the consumption basket. Therefore, the job creation inefficiency wedge contributes to the bulk of the inefficiency in the baseline market economy.

8 Market-Driven Allocation of Permits

In this section, I address the following question — what would be the impact of moving toward a market-driven allocation of visas (via auction of a fixed number of permits) for skilled foreign workers, compared to the baseline model with the current skilled immigration policy setup? In particular, would such a policy increase economic welfare and bring the economy's equilibrium closer to the social planner's first-best allocation?

The motivation for the market-driven allocation of permits stems from the advantages of this alternate skilled immigration policy highlighted in Peri (2012).[22] The main idea is that introducing a market-driven system of allocating permits to firms who hire immigrants would introduce a price mechanism to allocate visas efficiently and according to their most productive use. Importantly, the price of permits would be determined by the auction and would quantify the value attributed by the U.S. market to a foreign skilled worker. According to the proposal, the market-driven price of permits would also provide potential flexibility across the business cycle via price feedback and these prices could be a potential signal for raising/lowering total number of permits. For instance, if the price of the permit rises during expansionary times, it would signal a true shortage in the number of permits relative to firm demand for foreign skilled labor. This may be relevant for policy makers as the costs under the current immigration policy setup may distort firm demand of foreign skilled labor, and thus do not always give a true indication of firm demand (Section 5.1). Moreover, such a mechanism may generate additional income for the government, which could help compensate domestic households. The key idea in Peri's proposal is that "a simpler, more flexible, and more market-driven system of labor-sponsored permits for immigrants would enhance the economic benefits of employment-based visas".

8.1 Model with Market-Driven Allocation of Permits

The preferences, technology, and the economic setup are exactly the same as in the baseline model (Section 3). The main difference is that the immigration policy-imposed cost of hiring skilled foreign workers will vary with economic conditions via the optimally varying price of permits. Also, there is no lottery and hence firms get allocated their optimal demand of permits at the market clearing price of permits.

I evaluate the impact of moving from the baseline skilled immigration policy toward an alternate market-driven policy, in a simple framework with no informational asymmetries, and one in which firms bid for permits according to their demand schedule. I initially set the number of permits to be allocated to be the same as the cap imposed under the baseline

[22] Peri (2012) focuses on reform of all immigration policies rather than only on skilled immigration policy.

policy.[23] The timing is as follows: the state of the economy is realized and a fraction δ of foreign skilled workers separate from the domestic labor market. Wages are determined competitively, and firms produce period-t output. The Government announces the period t auction of a fixed mass of permits — each permit would allow firms in Sector 1 to hire a foreign skilled worker that will become productive in period $t+1$. Each firm submits a schedule of permit prices for varying quantities of permits. Since there are no informational asymmetries across firms, they have an incentive to submit no other price and quantity bid other than according to their demand schedule. The government evaluates the permit price at which the total number requested by firms matches the total supply of permits. Firms then pay the government the market clearing price of permits and receive the permits. The government collects the revenue and rebates it back to skilled and unskilled domestic households in a lump-sum and symmetric manner. Firms then hire foreign skilled workers according to the number of permits they hold and also incur the technological hiring costs. There are no secondary market sales for permits.

The optimization problem of firm z for deriving its demand schedule of permits as a function of price of permits is given below. In this derivation, the assumption is that firms anticipate that the number of workers they hire will be equal to the number of permits they own, and they choose their permit demand accordingly.

$$\max_{N^p_{e,t}(z)} E_t \sum_{\tau=t}^{\infty} \beta_{\tau,t} \left[\rho_\tau(z) Z_\tau z l_{s,\tau}(z) - w_{s,\tau} l_{s,\tau}(z) - \zeta^p_\tau N^p_{e,\tau}(z) - f_T N^p_{e,\tau}(z) \right]$$

subject to

1. $l_{s,t}(z) = l^s_{h,t}(z) + l^s_{f,t}(z)$
2. $l^s_{f,t+1}(z) = (1-\delta)(l^s_{f,t}(z) + N^p_{e,t}(z))$

The equilibrium condition for the optimal permit demand is given by:

$$\zeta^p_t + f_T = (1-\delta) E_t \{ \beta (C_{s,t+1}/C_{s,t})^{-1} [\rho_{1,t+1}(z) Z_{t+1} z - w_{s,t+1} + \zeta^p_{t+1} + f_T] \} \tag{6}$$

Goods market clearing in each period t for each firm is given by $Z_t z l_{s,t}(z) = (\frac{\rho_{1,t}(z)}{\tilde{\rho}_{1,t}})^{-\theta} Y^c_t / \tilde{\rho_{1,t}}$, where Y^c_t is the aggregate demand.

Since wages, and therefore prices in period $t+1$ will be a function of permits held in period t (which determines the foreign labor stock and hence firm output in $t+1$), (6) determines the

[23] According to Peri (2012), permits for the H-1B category (as well as the L-1A, L-1B (intra-company transfers) and TN visas (professionals from NAFTA), all included as one category) should be sold in a quarterly electronic auction and the number of permits could initially be set equal to the number of annual average temporary visas issued over the past ten years.

demand schedule for permits. To get the market clearing permit price, the demand schedules of all firms can be aggregated to obtain the following.[24]

$$\zeta_t^p + f_T = (1-\delta)E_t\{\beta(C_{s,t+1}/C_{s,t})^{-1}[\tilde{\rho}_{1,t+1}(z)Z_{t+1}\tilde{z} - w_{s,t+1} + \zeta_{t+1}^p + f_T]\} \tag{7}$$

The equilibrium market-clearing permit price that firms pay is such that the aggregate demand for permits is equal to the aggregate supply i.e. $\tilde{N}_{e,t}^p = \bar{N}_e$, where $\bar{N}_e$ is the exogenous number of permits allocated by the government. In Appendix G, I derive the steady-state demand schedule and the equilibrium price of permits. Figure 6 plots the steady-state aggregate demand schedule of permits and the equilibrium in the permit market. An increase in aggregate productivity Z raises demand by all firms and therefore shifts the aggregate demand schedule to the right. Given an exogenously fixed number of permits, this raises the equilibrium permit price paid by each firm to the government.

For a given price of permits, the permit demand of each firm (and hence for the economy as a whole) is increasing in productivity, unskilled labor available in the economy, the share of the sector 1 good in consumption, and decreasing in δ (which is a rough measure of how long each permit is allowed to be valid)[25], θ (as this influences profit from hiring each foreign skilled worker), and in the availability of domestic skilled workers.

Moreover, since the permit price is endogenous, an increase in the permit price decreases the optimal demand of permits by each firm, and hence for the economy as a whole. Any factor that increases the demand for permits, including aggregate productivity in the economy, would increase the market clearing price of permits, for a given supply. Thus, as economic conditions vary, the market-clearing price of permits would adjust to reflect varying firm demand of foreign skilled labor.

The rest of the equilibrium conditions are similar to the baseline model. Aggregate accounting requires $\tilde{\rho}_{1,t}Z_t\tilde{z}\tilde{l}_{s,t} + \rho_{2,t}Z_t\bar{l}_u = C_{s,t} + C_{i,t} + C_{u,t} + f_T\bar{N}_e$. Consumption of skilled domestic households is given by $C_{s,t} = w_{s,t} + \tilde{d}_t + T_t/2$, where T_t is the government transfer, which is now equal to $\bar{N}_e\bar{\zeta}_t^p$ (the new revenue from immigration policy). Unskilled domestic households are assumed to receive the rest of the transfers, and immigrants consume their labor income, as before. Average firm profits are given by $\tilde{d}_t = \tilde{\rho}_{1,t}\tilde{y}_{1,t}/\theta - \zeta^p\bar{N}_{e,t} - f_T\bar{N}_{e,t}$.

[24]The tilde variables denote aggregates across firms. Aggregation in this model is similar to the baseline model.

[25]I take δ to be the same as in the baseline model even though the interpretation is different. If each permit is allowed to be valid for the same time as current skilled immigration visas, than the rate at which foreign skilled workers leave the firm would roughly be the same as the rate at which each permit lasts for the firm.

8.2 Welfare Implications of Moving Toward the Market-Driven Allocation of Permits

In this section, I describe the welfare implications for each domestic household, including transitional dynamics, of a perfect foresight change in policy from the current immigration policy setup, toward the market-driven allocation of permits. In order to measure the welfare change, I calibrate the model with the alternate immigration policy by setting the sunk regulatory costs f_R and application costs g to zero. The rest of the calibration is the same as in the baseline case. The equilibrium permit price is determined endogenously.

The welfare change is measured as the percentage change (Δ) in consumption that would leave the household indifferent between the baseline skilled immigration policy and the policy with the market-driven allocation of permits for hiring skilled foreign workers:

$$u\left[C_i(1+\frac{\Delta_j}{100})\right] = (1-\beta)\sum_{t=0}^{\infty}\beta^t u(C_{j,t}) \quad \forall\{j \in s, u\}$$

The first impact of the policy change is that government revenue increases by 35.6 percent. This is because the market clearing permit price $\bar{\zeta}_t^p$ that is collected by the government for each of the $\bar{N}_e$ permits is higher than the arbitrarily set application cost g. Since the cap is the same, the revenue collected under this alternate policy will always be higher than the revenue collected under the baseline skilled immigration policy. If this revenue is rebated back to domestic households in a symmetric manner, both households witness a welfare gain as a result of the policy implementation. The welfare gain amounts to 0.1803 percent of annualized steady-state consumption for unskilled workers. For skilled workers, the welfare gain amounts to 0.1770 percent of annualized steady-state consumption.

I next turn to analyzing the inefficiency wedges that would exist in this alternate framework. Since this is a second-best policy relative to the social planner's optimal solution, inefficiency wedges would still exist in this framework. However, since there are welfare gains, this alternate policy would bring the market economy's allocation closer to the social planner's allocation. The inefficiency wedges help identify the mechanisms involved.

In this alternate framework, $\Upsilon_{R,t} = 0$ as this distortion was associated with the baseline immigration policy. However, there is a new distortion associated with the alternate market-driven system – the equilibrium permit price differs from the technological hiring cost in the social planner's framework. This distortion is $\Upsilon_{\zeta,t} = \bar{\zeta}^p$. The new job creation wedge, as derived in the appendix, can be expressed as:

$$\Sigma_{jp,t} = E_t\left[B_{t,t+1}(1-\delta)\left(\rho_{1,t+1}Z_{t+1}\tilde{z}\frac{(\Upsilon_{\zeta^p,t}+f_T\Upsilon_\theta)}{f_T(\Upsilon_{\zeta^p,t}+f_T)} + \frac{\Upsilon_{\zeta^p,t}-\Upsilon_{\zeta^p,t+1}}{\Upsilon_{\zeta^p,t}+f_T}\right)\right] \tag{8}$$

If $\Upsilon_{\zeta^p,t} = \Upsilon_\theta = 0$, then the job creation wedge is equal to 0. The resource constraint wedge in this alternate policy is 0 as the only sunk hiring cost that is not rebated back to households is the technological component of the hiring cost, which is also the case under the social planner's problem. Thus the alternate policy closes the consumption resource constraint wedge.

As for the job creation wedge, on comparing the optimal hiring condition in the baseline case with the optimal permit demand under the alternate policy (equation 7), we see that as long as prices, wages, and the technological hiring costs under the two cases are the same, then it must be that $f_R/q_t + g = \bar{\zeta}_t^p$ i.e. $\Upsilon_{R,t} = \Upsilon_{\zeta^p,t}$. Prices and wages in this case depend only on the state variables in the economy i.e. on the realization of aggregate productivity Z_t and on the stock of foreign skilled labor $\tilde{l}^s_{f,t}$. Since the cap under the two policies is the same, the entry and stock of foreign labor is the same in both cases and therefore prices (and wages) in sector 1 are also the same.[26] In this case, the job creation wedges under the two policies are the same i.e. $\Sigma_{jc,t} = \Sigma_{jp,t}$. Therefore, the mechanism through which this alternate policy increases welfare and brings the market economy closer to the planner's allocation is by closing the consumption resource constraint wedge.

In the current framework, closing the job creation wedge, which is the larger of the two inefficiency wedges in the decentralized economy, requires policy makers to increase the cap. However, even in the absence of a cap increase, easing some of the burdensome immigration policy procedures, and moving towards a simpler market-driven mechanism could potentially bring the market economy's allocation closer to the social planner's allocation. An implication of this is that the recent political discussions on tightening the procedures related to skilled immigration policy in the U.S. would have the opposite impact of widening the gap between the market economy's allocation and the social planner's optimal allocation and therefore would increase inefficiency in the economy.

9 Model Extension: Search and Matching Framework

In the baseline model, there is no unemployment of domestic skilled workers. Yet, one of the political arguments against increasing skilled immigration is the displacement effect of immigrants on domestic workers. Another argument against immigration is that foreign workers are paid lower wages compared to domestic workers. In this section, I extend the model to include search and matching frictions and non-competitive wage-setting, as such a setup can potentially account for these features. I then re-evaluate the welfare impacts of

[26] When the cap does not bind in the baseline model, entry will be lower than the cap. However, in this case, even the optimal permit demand will be lower.

an immigration cap change.

In the framework with search frictions, firms in the skill-intensive sector now post vacancies and they can be matched with either a skilled foreign worker or to a a skilled domestic worker. The probability of getting matched to a domestic or foreign worker depends on the relative fraction of each type of worker searching for jobs. However, there is still a policy-imposed cap and additional costs of hiring foreign workers. For each foreign worker that is matched, firms have to pay immigration policy and technological hiring costs as in the baseline model. Also, if the aggregate number of matches with foreign workers (which is now the aggregate demand for foreign workers) exceeds the cap, there is a probability that each application will be selected, similar to before.

Search and Matching in the skill-intensive sector

Suppose firm z posts $v_t(z)$ vacancies for skilled workers in period t. The cost of posting a vacancy is κ. Given a standard constant returns to scale matching technology with unemployment elasticity ϵ and matching efficiency χ, the probability that the firm will be matched with a skilled worker (domestic or foreign) is given by $\mu_t = \chi(\frac{V_t}{U_t})^{-\epsilon}$, where V_t denotes the aggregate vacancies posted, and $U_t = U_{d,t}+U_{f,t}$ is the aggregate mass of domestic and foreign skilled workers searching for a job. Since the H1-B policy in the U.S. does not allow foreign workers to remain in the U.S. for an extended period if they are unemployed, I interpret $U_{f,t}$ not necessarily as the unemployment of foreign skilled workers in the domestic economy, but instead as the aggregate number of foreign skilled workers who are seeking a job in the domestic labor market, and can be located either in the domestic economy, or abroad.

The market tightness is given by $\frac{V_t}{U_t}$. The probability that the firm is matched to worker type $j \in \{d, f\}$ is $q_{j,t} = \frac{U_{j,t}}{U_{d,t}+U_{f,t}}\chi(\frac{V_t}{U_t})^{-\epsilon}$, where d and f denote domestic and foreign respectively, and $\frac{U_{j,t}}{U_{d,t}+U_{f,t}}$ is the relative share of job searchers of each type. Note that $q_{d,t}+q_{f,t} = \mu_t$. However, firms that match with foreign workers have to pay an additional sunk cost for each foreign worker they apply for, f_R, as well application costs g and technological hiring costs f_T, for foreign workers that they are eventually able to bring to the firm.

The second immigration policy restriction, as before, is the cap on the total number of foreign workers that can be hired each period, $\bar{N}_{e,t}$. Let $N_{e,t}(z) = q_{f,t}v_t(z)$ be the demand of foreign workers (determined by the number of foreign matches) at firm z. Then, the firm will apply for these workers to join the firm and the probability that each foreign worker that was matched would eventually be able to join the firm is $q_t = min[\frac{\bar{N}_{e,t}}{(\int_{z_{min}}^{\infty} N_{e,t}(z)dG(z))}, 1]$, which is endogenously determined. Therefore, if the flow of matches for foreign workers is $q_{f,t}v_t(z)$, the mass of foreign workers that eventually join firm z is $q_tq_{f,t}v_t(z)$. Each firm is of measure 0 and takes q_t as given when making its vacancy decision.

The exogenous separation rate for domestic workers is δ_d, and that of foreign workers is δ_f. As a significant proportion of foreign workers are likely to be temporary workers due to the nature of immigration policy, one can postulate that $\delta_f > \delta_d$ (Battisti et al., 2014). Workers hired this period join the firm in the next period and the separation shock is realized at the beginning of every period. Thus the stock of employed domestic workers at firm z is given by $l_{d,t}(z) = (1 - \delta_d)(l_{d,t-1}(z) + q_{d,t-1}v_{t-1}(z))$, and the stock of employed foreign workers is $l_{f,t}(z) = (1 - \delta_f)(l_{f,t-1}(z) + q_{t-1}q_{f,t-1}v_{t-1}(z))$. The timing is as follows — in each period, first the separations are realized, the aggregate productivity shock is realized, wages are negotiated by a surplus-sharing rule, and firms produce period-t output. Firms then post vacancies and workers are matched. For foreign matches, firms pay the immigration policy costs and submit applications for the workers to join the firm. If the cap binds, only a fraction q_t of the foreign matches are allocated to the firms. Finally, firms pay the technological cost f_T for the workers who are able to join the firm.

There are three households - skilled domestic, unskilled domestic, and skilled foreign. Each household consists of a continuum of workers and the measure of workers that are employed in Sector 1 is determined by the matching process. Let the total measure of domestic skilled workers in the labor force be $\bar{L}_d$ and that of foreign be $\bar{L}_f$ (both fixed to begin with). Then $U_{d,t} = \bar{L}_d - L_{d,t}$ and $U_{f,t} = \bar{L}_f - L_{f,t}$ are the domestic and foreign unemployed/job searchers in each period. $L_{d,t}$ and $L_{f,t}$ are the aggregate domestic and foreign workers that are employed, respectively. Employed and unemployed households of each type pool labor income, as is standard. Thus, the budget constraints are similar to the simple model except that now labor income is earned only by the measure of employed households of each type. Household preferences and optimal consumption choices are exactly the same as in the simple version of the model.

I present details of the optimization problem, the equilibrium conditions, and the wage setting in Appendix H.

The first order condition from firms' optimization problem shows that in equilibrium, the cost of posting a vacancy is equal to the expected discounted surplus from a domestic match plus the expected discounted surplus from a foreign match, both weighed by the probability of each match, net of sunk hiring costs for foreign matches. The surplus from each match is just the additional value generated from a skilled labor net of the real wage paid, plus the continuation value of the match.

From the wage setting equations ((H.5), and (H.6)) derived in the appendix, we see that since domestic and foreign workers are perfect substitutes and contribute equally to production, any differences in wages between the two skilled workers has to be because of potential differences in the bargaining power or outside options of the two workers. Evidence shows that immigrants and domestic workers differ according to their outside options (Chassamboulli and Palivos, 2014), and in their separation rates (Battisti et al., 2014).

Welfare Impact of an Immigration Policy Cap Change in the Search and Matching Framework:
In order to get some intuition regarding how an immigration cap change would impact employment of domestic skilled workers and therefore their welfare in this framework, in the appendix, I derive the steady-state relationship between domestic skilled labor employed by firms, L_d, and the immigration policy cap, $\bar{N}_e$ as:

$$L_d = \bar{L}_d \left(\frac{\bar{L}_f \delta_d}{(1-\delta_d)\bar{N}_e} - \frac{\delta_d(1-\delta_f)}{\delta_f(1-\delta_d)} \right) + 1 \Bigg)^{-1}$$

The equation above shows that as the cap on foreign skilled workers increases, long-run aggregate domestic workers employed, L_d, increases, for a given mass of aggregate domestic and foreign labor in the labor force (i.e. for a given $\bar{L}_d$ and $\bar{L}_f$). Intuitively, the increase in the entry cap increases firms' incentive to post more vacancies as there is a higher probability that a foreign worker that was matched with the firm would eventually be able to join the firm. Another way to see this is from the vacancy posting condition derived in the appendix:

$$\kappa = (1-\delta_d)E_t[B_{t,t+1}\Gamma_{zd,t+1}]q_{d,t} + (1-\delta_f)E_t[B_{t,t+1}\Gamma_{zf,t+1}]q_{f,t}q_t - f_R q_{f,t} - q_{f,t}q_t(f_T + g)$$

Part of the expected benefit from posting a vacancy for firm z is the surplus from a foreign match ($\Gamma_{zf,t}$), as with probability $q_{f,t}$, a firm may be matched with a foreign worker. However, because of the cap on foreign workers, the firm will only be able to retain a fraction q_t of its foreign matches. When the cap increases, the probability of being able to bring the worker to the firm increases as q_t is a positive function of the cap. This increases firms' expected discounted benefit from posting a vacancy and firms end up posting more vacancies. More vacancies posted increases firm matches with domestic skilled workers as well, and therefore, their employment. This effect is captured in Figure 7, which plots the transitional dynamics following a 10 percent perfect foresight cap increase in the search and matching framework.

In order to numerically compute the transitional dynamics following a 10 percent cap increase, I calibrate the model as follows. I choose the matching elasticity ϵ as 0.4 as is standard in the literature (Blanchard and Diamond (1989)), and the bargaining power of both workers, $\eta_d = \eta_f = 0.4$, as the same as ϵ so that the Hosios condition holds. Vacancy posting costs κ are normalized to 1. I normalize the aggregate domestic skilled workers in the labor force ($\bar{L}_d$) to 1 and I calibrate the mass of skilled foreign workers in the labor force ($\bar{L}_f$) to 0.2 in order to match the average ratio of skilled foreign workers to skilled domestic workers in the U.S. labor force. In order to facilitate comparison between the baseline model with no search frictions and this extended model, I take the immigration policy-imposed costs and exogenous return of foreign skilled workers to be the same as in the baseline model. The exogenous separation of domestic skilled workers is set to 0.06 to match the average annual unemployment rate of domestic skilled workers in the U.S. Therefore, foreign skilled workers face a higher separation rate compared to domestic workers which is consistent with Battisti et al. (2014) and Kingi (2015). The outside option of domestic skilled workers is determined endogenously and depends on the job finding probability i. The outside option of foreign workers is taken as zero in the baseline calibration. This is in line with evidence that immigrants and natives differ according to their outside options (Chassamboulli and Palivos, 2014). Moreover, according to the U.S. skilled immigration policy, a foreign skilled worker can only stay for a short duration without being employed.[27]

Figure 7 shows that following an increase in the immigration cap, firms post more vacancies, which increases not only the foreign skilled labor employed, but also the domestic skilled labor employed. Therefore, even though domestic skilled wages fall as in the baseline model, an increase in domestic employment offsets the negative impact on domestic skilled workers and welfare impacts on domestic skilled workers are actually slightly positive (Table 2). Higher vacancies posted also increase the outside option of domestic workers and mitigate part of the decline in skilled domestic wages. As the stock of skilled labor employed in sector 1 builds up, output and profit of sector 1 firms increase. Since a larger skilled labor stock employed implies a higher demand for sector 2 output, unskilled wages increase, similar to the baseline model. Therefore, the impact of an immigration cap increase on the welfare of unskilled natives is still positive. In fact, the welfare gain for unskilled domestic households is twice compared to the welfare gain in the baseline model (Table 2). The other main difference in this case is that the welfare of skilled native households is slightly positive since their employment increases.

This result confirms the fact that firms' endogenous response to the immigration cap change depends on the structure of the labor market. An immigration cap increase can lead to a welfare gain for domestic skilled workers (through employment gains) if labor markets are imperfectly competitive, even if domestic and foreign workers are perfect substitutes. This is in contrast to the case of a perfectly competitive labor market where the welfare of domestic skilled workers unambiguously falls due to a wage decrease. Therefore, it is important to take into account the structure of the labor market when determining the

[27]Since I do not model the foreign economy explicitly in this paper, I do not calibrate the outside options that would exist for foreign workers in the foreign economy. However, positive outside options for foreign workers do not change the results as long as these outside options are lower than those available to domestic workers.

welfare impact of an increase in skilled immigration on domestic workers.

In a separate paper, I study the dynamics, additional welfare implications, and also the distortions that exist in this search and matching framework in greater detail.

10 Conclusion

This paper takes a step in the direction of studying the impact of skilled immigration and skilled immigration policy changes in a macroeconomic general equilibrium framework by explicitly focusing on the role of firm demand for foreign skilled labor. The framework I propose — featuring monopolistically competitive firms and a realistic modeling of the current skilled immigration policy setup — facilitates a better understanding of the determinants of firm demand for foreign skilled labor and the aggregate impacts of changes in current skilled immigration policies. This framework leads to some new insights — the realized state of the economy and the structure of the labor market matter, for evaluating the welfare impact of current immigration policy changes. This is because both of these factors influence how firm hiring of skilled workers changes in response to an immigration cap increase. These insights are lost if we evaluate the impact of a cap increase as a pure labor supply change. Also, even if the government does not want to change the cap on foreign workers, an allocation of the same quota of foreign workers via an alternate mechanism — a market-driven allocation of permits for hiring skilled foreign workers, brings the economy closer to the social planner's optimal allocation. However, an increase in the immigration cap is required to close much of the gap between the decentralized economy and the social planner's equilibrium allocation.

Extensive work remains. Current debates surrounding skilled immigration focus on moving toward a merit-based skilled immigration policy on the lines of Canada. In order to evaluate the impact of this, it is important to include heterogeneity within skill groups. Moreover, it is important to study the implications of firm heterogeneity and the misallocation effect of the current immigration policy across firms in greater detail. I leave this for future work.

Table 1: Baseline Model Summary

Economic Variable	**Equilibrium Condition**
Sector 1 Output	$Y_{1,t} = Z_t \tilde{z}(1 + \tilde{l}^s_{f,t})$
Sector 2 Output	$Y_{2,t} = Z_t \bar{l}_u$
Price Index	$1 = (\tilde{\rho}_{1,t})^{\alpha}(\rho_{2,t})^{1-\alpha}$
Hiring Condition	$f_{R,t}/q_t + g_t + f_{T,t} = (1-\delta)E_t\{\beta(C_{s,t+1}/C_{s,t})^{-1}[\tilde{\rho}_{1,t+1}Z_{t+1}\tilde{z} - w_{s,t+1} + \frac{f_{R,t+1}}{q_{t+1}} + g_{t+1} + f_{T,t+1}]\}$
Hiring Probability	$q_t = min[\frac{\bar{N}_{e,t}}{\tilde{N}_{e,t}}, 1]$
Foreign Labor Stock	$\tilde{l}^s_{f,t+1} = (1-\delta)(\tilde{l}^s_{f,t} + q_t \tilde{N}_{e,t})$
Sector 1 Profits	$\tilde{d}_t = \tilde{\rho}_{1,t} y_{1.t}/\theta - f_{R,t}\tilde{N}_{e,t} - (c_t + f_{T,t})q_t \tilde{N}_{e,t}$
Skilled Wages	$w_{s,t} = \tilde{\psi}_{1,t} Z_t z$
Sector 1 Prices	$\tilde{\rho}_{1,t} = \frac{\theta}{\theta-1}\tilde{\psi}_{1,t}$
Sector 2 Prices	$\rho_{2,t} = w_{u,t}/Z_t$
Aggregate Demand	$Y^c_t = C_{s,t} + C_{u,t} + C_{i,t} + f_{R,t}\tilde{N}_{e,t} + f_{T,t} q_t \tilde{N}_{e,t}$
Market Clearing	$\rho_{1,t} Y_{1,t}/\alpha = \rho_{2,t} Y_{2,t}/(1-\alpha)$
Consumption by Unskilled	$C_{u,t} = w_{u,t}\bar{l}_u + T_t/2$
Consumption by Domestic Skilled	$C_{s,t} = w_{s,t} + \tilde{d}_t + T_t/2$
Consumption by Immigrants	$C_{i,t} = w_{s,t}\tilde{l}_{f,t}$
Government Budget Constraint	$T_t = R_t$
Immigration Revenue	$R_t = g_t q_t \tilde{N}_{e,t}$

Table 2: Summary of Welfare Impacts

Experiment	**Skilled Domestic**	**Unskilled Domestic**
10 percent cap increase: Baseline model	-0.0502%	0.0697%
70 percent cap increase: Baseline model (more binding cap)[a]	-0.8231%	0.8427%
70 percent cap increase: Baseline model (less binding cap)[b]	-0.1353%	0.1396%
Policy change: Baseline policy to alternate policy	0.1770%	0.1803%
10 percent cap increase: Search and matching model	0.0542%	0.1487%

[a]Calibrated to mimic the 2001-2003 policy change
[b]Calibrated to mimic the 2001-2003 policy change

Note: The welfare change of each household from the policy change is computed as the percentage increase in consumption (including transitional dynamics) that would leave the household indifferent between the initial policy and the new policy.

Table 3: Efficient Allocation

Economic Variable	**Equilibrium Condition**
Consumption	$C_{u,t} = C_{s,t}$
Demand for sector 1 output	$Y_{1,t} = \alpha \frac{\varsigma_t}{\nu_t} Y_t^c$
Demand for sector 2 output	$Y_{2,t} = (1-\alpha) \frac{\varsigma_t}{\eta_t} Y_t^c$
Sector 1 output	$Y_{1.t} = Z_t \tilde{z}(1 + L_{f,t}^s)$
Sector 2 output	$Y_{2.t} = Z_t \bar{L}_u$
Aggregate accounting	$C_{u,t} + C_{s,t} + f_{T,t} N_{e,t} = Y_t^c$
Price index	$1 = \left(\frac{\nu_t}{\varsigma_t}\right)^{\alpha} \left(\frac{\eta_t}{\varsigma_t}\right)^{1-\alpha}$
Entry condition of foreign workers	$f_{T,t} = E_t \left[\beta(1-\delta) \frac{C_{s,t}}{C_{s,t+1}} \left(\frac{\nu_{t+1}}{\varsigma_{t+1}} Z_{t+1} \tilde{z} + f_{T,t+1}\right)\right]$
Stock of foreign workers	$L_{f,t+1}^s = (1-\delta)(L_{f,t}^s + N_{e,t})$

Note: $\frac{\nu_t}{\varsigma_t} = \varrho_{1,t}$, and $\frac{\eta_t}{\varsigma_t} = \varrho_{2,t}$ are the relative prices of sector 1 and 2 output in the planner's equilibrium.

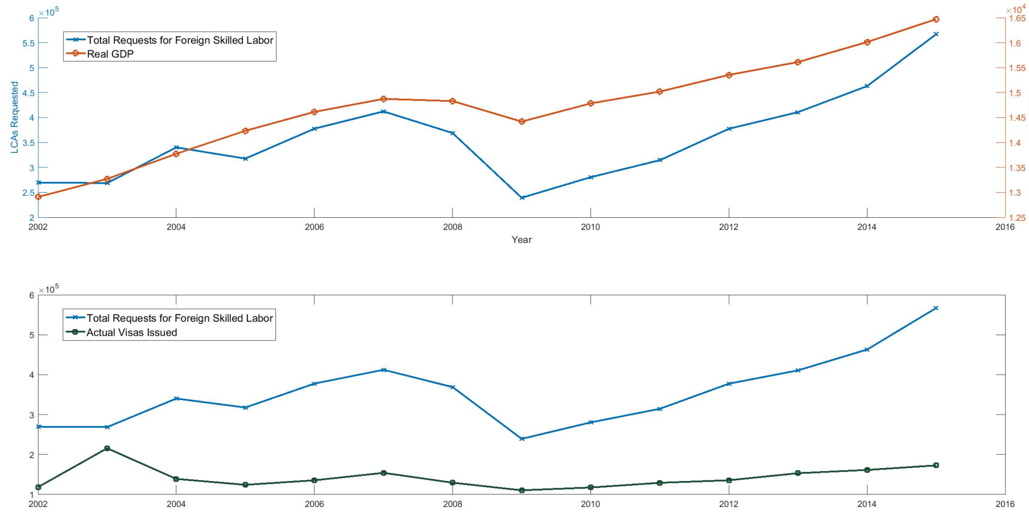

Figure 1: Firm demand of H1-B foreign skilled labor over the business cycle vs actual visas issued. Source: LCA database, Department of Labor. Visa data from the Department of State.

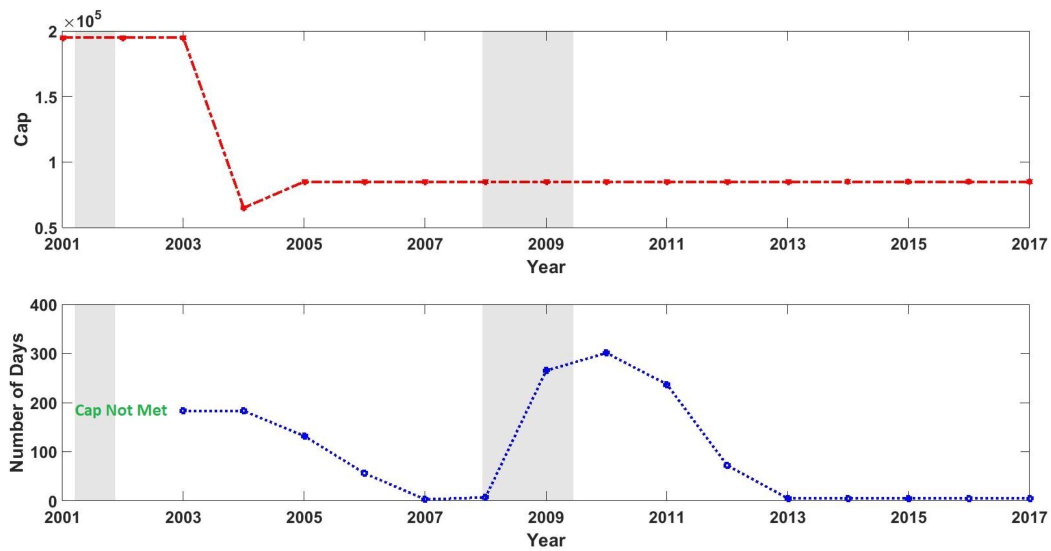

Figure 2: H1-B visa cap (top panel) and number of days in which cap was met (bottom panel)

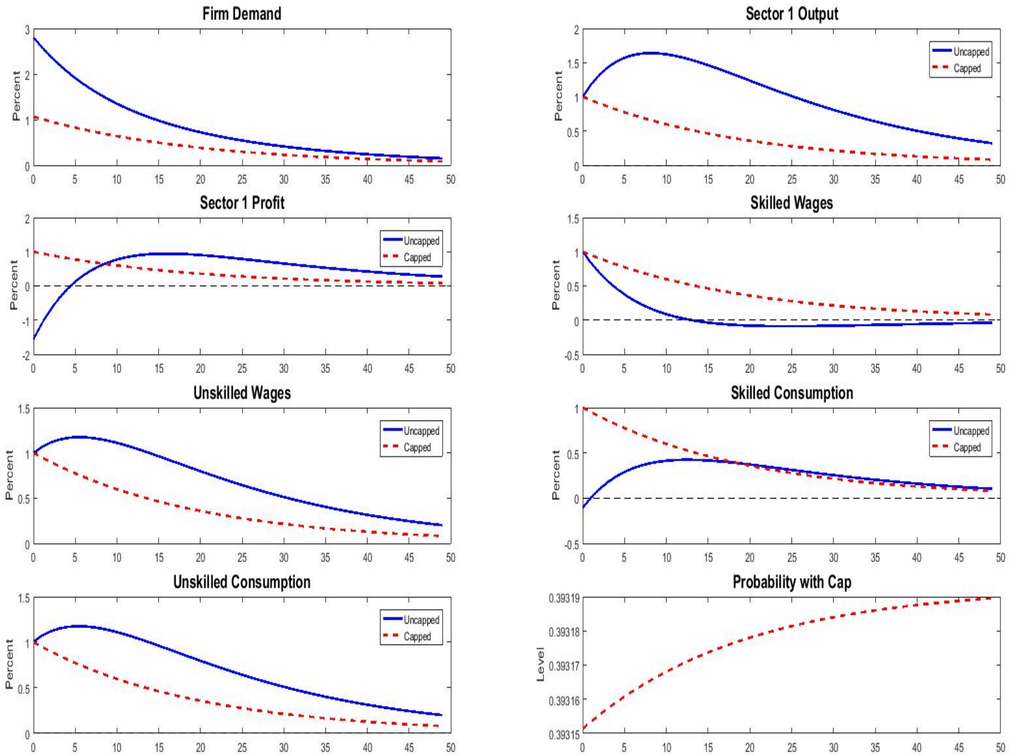

Figure 3: Response to a temporary productivity shock in the presence of an occasionally binding constraint (solid blue line) (baseline case) vs the case without an entry cap (dotted red line). All variables (except probability) are in percent deviation from steady state.

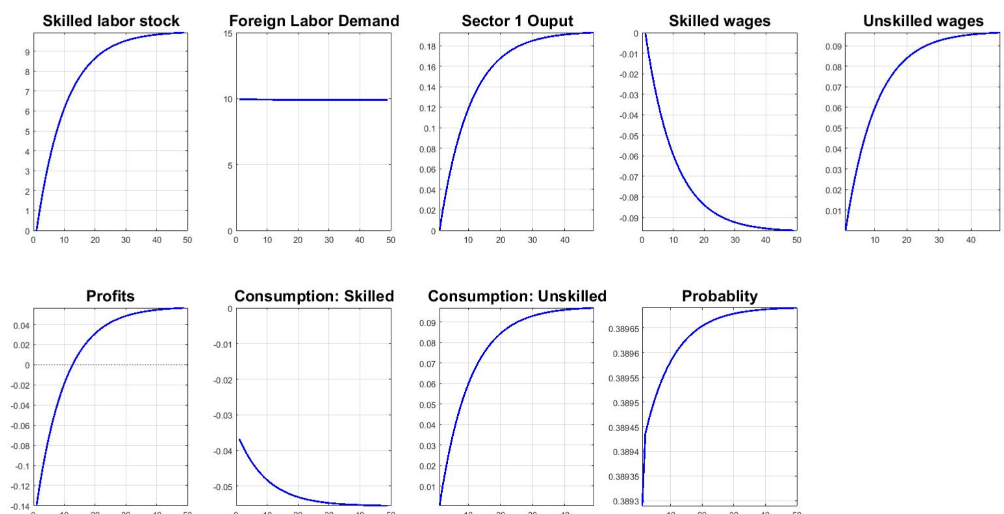

Figure 4: Transitional dynamics after a 10 percent skilled immigration cap increase in the baseline model. All variables (except probability) are in percent deviation from the steady state.

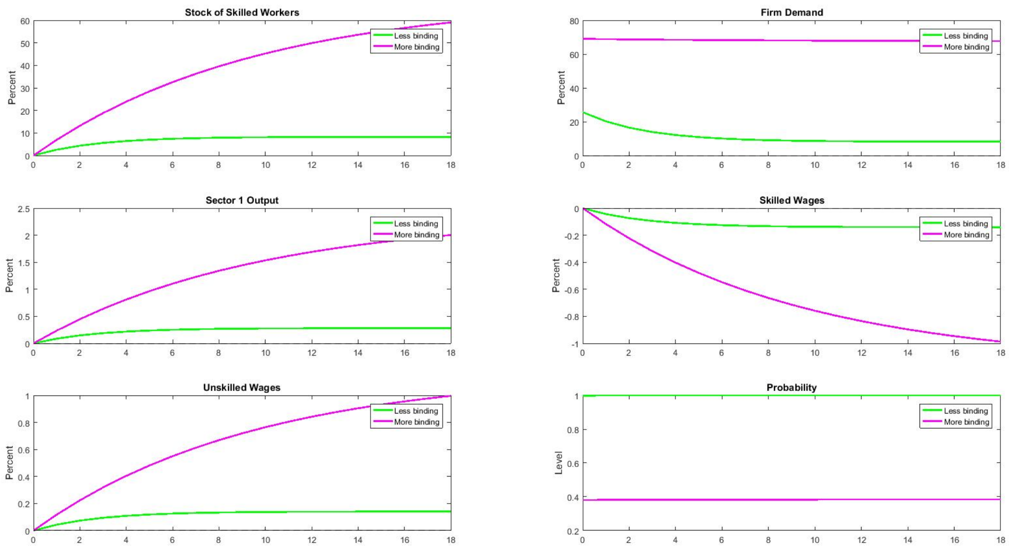

Figure 5: Transitional dynamics after a 10 percent skilled immigration cap change in the baseline model when the cap is 'more binding' vs the case when it is 'less binding'. All variables (except probability) are in percent deviation from the steady state.

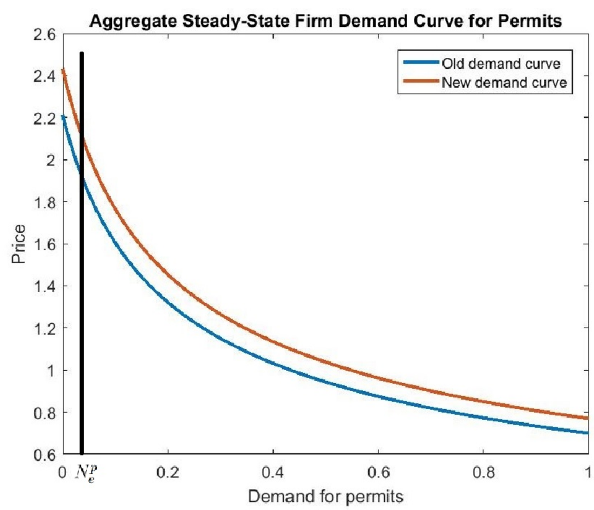

Figure 6: Aggregate steady-state demand schedule for permits in the model with market-driven allocation of permits

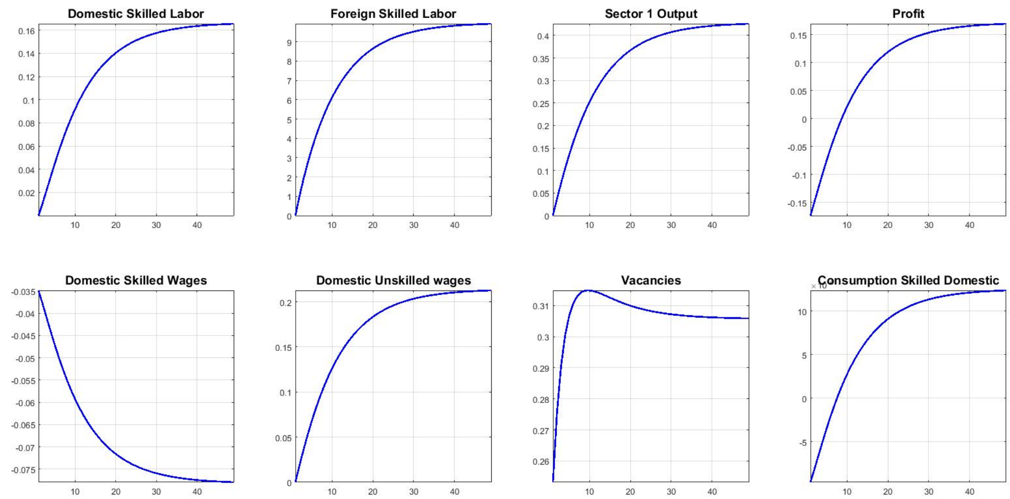

Figure 7: Transition dynamics after a 10 percent skilled immigration cap change in the search and matching model. All variables are in percent deviation from the steady state.

Appendix A Figures

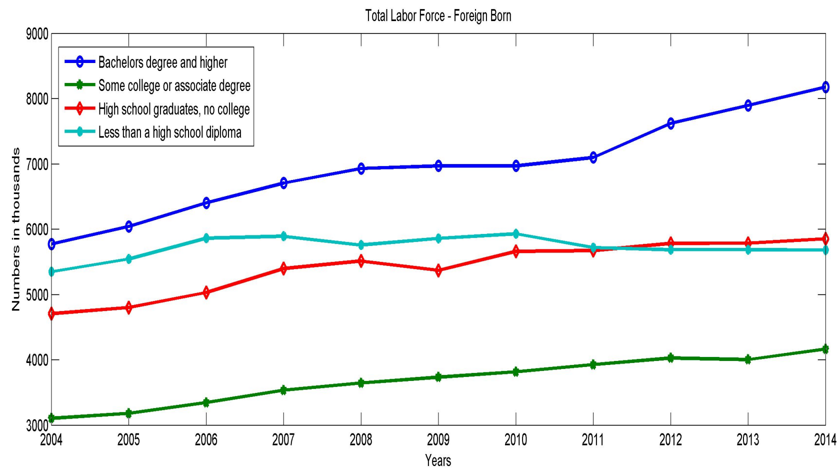

Figure A.1: Source: Current Population Survey (CPS) on foreign-born population in the U.S.

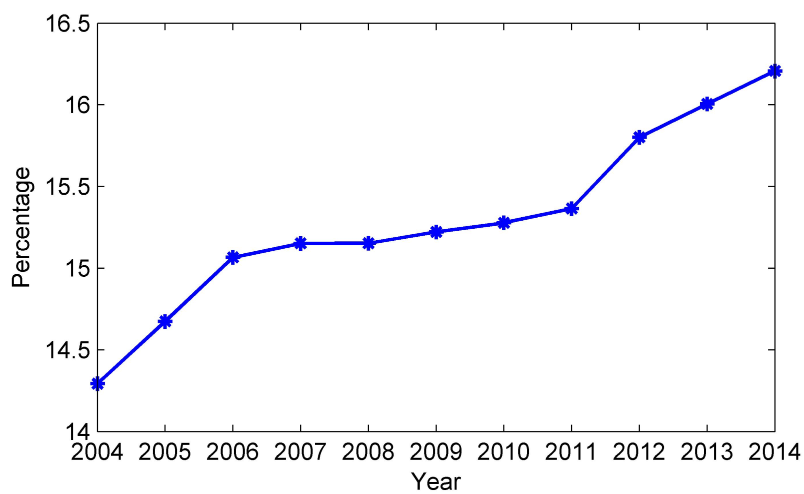

Figure A.2: Proportion of foreign born in the U.S. skilled labor force

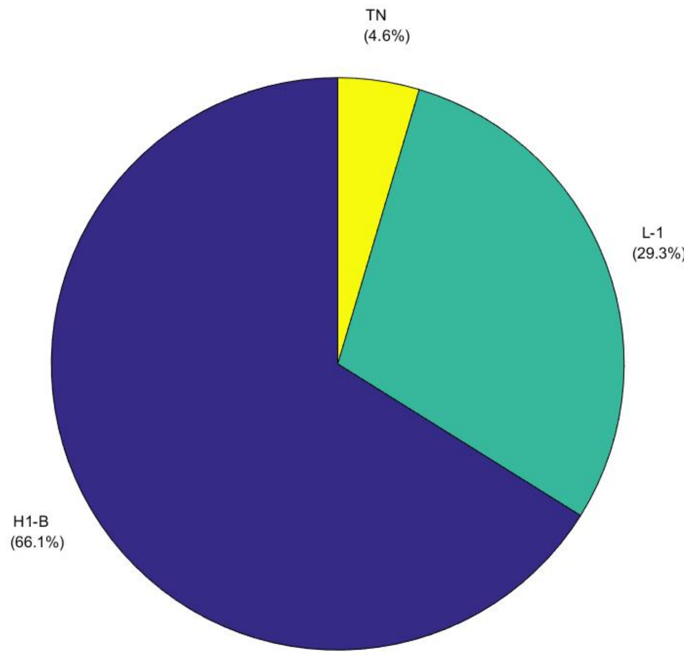

Figure A.3: Entry of foreign skilled workers by visa category

Source: U.S. Department of State.

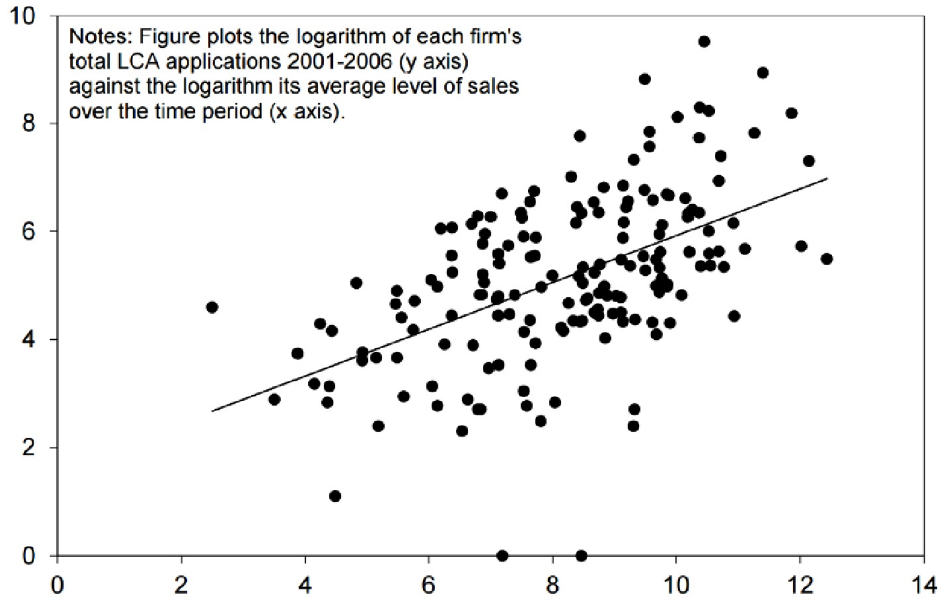

Figure A.4: Firm sales and Labor Condition Applications (LCA)

Source: Kerr et. al. (2014)

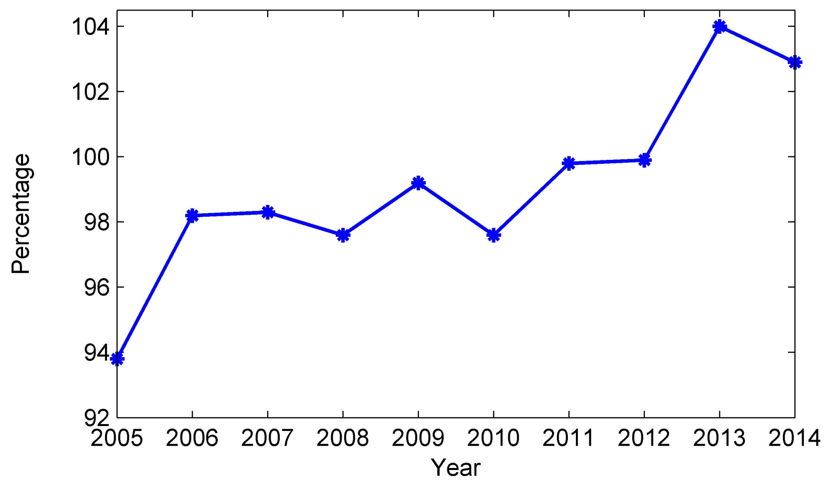

Figure A.5: Earnings of foreign born as percent of native born: Bachelor's degree and higher, 25 years and over

Source: Federal Reserve Bank of St. Louis

Appendix B H1-B Program: Institutional Framework and Background

Since the implementation of the H1-B visa program in 1990, it has been the main method of entry into the U.S. workforce for foreign college educated professionals. Table 6 shows that H1-B visa holders constituted 66 percent of all skilled foreign entrants in 2014. A significant proportion of H1-B recipients (over 70 percent) are from emerging economies — India and China. The other major visa categories for foreign skilled workers are L-1 (for transfer of employees across multinational firms) and TN (visas for Canadian and Mexican NAFTA professional workers). The proportion of entrants from the latter two visa categories has been increasing since 2001, but the H1-B visa program still remains the dominant entry mode. Thus, most studies that analyze the impact of skilled foreign workers in the U.S. focus on the H1-B visa program. Though the H-1B visa is a temporary visa as it is issued for three years (and can be renewed for another three years), it is a dual intent visa as it can lead to permanent residency if the employer is willing to sponsor the worker for a green card.

Table 4: Major Entry Routes for Foreign Skilled Workers (2014)

Visa Category	Proportion of Total
H1-B	66.1 %
L-1	29.3 %
TN	4.6 %

Source: U.S. Department of State

The H1-B 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 for 2011 through 2003. In 2001, cap exemptions were introduced for employees of higher education, non-profit, 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 U.S. institution. These exceptions raise the actual number of visas issued to over 120,000 each year. The cap applies only to new H1-B visa issuances for for-profit firms.

In order to obtain an H1-B visa, there are several steps to be followed and firms are central to this process. The first step requires the firm that wants to hire a foreign worker to file a Labor Condition Application (LCA) with the department of labor. In the application, the firm specifies the nature of occupation and attests that the firm will pay the worker 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.

LCA forms can request for one or more foreign workers for a particular occupation and thus they signal firm vacancies in specific occupations for foreign workers. However, there are some limitations of using the LCA database. The LCA database contains records for every request submitted, but this is only an intermediate step in the process towards the final visa approval. An LCA is submitted for every H-1B request, whether new or a renewal, and each LCA can contain multiple H-1B workers. A more conservative estimate of the demand for foreign skilled workers would be to count each LCA filed as a request for one employee. Plotting the total number of LCAs filed as compared to Figure 1 that plots the total number of employees requested in LCAs filed each year does not change the main motivation regarding business cycle correlation and rising excess demand during expansionary periods.

Once the LCA has been approved by the Department of Labor, it is sent to the United States Citizenship and Immigration Services (USCIS), along with the I-129 form[28] and the required visa fees. The crucial fact is that employees can apply for an H1-B visa only if they have a job offer from an employer with an LCA approval. The employer cannot file more than one I-129 for the same prospective employee. Most of the filing and legal fees are borne by the employer. If the number of H1-B visa petitions (I-129 forms) that fall within the non-exempt category exceed the cap, the USCIS randomly selects visas for processing via a lottery system, until the 65,000 cap is reached. The filing fees paid for the unsuccessful visa applications is returned (unless it is discovered that multiple H1-B petitions are submitted for the same employee). In the last few years, the number of H1-B petitions filed by employers have reached the annual limit within days after the beginning of the filing period which is the first day of April..

[28]This proves the worker's qualifications

Appendix C Model with Complementarities between Domestic and Foreign Skilled Labor

The production technology is now given by: $y_{1,t}(z) = zZ_t l_{s,t}(z)$ where:

$$l_{s,t}(z) = \left(\lambda^{(1-\gamma)}(l^s_{h,t}(z))^\gamma + (1-\lambda)^{(1-\gamma)}(l^s_{f,t}(z))^\gamma\right)^{1/\gamma}.$$

$\gamma = 1 \rightarrow$ gives the baseline model. When $\gamma < 1$, foreign and domestic skilled labor are imperfect substitutes. As γ falls, foreign and domestic skilled labor increasingly become complementary. Now wages of domestic and foreign skilled households are given by:

$$w^s_{h,t} = \psi_{1,t}(z)Z_t z(l_{s,t}(z))^{(1-\gamma)}(l^h_{s,t}(z))^{(\gamma-1)}\lambda^{(1-\gamma)}$$

$$w^s_{f,t} = \psi_{1,t}(z)Z_t z(l_{s,t}(z))^{(1-\gamma)}(l^h_{f,t}(z))^{(-\gamma)}(1-\lambda)^{(1-\gamma)}$$

and the hiring condition of foreign skilled labor is given by:

$$f_R/q_t + g + f_T = (1-\delta)E_t\{\beta(C_{s,t+1}/C_{s,t})^{-1}[\rho_{t+1}(z)Z_{t+1}z(l_{s,t+1}(z))^{(1-\gamma)}(l^f_{s,t+1}(z))^{(\gamma-1)}(1-\lambda)^{(1-\gamma)} - w^s_{f,t+1} + f_R/q$$

Rest of the model is the same as the baseline case with perfect substitutes.

Appendix D Steady-State Solution in the Baseline Model

Since the model features an occasionally binding constraint, the model is equivalent to one with two regimes. The constraint is binding under one regime and slack under the other and each regime has a separate non-stochastic steady state. I first solve for the steady state when the cap is not binding.

When the Cap does Not Bind ($q = 1$):

Here, I derive expressions for the average real prices in sector 1, $\tilde{\rho}_1$, and the average stock of foreign skilled workers, $\tilde{l}^s_f$, and all other endogenous steady-state variables can be expressed as a function of these.

From the market clearing condition in sector 1 we can get: $Z\tilde{z}(1+\tilde{l}^s_f) = \alpha Y^c/\tilde{\rho}_1$, where the aggregate demand $Y^c = C_s + C_i + C_u + (f_R + f_T + g)\tilde{N}_e$. Substituting the households' budget constraint, we can write aggregate demand as $Y^c = w_s(1+\tilde{l}^s_f) + \tilde{d} + T + w_u\bar{l}_u + (f_R + f_T)\tilde{N}_e$.

Consider the average price setting equation in sector 1, $w_s = \frac{\theta-1}{\theta} Z\tilde{z}\tilde{\rho}_1$, the average firm profit in sector 1, $\tilde{d} = \tilde{\rho}_1\tilde{y}_1/\theta - (f_R + g + f_T)\tilde{N}_e$, and the average transfers to households, $T = g\tilde{N}_e$, we can obtain $Y^c = \frac{\theta-1}{\theta} Z\tilde{z}\tilde{\rho}_1(1+\tilde{l}_f^s) + \tilde{\rho}_1 Z\tilde{z}(1+\tilde{l}_f^s)/\theta + w_u\bar{l}_u$, and therefore $Y^c = \rho_1 Z\tilde{z}(1+\tilde{l}_f^s) + w_u\bar{l}_u$.

Then substituting for Y^c in the market clearing for sector 1, we have: $\tilde{\rho}_1 Z\tilde{z}(1+\tilde{l}_f^s) = \alpha(\tilde{\rho}_1 Z\tilde{z}(1+\tilde{l}_f^s) + w_u\bar{l}_u)$. Using expression for real wages in sector 2, $w_u = \rho_2 Z$, and also the price index equation, $\tilde{\rho}_1^{\alpha}\rho_2^{1-\alpha} = 1$, we can write:

$$(1-\alpha)\tilde{z}(1+\tilde{l}_f^s) = \alpha(\tilde{\rho}_1)^{\frac{1}{\alpha-1}}\bar{l}_u \tag{D.1}$$

Consider the steady state aggregate hiring condition which can be written as:

$$(f_R + g + f_T) = \frac{(1-\delta)\beta\tilde{\rho}_1 Z\tilde{z}}{\theta(1-(1-\delta)\beta)} \tag{D.2}$$

Using D.2, we can get the average price of sector 1 output as $\tilde{\rho}_1 = \frac{\theta(f_R+g+f_T)(1-(1-\delta)\beta)}{(1-\delta)\beta Z\tilde{z}}$. Then, using the price setting equation for sector 1, real wages of skilled workers can be written as $w_s = \frac{(\theta-1)(f_R+g+f_T)(1-(1-\delta)\beta)}{(1-\delta)\beta}$. Finally, substituting the expression obtained for average sector 1 prices in D.1, we can get:

$$\tilde{l}_f^s = \frac{\alpha\tilde{z}^{\frac{\alpha}{(1-\alpha)}}\bar{l}_u}{(1-\alpha)}\left[\frac{(1-\delta)\beta Z}{\theta(f_R+f_T+g)(1-(1-\delta)\beta)}\right]^{\frac{1}{(1-\alpha)}} - 1$$

Once we have $\tilde{\rho}_1$ and $\tilde{l}_f^s$, we can obtain the other steady state expressions for $\tilde{N}_e, Y_1, \rho_2, w_u, C_u, C_s, C_i, Y^c$, since these variables can be expressed as a function of $\tilde{l}_f^s$, and $\tilde{\rho}_1$:

$$\begin{aligned}
\rho_2 &= \tilde{\rho}_1^{\frac{\alpha}{\alpha-1}} \\
Y_1 &= Z\tilde{z}(1+\tilde{l}_f^s) \\
\tilde{N}_e &= \delta\tilde{l}_f^s/(1-\delta)
\end{aligned}$$

All other variables can be expressed as a function of these variables.

Steady-State Firm Profit and Skilled Wages as a Function of Skilled Foreign Labor Stock:

Using the expression for average firm profit in sector 1 i.e. $\tilde{d} = \tilde{\rho}_1\tilde{y}_1/\theta - (f_R + g + f_T)\tilde{N}_e$, and substituting $\tilde{\rho}_1 = \left(\frac{\alpha \bar{l}_u}{(1-\alpha)\tilde{z}(1+\tilde{l}_f^s)}\right)^{1-\alpha}$ (from D.1) and $Y_1 = Z\tilde{z}(1+\tilde{l}_f^s)$, we get:

$$\tilde{d} = \left(\frac{\alpha \bar{l}_u}{(1-\alpha)}\right)^{1-\alpha} Z\tilde{z}^{\alpha}(1+\tilde{l}_f^s)^{\alpha}/\theta - (f_R + g + f_T)\delta\tilde{l}_f^s/(1-\delta) \tag{D.3}$$

Also, using the price setting equation for sector 1 and equation D.2, we can obtain steady-state skilled wages as a function of skilled labor stock as:

$$w_s = (\theta - 1)Z\tilde{z}^{\alpha}\left(\frac{\alpha \bar{l}_u}{(1-\alpha)(1+\tilde{l}_f^s)}\right)^{1-\alpha}/\theta \tag{D.4}$$

Steady State When the Cap Binds

When the cap binds, $q\tilde{N}_e = \bar{N}_e$ and therefore from the law of motion for foreign skilled workers, we have $\tilde{l}_f^s = \delta\bar{N}_e/(1-\delta)$. Since the market clearing in sector 1 still needs to hold for the aggregate, we have $\tilde{\rho}_1 = \left(\frac{(1-\alpha)\tilde{z}(1+\tilde{l}_f^s)}{\alpha \bar{l}_u}\right)^{\alpha-1}$. Then, using the optimal hiring condition, we can get: $q = f_R\left[\frac{(1-\delta)\beta(\alpha\delta\bar{l}_u)^{1-\alpha}Z\tilde{z}}{\theta(1-(1-\delta)\beta)((1-\alpha)\tilde{z}(\delta+(1-\delta)\bar{N}_e))^{1-\alpha}} - f_T - g\right]^{-1}$.

Appendix E Social Planner Allocation

Consider the problem of a Social Planner who maximizes welfare of domestic households and chooses the optimal entry of foreign skilled workers in the domestic labor force, taking the firm size distribution, preferences, technology, and resources available in the economy as given. Let f_T be the technologically imposed cost of hiring skilled immigrants in the economy. The planner's problem is given by:

$$\max_{\left\{(C_{u,t}, C_{s,t}, N_{e,t}(z), l^s_{f,t+1}(z), y_{1,t}(z), y_{2,t}\right\}_{t=0}^{\infty}} E_t\left[a\sum_{t=0}^{\infty}\beta^t \ln(C_{u,t}) + (1-a)\sum_{t=0}^{\infty}\beta^t \ln(C_{s,t})\right]$$

s.t.

$$C_{u,t} + C_{s,t} + f_T\int_{z_{min}}^{\infty} N_{e,t}(z)dG(z) = \left(\frac{\int_{z_{min}}^{\infty} y_{1,t}(z)^{\frac{\theta-1}{\theta}}dG(z)^{\frac{\theta}{\theta-1}}}{\alpha}\right)^{\alpha}\left(\frac{y_{2,t}}{1-\alpha}\right)^{1-\alpha} \tag{E.1}$$

$$\int_{z_{min}}^{\infty} y_{1,t}(z)dG(z) = \int_{z_{min}}^{\infty}(Z_t z(l^s_{f,t}(z) + l^s_{h,t}(z)))dG(z) \tag{E.2}$$

$$Y_{2,t} = Z_t \bar{L}_u \tag{E.3}$$

$$\int_{z_{min}}^{\infty} l^s_{f,t+1}(z)dG(z) = (1-\delta)(\int_{z_{min}}^{\infty} l^s_{f,t}(z)dG(z) + \int_{z_{min}}^{\infty} N_{e,t}(z)dG(z)) \tag{E.4}$$

$$\int_{z_{min}}^{\infty} l_{h,t}{}^s(z)dG(z) = \tilde{l}^s_{h,t} = 1 \tag{E.5}$$

where $a \in [0,1]$ is the planner's weight on domestic unskilled households' welfare.

The first constraint is the aggregate resource constraint — the total output can be used for aggregate consumption and incurring the technological component of the sunk cost for hiring foreign skilled workers. The Lagrange multiplier on this constraint, ς_t, represents the social marginal utility of consumption resources. In the Planner's environment, $Y^c_t = C_{u,t} + C_{s,t} + f_T N_{e,t}$.

The second constraint defines production technology for firms in sector 1. The Lagrange multiplier on this, ν_t, denotes the social marginal cost of producing one more unit of sector 1 output. Similarly, the Lagrange multiplier on constraint (E.3), η_t gives the social marginal cost of producing one more unit of sector 2 output.

Constraint (E.4) gives the law of motion of foreign skilled workers. Note that the social planner's constraint does not include the probability of an application being selected, q_t, because the Social Planner does not face a cap. The Lagrange multiplier associated to this constraint, ξ_t, denotes the real marginal value of a foreign skilled worker to society.

The first order conditions are given by:

$$
\begin{aligned}
\frac{a}{C_{u,t}} &= \varsigma_t \quad \forall \text{t} \\
\frac{(1-a)}{C_{s,t}} &= \varsigma_t \quad \forall \text{t} \\
E_t[\beta(\nu_{t+1}Z_{t+1}\tilde{z} + (1-\delta)\xi_{t+1})] &= \xi_t \quad \forall \text{t} \\
-\varsigma_t f_T + (1-\delta)\xi_t &= 0 \quad \forall \text{t} \\
\varsigma_t \left(\frac{Y_{1,t}}{\alpha}\right)^{\alpha-1} \left(\frac{Y_{2,t}}{1-\alpha}\right)^{1-\alpha} \left(\frac{y_{1,t}(z)}{Y_{1,t}}\right)^{\frac{-1}{\theta}} &= \nu_t \quad \forall \text{t} \\
\varsigma_t \left(\frac{Y_{1,t}}{\alpha}\right)^{\alpha} \left(\frac{Y_{2,t}}{1-\alpha}\right)^{-\alpha} &= \eta_t \quad \forall \text{t} \\
C_{u,t} + C_{s,t} + f_T N_{e,t} &= \left(\frac{y_{1,t}}{\alpha}\right)^{\alpha} \left(\frac{y_{2,t}}{1-\alpha}\right)^{1-\alpha} \quad \forall \text{t} \\
Y_{1.t} &= Z_t \tilde{z}(1 + L^s_{f,t}) \quad \forall \text{t} \\
Y_{2.t} &= Z_t \bar{L}_u \quad \forall \text{t} \\
L^s_{f,t+1} &= (1-\delta)(L^s_{f,t} + N_{e,t}) \quad \forall \text{t}
\end{aligned}
$$

where $Y_{1,t} = \int_{z_{min}}^{\infty} y_{1,t}(z)^{\frac{\theta-1}{\theta}} dG(z)$. Also, $L^s_{f,t} = \int_{z_{min}}^{\infty} {l_{f,t}}^s(z) dG(z)$. $L^s_{h,t}$ and $N_{e,t}$ are similarly aggregate values across all firms.

Using $C_{u,t} + C_{s,t} + f_T N_{e,t} = Y^c_t$, and $a = 1/2$ (Assuming that the social planner puts equal weights on skilled and unskilled domestic workers), the first order conditions can be expressed as:

$$
\begin{aligned}
C_{u,t} &= C_{s,t} && \text{(E.6)}\\
Y_{1,t} &= \alpha\frac{\varsigma_t}{\nu_t}Y_t^c && \text{(E.7)}\\
Y_{2,t} &= (1-\alpha)\frac{\varsigma_t}{\eta_t}Y_t^c && \text{(E.8)}\\
1 &= \left(\frac{\nu_t}{\varsigma_t}\right)^{\alpha}\left(\frac{\eta_t}{\varsigma_t}\right)^{1-\alpha} && \text{(E.9)}\\
f_{T,t} &= E_t\left[\beta(1-\delta)\frac{C_{s,t}}{C_{s,t+1}}\left(\frac{\nu_{t+1}}{\varsigma_{t+1}}Z_{t+1}\tilde{z}+f_T\right)\right] && \text{(E.10)}\\
Y_{1.t} &= Z_t\tilde{z}(1+L^s_{f,t}) && \text{(E.11)}\\
Y_{2.t} &= Z_t\bar{L}_u && \text{(E.12)}\\
L^s_{f,t+1} &= (1-\delta)(L^s_{f,t}+N_{e,t}) && \text{(E.13)}\\
C_{u,t}+C_{s,t}+f_TN_{e,t} &= Y_t^c && \text{(E.14)}
\end{aligned}
$$

(E.7) and (E.8) give the demand schedules for sector 1 and sector 2 output. To facilitate the comparison between the planned and decentralized economy, define the following relative prices in sector 1 and 2 for the planner's equilibrium: $\frac{\nu_t}{\varsigma_t}=\varrho_{1,t}$ and $\frac{\eta_t}{\varsigma_t}=\varrho_{2,t}$. Then (E.9) is the price index in the planner's economy.

Equation (E.10), obtained by combining the first order condition w.r.t $N_{e,t}$ with $L^s_{f,t+1}$ is the social planner's entry condition for foreign skilled workers. It shows that the social planner will allow entry of foreign skilled workers till the technological cost of hiring foreign workers is equal to the social expected benefit — the expected discounted value of output produced by an additional foreign skilled worker.

The social planner's equilibrium allocation can be solved using the 9 equations (E.6 - E.14) and 9 variables — $\varrho_{1,t},\varrho_{2,t},y_{1,t},y_{2,t},C_{s,t},C_{u,t},Y_t^c,L^s_{f,t},N_{e,t}$.

Appendix F Distortions and Inefficiency Wedges in the Baseline Model's Decentralized Economy

Comparing the equilibrium conditions in the decentralized economy (Table 1) to those for the planned economy (Table 4) allows us to identify the distortions in the model economy and define inefficiency wedges relative to the efficient allocation.

Entry condition under the planned economy:

$$f_T = E_t \left[\beta_{t,t+1}(1-\delta) \left(\frac{\nu_{t+1}}{\chi_{t+1}} Z_{t+1}\tilde{z} + f_T \right) \right]$$

Entry condition under the decentralized economy:

$$\frac{f_R}{q_t} + f_T + g = E_t \left[\beta_{t,t+1}(1-\delta) \left(\frac{1}{\theta}\tilde{\rho}_{1,t+1} Z_{t+1}\tilde{z} + \frac{f_R}{q_{t+1}} + f_T + g \right) \right]$$

The major distortions in the decentralized economy are[29]:

- Distortion because of the cap: When the cap binds, the probability that a firm will be able to hire a worker is $q_t < 1$. Hence the distortion because of the cap is $\Upsilon_{q,t} = 1 - q_t$.

- Distortion because of immigration policy costs: Let $\Upsilon_{x,t} = f_R + g$. This is the difference between costs that firms face and the technological hiring cost. If there was no cap and the probability of being able to hire a foreign worker is always 1, $\Upsilon_{x,t}$ would be the only distortion w.r.t. immigration policy in this framework. However, when the cap binds and $q_t < 1$ in the decentralized economy, firms face additional costs. To hire one worker, they need to submit $1/q_t$ applications and hence incur f_R/q_t as regulatory costs. Let $\Upsilon_{R,t} = f_R/q_t + g$. Then, simple manipulation yields

$$\Upsilon_{R,t} = \Upsilon_{x,t} + \frac{f_R \Upsilon_{q,t}}{1 + \Upsilon_{q,t}} \tag{F.1}$$

  which decomposes the overall distortion due to immigration policy into distortions due to costs alone and into additional distortions that the cap imposes. Thus, if $\Upsilon_{q,t} = 0$ (no distortion because of the cap), then $\Upsilon_{R,t} = \Upsilon_{x,t}$. However, a binding cap increases the distortions due to immigration policy costs and $\Upsilon_{R,t} > \Upsilon_{x,t}$ is the relevant distortion due to skilled immigration policy costs.

- Distortion because of monopoly power: Monopoly power distorts the job creation decision by inducing a lower return from hiring foreign skilled workers, captured by: $\Upsilon_\theta = 1 - 1/\theta$.

[29]In this model with a C.E.S consumption basket for sector 1, there is no distortion in the relative allocation across firms. This is because a social planner would chose an allocation across firms such that the marginal rate of substitution is equal to the marginal rate of transformation i.e. $z_1/z_2 = (\frac{y(z_1)}{y(z_2)})^{1/\theta}$. This is preserved under the decentralized equilibrium due to market clearing in each firm z: $y_1(z) = \frac{\rho_1(z)}{\rho_1} Y^c/\rho_1$. Therefore, as long as market clearing holds under the decentralized economy, the relative allocation across firms is efficient. Market clearing across firms facilitates aggregation as $\frac{y_1(z)}{y_1(\tilde{z})} = (\frac{z}{\tilde{z}})^\theta$ is the condition that allows us to interpret $\tilde{z}$ as the weighted average of productivity of firms, where weights reflect relative output share of firms.

The major inefficiency wedges are:

- Job Creation Margin: Comparing the entry condition for foreign skilled workers under the decentralized economy with the entry condition under the social planner's equilibrium implicitly defines the inefficiency wedge under the market economy's job creation margin for foreign skilled workers:

$$\Sigma_{jc,t} \quad = \quad E_t\left[B_{t,t+1}(1 \quad - \quad \delta)\left(\rho_{1,t+1}Z_{t+1}\tilde{z}\frac{(\Upsilon_{R,t}+f_T\Upsilon_\theta)}{f_T(\Upsilon_{R,t}+f_T)} \quad + \quad \frac{\Upsilon_{R,t}-\Upsilon_{R,t+1}}{\Upsilon_{R,t}+f_T}\right)\right]$$

  where $\Upsilon_{R,t}$ is defined in equation (F.1) and all variables are evaluated at the decentralized allocations. If $\Upsilon_{x,t} = \Upsilon_{q,t} = \Upsilon_\theta = 0$ (which implies that $\Upsilon_{R,t}$ as well), the job creation wedge is equal to 0.

- Consumption Resource Constraint: Sunk regulatory costs imply a diversion of resources from consumption, leading to a consumption-output efficiency wedge. By comparing the aggregate resource constraints under the decentralized economy with the resource constraint faced by the planner, we get $\Sigma_{r,t} = \Upsilon_{R,t}\tilde{N}_{e,t} - f_T\tilde{N}_{e,t}\Upsilon_{q,t}$.

Appendix G Alternate Policy: Market Driven Allocation of Permits

Steady-State Aggregate Demand Schedule and Equilibrium for Permits

Using the aggregate goods market clearing expression for sector 1, the permit demand equation (7) in the main text, and the law of motion for foreign skilled workers in sector 1, we can obtain the permit price as a function of aggregate permit demand. In steady state this is given by:

$$\zeta^p = \frac{(1-\delta)\beta Z\tilde{z}}{\theta(1-(1-\delta)\beta)}\left(\frac{\alpha\delta\bar{l}_u}{(1-\alpha)\tilde{z}(\delta+(1-\delta)N_e^p)}\right)^{1-\alpha} \tag{G.1}$$

The steady-state equilibrium permit price $\bar{\zeta^p}$ is obtained by substituting $N_e^p = \bar{N}_e$ in equation (G.1).

Inefficiency Wedges in the Alternate Framework:

In this framework, the equilibrium permit price differs from the technological hiring cost in the social planner's framework. This distortion is $\Upsilon_{\zeta,t} = \bar{\zeta}^p$ or the equilibrium price of a permit. Comparing equation (7) in the main text to the optimal entry condition under the social planner's equilibrium (E.10), we can derive the inefficiency wedge under the decentralized economy's job creation margin under the alternate immigration policy with market-driven allocation of permits as:

$$\Sigma_{jp,t} = E_t\left[B_{t,t+1}(1-\delta)\left(\rho_{1,t+1}Z_{t+1}\tilde{z}\frac{(\Upsilon_{\zeta^p,t} + f_T\Upsilon_\theta)}{f_T(\Upsilon_{\zeta^p,t} + f_T)} + \frac{\Upsilon_{\zeta^p,t} - \Upsilon_{\zeta^p,t+1}}{\Upsilon_{\zeta^p,t} + f_T}\right)\right] \tag{G.2}$$

If $\Upsilon_{\zeta^p,t} = \Upsilon_\theta = 0$, the job creation wedge is equal to 0.

The aggregate resource constraint in this decentralized economy is the same as the resource constraint under the planner's economy as $q = f_R = 0$. Therefore, the resource constraint wedge is 0 and the alternate policy closes the consumption resource constraint wedge.

Appendix H Extension to the Baseline Model: Search and Matching Frictions

Firm z's profit maximization problem (taking wages paid to domestic and foreign workers as given)[30] is given by:

$$\max_{\left\{(\rho_{1,t}(z),v_t(z),l_{f,t}(z),l_{d,t}(z)\right\}_{t=0}^{\infty}} E_t\sum_{\tau=t}^{\infty}\beta_{\tau,t}\Big[\rho_{1,\tau}(z)y_{1,\tau}(z) - w_{d,\tau}(z)l_{d,\tau}(z) - w_{f,\tau}(z)l_{f,\tau}(z) - \kappa v_\tau(z) - q_{f,\tau}v_\tau(z)f_R - q_{f,\tau}v_\tau(z)q_\tau(f_T + g_\tau)\Big]$$

s.t.

1. $y_{1,t}(z) = Z_t z(l_{f,t}(z) + l_{d,t}(z))$
2. $l_{d,t}(z) = (1-\delta_d)(l_{d,t-1}(z) + v_{t-1}(z)q_{d,t-1})$
3. $l_{f,t}(z) = (1-\delta_f)(l_{f,t-1}(z) + v_{t-1}(z)q_{f,t-1}q_{t-1})$
4. $y_{1,t}(z) = (\frac{\rho_{1,t}(z)}{\rho_{1,t}})^{-\theta}Y_t^c/\rho_{1,t}$

Let $\Xi_{z,t}$, $\Gamma_{zd,t}$, and $\Gamma_{zf,t}$ be the Lagrange multipliers on constraints 1, 2, and 3 respectively.

[30] I follow Cacciatore (2014) in assuming this.

F.O.Cs can be rearranged to get:

$$
\begin{aligned}
\kappa &= (1-\delta_d)E_t[B_{t,t+1}\Gamma_{zd,t+1}]q_{d,t} + (1-\delta_f)E_t[B_{t,t+1}\Gamma_{zf,t+1}]q_{f,t}q_t - f_R q_{f,t} - q_{f,t}q_t(f_T+g) \\
\Gamma_{zd,t} &= \Xi_{z,t}Z_t z - w_{d,t}(z) + (1-\delta_d)E_t[B_{t,t+1}\Gamma_{zd,t+1}] \\
\Gamma_{zf,t} &= \Xi_{z,t}Z_t z - w_{f,t}(z) + (1-\delta_f)E_t[B_{t,t+1}\Gamma_{zd,t+1}]
\end{aligned}
$$

The first equation implies that in equilibrium, the cost of posting a vacancy is equal to the expected discounted surplus from a domestic match and the expected discounted surplus from a foreign match, both weighed by the probability of each match, net of sunk hiring costs for foreign matches. The second and third equation give the surplus from each match — additional value generated from a skilled labor net of the real wage paid, plus the continuation value of the match.

Nash bargaining of wages: There is no match specific productivity so the surplus to the firm from every domestic match and from every foreign match is the same. However, the surplus from a foreign match may differ from the surplus from a domestic match. Equilibrium surplus sharing rule:

$$\eta_j S^F_{zj,t} = (1-\eta_j)S^W_{zj,t} \quad \forall j \in \{d,f\} \tag{H.1}$$

where η_j is bargaining power of worker j. $S^F_{zj,t}$ is the firm's surplus from a match with worker j, and $S^W_{j,t}$ is the worker's surplus from match with firm z. Workers' surplus is given by:

$$S^W_{zd,t} = w_{d,t}(z) - \varpi_{d,t} + (1-\delta_d)E_t B_{t,t+1} S^W_{zd,t+1} \tag{H.2}$$

$$S^W_{zf,t} = w_{f,t}(z) - \varpi_{f,t} + (1-\delta_f)E_t B^f_{t,t+1} S^W_{zf,t+1} \tag{H.3}$$

where $\beta^f_{t,t+1}$ takes into account that the stochastic discount factor of the firm and foreign skilled households may differ as foreign households are not firm owners in the baseline model. ϖ_{jt} is worker j's outside option. For now, I assume no unemployment benefits. The only outside option for domestic workers is the expected discounted value of searching for other jobs in the next period:

$$\varpi_{d,t} = E_t[\beta_{t,t+1}(1-\delta_d)i_{t+1}\int_{z_{min}}^{\infty}\frac{v_{,t+1}(z)}{V_{t+1}}S^W_{zd,t+1}dG(z)] \tag{H.4}$$

where i_t is the probability of finding a job for a skilled worker i.e. $i_t = \chi(\frac{V_t}{U_{d,t}+U_{f,t}})^{1-\epsilon}$. To begin with, I assume that a foreign worker would not be able to search for jobs and thus their outside option is 0. The rationale is that foreign workers cannot legally stay in the domestic economy beyond a certain point without a valid work visa. However, an alternate exercise with a positive outside option for foreign workers does not change the main implications.

Using $S^F_{zj,t} = \Gamma_{zj,t}$ (no firing costs) and the first order conditions — H.1, H.2, H.3, and H.4. we can get

$$\begin{aligned} w_{d,t}(z) &= \eta_d(\Xi_t Z_t z) + (1-\eta_d)\varpi_{d,t} && \text{(H.5)} \\ w_{f,t}(z) &= \eta_f(\Xi_t Z_t z) + (1-\eta_f)\varpi_{f,t} + (1-\delta_f)E_t[\Gamma_{f,t+1}(\beta_{t,t+1} - \beta^f_{t,t+1})] && \text{(H.6)} \end{aligned}$$

H.6 again takes into account the different stochastic discount factor of foreign workers and domestic firms (an implication of H.3). Since firms face the same costs and the same probabilities of being matched with workers, it can be shown as in Cacciatore (2014) that $\Xi_{z,t} = \Xi_t/z$ (the real marginal cost is symmetric across producers up to firm-specific productivity differentials). This facilitates aggregation as in the standard Melitz (2003) model.

Profit maximization with respect to prices and constraint (4) in the optimization problem gives $\rho_{1.t}(z) = \frac{\theta-1}{\theta}\Xi_{z,t}$, which can be aggregated to give average sector 1 prices: $\tilde{\rho}_{1.t} = \frac{\theta-1}{\theta}\tilde{\Xi}_t$. Aggregate accounting is given by $\tilde{\rho}_{1,t}Y_{1,t} + \rho_{2,t}Y_{2,t} = C_{u,t} + C_{s,t} + C_{i,t} + \kappa V_t + f_R q_{f,t} V_t + f_T q_{f,t} q_t$, where upper case letters and variables with tilde denote aggregate variables.

The summary of conditions can be written as:

$$Y_{1,t} = Z_t \tilde{z}(L_{f,t} + L_{d,t}) \tag{H.7}$$

$$L_{d,t} = (1-\delta_d)(L_{d,t-1} + V_{t-1}q_{d,t-1}) \tag{H.8}$$

$$L_{f,t} = (1-\delta_f)(L_{f,t-1} + V_{t-1}q_{f,t-1}q_{t-1}) \tag{H.9}$$

$$w_{d,t} = \eta_d(\tilde{\Xi}_t Z_t \tilde{z}) + (1-\eta_d)\varpi_{d,t} \tag{H.10}$$

$$w_{f,t} = \eta_f(\tilde{\Xi}_t Z_t \tilde{z}) + (1-\eta_f)\varpi_{f,t} + (1-\delta_f)E_t[\Gamma_{f,t+1}(\beta_{t,t+1} - \beta^f_{t,t+1})] \tag{H.11}$$

$$\begin{aligned}\kappa = {} & (1-\delta_d)E_t[B_{t,t+1}\Gamma_{d,t+1}]q_{d,t} + (1-\delta_f)E_t[B_{t,t+1}\Gamma_{f,t+1}]q_{f,t}q_t - f_R q_{f,t} \\ & - q_{f,t}q_t(f_T + g_t)\end{aligned} \tag{H.12}$$

$$\Gamma_{d,t} = \tilde{\Xi}_t Z_t \tilde{z} - w_{d,t} + (1-\delta_d)E_t[B_{t,t+1}\Gamma_{d,t+1}] \tag{H.13}$$

$$\Gamma_{f,t} = \tilde{\Xi}_t Z_t \tilde{z} - w_{f,t} + (1-\delta_f)E_t[B_{t,t+1}\Gamma_{f,t+1}] \tag{H.14}$$

$$q_{d,t} = \frac{U_{d,t}}{U_{d,t} + U_{f,t}}\chi(\frac{V_t}{U_t})^{-\epsilon} \tag{H.15}$$

$$q_{f,t} = \frac{U_{f,t}}{U_{d,t} + U_{f,t}}\chi(\frac{V_t}{U_t})^{-\epsilon} \tag{H.16}$$

$$q_t = min[\frac{\bar{N}_{e,t}}{\tilde{N}_{e,t}}, 1] \tag{H.17}$$

$$\tilde{\rho}_{1.t} = \frac{\theta - 1}{\theta}\tilde{\Xi}_t \tag{H.18}$$

$$1 = (\tilde{\rho}_{1,t})^{\alpha}(\rho_{2,t})^{1-\alpha} \tag{H.19}$$

$$Y_{2,t} = Z_t \bar{L}_u \tag{H.20}$$

$$\tilde{\rho}_{1,t}Y_{1,t}/\alpha = \rho_{2,t}Y_{2,t}/(1-\alpha) \tag{H.21}$$

$$U_{d,t} = \bar{L}_d - L_{d,t} \tag{H.22}$$

$$U_{f,t} = \bar{L}_f - L_{f,t} \tag{H.23}$$

$$U_t = U_{d,t} + U_{f,t} \tag{H.24}$$

$$\rho_{2,t} = w_{u,t}/Z_t \tag{H.25}$$

$$\tilde{d}_t = \tilde{\rho}_{1,t}Y_{1,t} - w_{d,t}L_{d,t} - w_{f,t}L_{f,t} - \kappa_t V_t - f_R q_{f,t}V_t - (f_T + c)q_{f,t}q_t V_t \tag{H.26}$$

$$\varpi_{d,t} = E_t[\beta_{t,t+1}(1-\delta_d)i_{t+1}\frac{\eta_d}{1-\eta_d}\tilde{\Gamma}_{d,t+1}] \tag{H.27}$$

$$i_t = \chi(\frac{V_t}{U_{d,t} + U_{f,t}})^{1-\epsilon} \tag{H.28}$$

There are 22 equations in 22 endogenous variables: $Y_{1,t}$, $Y_{2,t}$, $L_{f,t}$, $L_{d,t}$, V_t, q_t, $U_{d,t}$, $U_{f,t}$, $\tilde{\rho_{1,t}}$, $\rho_{2,t}$, $q_{d,t}$, $q_{f,t}$, $w_{d,t}$, $w_{f,t}$, $\Gamma_{d,t}$, $\Gamma_{f,t}$, $\tilde{\Xi}_t$, $w_{u,t}$, U_t, d_t, $\varpi_{d,t}$, i_t.

To see how this model relates to the baseline model with no search and matching frictions, note that in the simple version, $q_{d,t} = 0$, $q_{f,t} = 1$, and the matching function is such that matches are formed instantaneously. In other words, $U_{d,t} = 0$ and therefore, $\bar{L}_d = L_{d,t}$. All posted vacancies are filled by foreign workers, $\chi = 1$, and vacancy posting cost $\kappa = 0$ in baseline model.

Steady State for the Search and Matching Model

The steady state for the search and matching model boils down to a system of 8 equations in 8 endogenous variables — L_f, L_d, V, $\tilde{\rho}_1$, q_d, q_f, ϖ_d, and i. I show this for the steady state under the non-binding regime.[31] Using (H.10), (H.11), (H.13), and (H.14), we can write the steady-state surplus obtained from domestic and foreign workers as $\Gamma_d = \frac{(1-\eta_d)(\tilde{\Xi} Z\tilde{z}-\varpi_d)}{(1-(1-\delta_d)\beta)}$, and $\Gamma_f = \frac{(1-\eta_f)(\tilde{\Xi} Z\tilde{z}-\varpi_f)}{(1-(1-\delta_f)\beta)}$. Then substituting these expressions in the steady state hiring condition we get:

$$\kappa = (1-\delta_d)\beta\frac{(1-\eta_d)(\frac{\theta-1}{\theta}\tilde{\rho}_1 Z\tilde{z}-\varpi_d)}{(1-(1-\delta_d)\beta)}q_d + (1-\delta_f)\beta\frac{(1-\eta_f)(\frac{\theta-1}{\theta}\tilde{\rho}_1 Z\tilde{z}-\varpi_f)}{(1-(1-\delta_f)\beta)}q_f q - (f_R+f_T+g)q_f \tag{H.29}$$

Using the market-clearing equation for sector one, and aggregate accounting, we can get (similar to the derivation of (D.1)):

$$(1-\alpha)\tilde{z}(L_d+L_f) = \alpha(\tilde{\rho}_1)^{\frac{1}{(1-\alpha)}}\bar{l}_u \tag{H.30}$$

(H.8), (H.9), (H.15), and (H.16) can be written as:

$$L_d = (1-\delta_d)Vq_d/\delta_d \tag{H.31}$$

$$L_f = (1-\delta_f)Vq_f q/\delta_f \tag{H.32}$$

$$q_d = \frac{\bar{L}_d - L_d}{\bar{L}_d+\bar{L}_f-L_d-L_f}\chi\left(\frac{V}{\bar{L}_d+\bar{L}_f-L_d-L_f}\right)^{-\epsilon} \tag{H.33}$$

$$q_f = \frac{\bar{L}_f - L_f}{\bar{L}_d+\bar{L}_f-L_d-L_f}\chi\left(\frac{V}{\bar{L}_d+\bar{L}_f-L_d-L_f}\right)^{-\epsilon} \tag{H.34}$$

Also, using the surplus sharing rule (H.1) and the outside-option equation (H.4), the outside option of domestic skilled workers can be written as:

[31] For the binding case, there will be 10 equations in 10 variables with $q = \bar{N}_e/(q_f V)$ as an additional equation for determining q.

$$\varpi_d = \beta(1-\delta_d)i\frac{\eta_d}{1-\eta_d}\Gamma_d \tag{H.35}$$

The steady-state job finding probability of skilled workers is given by:

$$i = \chi\left(\frac{V}{\bar{L}_d + \bar{L}_f - L_d - L_f}\right)^{1-\epsilon} \tag{H.36}$$

(H.29), (H.30), (H.31), (H.32), (H.33), (H.34), (H.35), and (H.36) constitute a system of 9 equations in 9 variables.

Relationship Between Domestic Skilled Employment and the Immigration Cap

When the cap binds in steady state, the probability of hiring a foreign skilled worker is given by $q = \frac{\bar{N}_e}{q_f V}$, and therefore from the law of motion of foreign labor, we get that the steady-state foreign labor employed is $L_f = (1-\delta_f)\bar{N}_e/\delta_f$. To see relationship between L_d and $\bar{N}_e$, take the ratio of (H.31) and (H.32), and use $L_f = (1-\delta_f)\bar{N}_e/\delta_f$, and after rearranging terms, we get:

$$L_d = \bar{L}_d\left(\frac{\bar{L}_f\delta_d}{(1-\delta_d)\bar{N}_e} - \frac{\delta_d(1-\delta_f)}{\delta_f(1-\delta_d)}\right) + 1\Big)^{-1} \tag{H.37}$$

i.e. as the cap on foreign skilled workers increases, for a given pool of domestic and foreign labor, aggregate domestic skilled workers employed increases. Intuitively, an increase in the entry cap increases firms' incentive to post more vacancies as there is a higher probability that a foreign worker that was matched would eventually be able to join the firm.

References

Battisti, M., Felbermayr, G., Peri, G., and Poutvaara, P. (2014). Immigration, search, and redistribution: A quantitative assessment of native welfare. Technical report, National Bureau of Economic Research.

Blanchard, O. J. and Diamond, P. A. (1989). The aggregate matching function. Technical report, National Bureau of Economic Research.

Borjas, G. J. (1989). Economic theory and international migration. *International migration review*, pages 457–485.

Borjas, G. J. (2003). The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. Technical report, National Bureau of Economic Research.

Borjas, G. J. and Doran, K. B. (2012). The collapse of the soviet union and the productivity of american mathematicians. Technical report, National Bureau of Economic Research.

Borjas, G. J., Grogger, J., and Hanson, G. H. (2008). Imperfect substitution between immigrants and natives: a reappraisal. Technical report, National Bureau of Economic Research.

Bound, J., Braga, B., Golden, J. M., and Khanna, G. (2015). Recruitment of foreigners in the market for computer scientists in the united states. *Journal of Labor Economics*, 33(S1 Part 2):S187–S223.

Bound, J., Khanna, G., and Morales, N. (2017). Understanding the economic impact of the h-1b program on the us. In *High-Skilled Migration to the United States and its Economic Consequences*. University of Chicago Press.

Cacciatore, M. (2014). International trade and macroeconomic dynamics with labor market frictions. *Journal of International Economics*, 93(1):17–30.

Card, D. (2009). Immigration and inequality. *American Economic Review*, 99(2):1–21.

Chassamboulli, A. and Palivos, T. (2014). A search-equilibrium approach to the effects of immigration on labor market outcomes. *International Economic Review*, 55(1):111–129.

Clemens, M. A. (2012). The effect of international migration on productivity: Evidence from randomized allocation of us visas to software workers at an indian firm.

Doran, K., Gelber, A., and Isen, A. (2016). The effects of high-skilled immigration policy on firms: Evidence from visa lotteries1.

Ehrlich, I. and Kim, J. (2015). Immigration, human capital formation and endogenous economic growth. Technical report, National Bureau of Economic Research.

Ghironi, F. and Melitz, M. J. (2004). International trade and macroeconomic dynamics with heterogeneous firms. Technical report, National Bureau of Economic Research.

Giovanni, J., Levchenko, A. A., and Ortega, F. (2015). A global view of cross-border migration. *Journal of the European Economic Association*, 13(1):168–202.

Guerrieri, L. and Iacoviello, M. (2015). Occbin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics*, 70:22–38.

Hong, G. and McLaren, J. (2015). Are immigrants a shot in the arm for the local economy? Technical report, National Bureau of Economic Research.

Iranzo, S. and Peri, G. (2009). Migration and trade: Theory with an application to the eastern–western european integration. *Journal of International Economics*, 79(1):1–19.

Kerr, S. P. and Kerr, W. R. (2015). Global collaborative patents. Technical report, National Bureau of Economic Research.

Kerr, S. P., Kerr, W. R., and Lincoln, W. F. (2013). Skilled immigration and the employment structures of us firms. Technical report, National Bureau of Economic Research.

Kerr, S. P., Kerr, W. R., and Lincoln, W. F. (2014). Firms and the economics of skilled immigration. In *Innovation Policy and the Economy, Volume 15*. University of Chicago Press.

Kerr, W. R. and Lincoln, W. F. (2010). The supply side of innovation: H-1b visa reforms and us ethnic invention. Technical report, National Bureau of Economic Research.

Kingi, H. R. (2017). Essays in migration and taxation.

Klein, P. and Ventura, G. (2009). Productivity differences and the dynamic effects of labor movements. *Journal of Monetary Economics*, 56(8):1059–1073.

Mandelman, F. S. and Zlate, A. (2012). Immigration, remittances and business cycles. *Journal of Monetary Economics*, 59(2):196–213.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Ottaviano, G. I. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Ottaviano, G. I., Peri, G., and Wright, G. C. (2015). Immigration, trade and productivity in services: evidence from uk firms. Technical report, National Bureau of Economic Research.

Peri, G. (2012). Rationalizing us immigration policy: Reforms for simplicity, fairness, and economic growth. *The Hamilton Project (May 2012)*, page 5.

Peri, G., Shih, K., and Sparber, C. (2013). Stem workers, h1b visas and productivity in us cities. *London: Norface Migration*.

Peri, G., Shih, K., and Sparber, C. (2015). Foreign and native skilled workers: What can we learn from h-1b lotteries? Technical report, National Bureau of Economic Research.

Peri, G., Shih, K. Y., and Sparber, C. (2014). Foreign stem workers and native wages and employment in us cities. Technical report, National Bureau of Economic Research.

Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3):135–169.

Urrutia, C. (1998). On the self-selection of immigrants. *Manuscript, Universidad Carlos III de Madrid*.

Waugh, M. E. (2017). Firm dynamics and immigration: The case of high-skilled immigration. Technical report, National Bureau of Economic Research.