NBER WORKING PAPER SERIES

WHAT IS THE OPTIMAL IMMIGRATION POLICY? MIGRATION, JOBS AND WELFARE

Joao Guerreiro
Sergio Rebelo
Pedro Teles

Working Paper 26154
http://www.nber.org/papers/w26154

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2019, Revised February 2020

We thank Laurence Ales and Chris Sleet as well as participants at the November 2019 CRNYU conference for their comments. Teles is grateful for the financial support of FCT. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research or the Banco de Portugal.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w26154.ack

What is the Optimal Immigration Policy? Migration, Jobs and Welfare
Joao Guerreiro, Sergio Rebelo, and Pedro Teles
NBER Working Paper No. 26154
August 2019, Revised February 2020
JEL No. F22,H21

ABSTRACT

We study the immigration policy that maximizes the welfare of the native population in an economy where the government designs an optimal redistributive welfare system and supplies public goods. We show that when the government can design different tax systems for immigrants and natives, free immigration is optimal. It is also optimal to use the tax system to encourage the immigration of high-skill workers and discourage that of low-skill workers. When immigrants and natives must be treated alike, banning low-skill immigration and allowing free immigration for high-skill workers is optimal. However, there might be no high-skill immigration when heavy taxes are levied on all high-skill workers, both natives and immigrants.

Joao Guerreiro
Department of Economics
Northwestern University
Evanston, IL 60208
jguerreiro@u.northwestern.edu

Sergio Rebelo
Northwestern University
Kellogg School of Management
Department of Finance
Leverone Hall
Evanston, IL 60208-2001
and CEPR
and also NBER
s-rebelo@northwestern.edu

Pedro Teles
Banco de Portugal
R. Francisco Ribeiro 2
1150 Lisboa
Portugal
and Univ Catolica Portuguesa and CEPR
pteles@ucp.pt

1 Introduction

In a televised address aired in 1977, Milton Friedman discussed the change in the U.S. attitude toward immigration. *"Suppose you go around and ask people: the United States, as you know, before 1914 had completely free immigration; [...]—was that a good thing or a bad thing? You will find hardly a soul who will say it was a bad thing. Almost everybody will say it was a good thing. But then suppose I say to the same people: but now what about today, do you think we should have free immigration? 'Oh no,' they'll say, 'we couldn't possibly have free immigration today.' [...] What's the difference? How can people be so inconsistent? Why is it that free immigration was a good thing before 1914 and free immigration is a bad thing today? [...]There is a sense in which free immigration in the same sense as we had it before 1914 is not possible today. Why not? Because it is one thing to have free immigration to jobs, it is another thing to have free immigration to welfare, and you cannot have both. If you have a welfare state, if you have a state in which every resident is promised a certain minimum level of income or a minimum level of subsistence regardless of whether he works or not, produces it or not, well then it really is an impossible thing."*[1]

The question of what is the optimal immigration policy and how it interacts with domestic redistribution programs has become even more important since Friedman's televised address. In both Europe and the United States, immigration policy has become a central political issue that is influencing electoral outcomes (see, e.g., Alesina, Miano, and Stantcheva, 2018).

In this paper, we study the immigration policy that maximizes the welfare of the na-

[1]Milton Frideman "What is America?" (lecture, University of Chicago, Chicago, IL, October 3, 1977). Transcript published in *The Economics of Freedom* (Cleveland: Standard Oil Company of Ohio, 1978). Immigration to the United States was not completely free prior to 1914. In 1882, the U.S. Congress passed the Chinese Exclusion Act, which restricted Chinese immigration. However, these restrictions had a minor effect on immigration flows. The Immigration Acts of 1917 and 1924 ended the era of relatively free immigration. In 1917, a literacy requirement was imposed. Visa requirements and nationwide immigration quotas were imposed in the act of 1924.

tive population in an economy where the government designs an optimal redistributive welfare system and supplies public goods.[2] The provision of public goods is assumed to be non-excludable and subject to congestion. In our theoretical results, we assume that native and workers are perfect substitutes for each skill type. To simplify, we abstract from other externalities associated with immigration.

We start by showing that *free immigration* is optimal, in the sense that there is no role for immigration quotas, in a first-best setting where the government can implement different transfers or taxes for low- and high-skill workers, natives and immigrants. In this case, immigrants are allowed to enter freely as long as they pay a levy that compensates for the congestion they create in the provision of public goods.[3] Immigrants are excluded from the welfare system. They do not receive transfers and do not pay domestic taxes, other than the public-goods congestion charge.

Next, we consider two second-best settings where the government faces Mirrlees (1971)-style information constraints in distinguishing between low- and high-skill workers. In the first setting, the government can discriminate between native and immigrant workers. In the second setting, immigrants cannot be excluded from the welfare system.

We show that free immigration is still optimal as long as the Mirrleesian planner can discriminate between native and immigrant workers. The reason for this result is that it is preferable to affect immigration flows using immigrant-specific taxes rather than quotas because taxes generate revenue. We consider both the cases in which skill types are perfect and imperfect substitutes. If skill types are perfect substitutes, as in the traditional Mirrleesian literature, the optimal immigration policy follows the same principles as in the first best: immigrants are allowed to enter freely as long as they pay taxes that compensate for congestion effects in the provision of public goods. When

[2]While attaching no weight to the welfare of immigrants may be an extreme assumption, the resulting optimal policy is a natural benchmark, since political systems are likely to target the welfare of the native population.

[3]This policy is similar to the one proposed by Gary Becker in Becker and Posner (2009), which involves charging immigrants for the right to enter the country.

low- and high-skill workers are imperfect substitutes, immigration affects the skill premium through general-equilibrium effects.[4] Taxes and subsidies on immigrants that encourage high-skill immigration and discourage low-skill immigration reduce the skill premium, improving the planner's ability to redistribute income from high-skill natives to low-skill natives.[5] The optimal immigration policy is to levy a tax on low-skill immigrants that is higher than their impact on the social cost of providing public goods and a tax on high-skill immigrants that is lower than their impact on the social cost of providing public goods. In our quantitative analysis, we find that the general-equilibrium effects of immigration on the skill premium play an important role in shaping optimal immigration policy.

When discriminating between immigrants and natives is infeasible, free immigration is not optimal and there is a role for immigration quotas. Since the planner wants to redistribute income toward low-skill native workers, and immigrants and natives must be treated alike, the planner chooses to ban low-skill immigration. The reason for this ban is that low-skill immigrants add to the pool of workers who receive transfers that need to be financed with distortionary taxes on high-skill workers. The optimal immigration policy can feature free immigration for high-skill workers. However, these workers may choose not to immigrate when heavy taxes are levied on all high-skill workers, natives and immigrants alike. These results hold regardless of whether immigration has general-equilibrium effects on the skill premium. However, they are reinforced if immigration affects the skill premium because low-skill immigration increases the skill premium, making it harder for the planner to redistribute in favor of low-skill native workers.

Milton Friedman partially anticipated these results in his 1977 address: "*Look at the obvious immediate, practical case of Mexican illegal immigration. Mexican immigration*

[4]The importance of general-equilibrium effects for the design of Mirrleesian tax systems has been emphasized by Stiglitz (1982) and Naito (1999), among others.

[5]A number of empirical studies have shown that low-skill immigration has a positive impact on the skill premium, see, e.g., Borjas, Freeman, and Katz (1992), Topel (1994), and Card (2009).

over the border is a good thing for the illegal immigrants and the United States. But it is only good so long as it's illegal. [...] As long as it's illegal people do not qualify for welfare, for social security, and for all the myriad of benefits that we pour out from our left pocket into our right pocket. As long as they don't qualify, they migrate to jobs" (Friedman, 1978).

Our analysis shows that the ability to exclude immigrants from the welfare system is critical in order for the native population to benefit from free immigration. However, we note several important nuances. Illegal immigration, in the sense of free and untaxed immigration, is not always good. On the one hand, immigration creates congestion in the provision of public goods. On the other hand, when different skill types are imperfect substitutes, low-skill immigration raises the skill premium, reducing the government's ability to redistribute income toward low-skill natives.

Our results are related to the literature on the optimality of production efficiency with Mirrleesian optimal taxation (see, e.g., Atkinson and Stiglitz, 1976). In the absence of general-equilibrium effects, production efficiency is optimal. In our model, this result translates into the optimality of free immigration combined with taxes that correct for congestion effects. In the presence of the general-equilibrium effects emphasized by Stiglitz (1982) and Naito (1999), taxes can affect relative wages. As a result, production efficiency ceases to be optimal. In our model, this result translates into the optimality of levying different taxes on low- and high-skill immigrant workers.

Our results are also related to the literature on the net benefits of immigration (see, e.g., Borjas, 1995). This literature, which abstracts from the implications of immigration for optimal fiscal policy, emphasizes the presence of an "immigration surplus." This surplus is the net benefit of immigration that results from increases in income to non-labor factors such as land. We show that the immigration surplus emerges in a version of our model in which workers are homogeneous, so there is no need to implement redistribution policies, and immigrants are excluded from the provision of public goods.

We illustrate and develop our results further using a calibrated version of our model to compute the optimal U.S. immigration policy for the 1994–2008 period. In this calibration, we allow immigrants of a given skill type to be imperfect substitutes for natives of the same skill type.

Both in the case of unrestricted taxes and in the case of Mirrleesian income taxes with discrimination between native workers and immigrants, we find that optimal total immigration flows are close to those of the data but the composition of optimal immigration differs from the one we observe.

We find an important quantitative role for the effect of immigration policy on the skill premium. Compared to the first-best solution, the case with Mirrleesian income taxes and discrimination of immigrants features higher high-skill immigration and lower low-skill immigration.

When discriminating between natives and immigrants is infeasible, the optimal immigration policy features zero quotas for low-skill immigrants. This finding agrees broadly with Friedman's intuition. Banning low-skill immigration is indeed optimal when the planner seeks to redistribute income toward low-skill natives. High-skill workers are still free to immigrate, but they may be discouraged by high taxes on high-skill workers.

We find that, if the outside options of high-skill immigrants are high, there is no high-skill immigration. In this *domestic redistribution regime*, high-skill workers are heavily taxed in order to redistribute income toward low-skill natives. As a result of these heavy taxes, high-skill immigrants choose not to immigrate. When the outside options of high-skill immigrants are lower, there is high-skill immigration. In this *immigration surplus regime*, high-skill workers pay low taxes. Transfers to low-skill workers are financed with tax collection on the high income from land.

The paper is organized as follows. In Section 2, we present the model. We discuss the properties of the solution with unrestricted taxes in Section 3. Section 4 contains the analysis of Mirrleesian optimal immigration policy with and without the ability to discriminate between immigrants and natives. Section 5 is devoted to the quantitative analysis. Section 6 summarizes the conclusions.

2 The model

We consider a simple static economy inhabited by a continuum of unit measure of workers, which we call *natives*. Native workers are heterogeneous with respect to their labor productivity. We assume that out of the total native population a share $\pi_{n,l}$ are low-skill workers and a share $\pi_{n,h}$ are high-skill workers. Each household is composed of a single worker. For simplicity, we do not consider the possibility of emigration by native workers.[6]

Native workers with ability $a \in \{l, h\}$ derive utility from consumption, $c_{n,a}$, and disutility from supplying labor, $n_{n,a}$. They also benefit from a publicly provided good, G.[7] For simplicity, we assume that the utility function is strictly separable in public-goods consumption,

$$U_{n,a} \equiv u(c_{n,a}, n_{n,a}) + v(G). \tag{1}$$

We make the standard assumptions that the utility function is twice continuously differentiable, strictly increasing in consumption, $u_c > 0$, and government spending, $v_G > 0$, and decreasing in hours worked, $u_n < 0$.[8] We also assume that the utility function satisfies the following *consumption-leisure normality* condition.[9]

Assumption 1 (Consumption-leisure normality condition)**.** *We assume that the utility function satisfies* $u_{cc}/u_c - u_{cn}/u_n \leq 0$ *and* $u_{cn}/u_c - u_{nn}/u_n \leq 0$*, with at least one strict*

[6]See Mirrlees (1982) for a treatment of optimal income taxation with emigration.

[7]The variable G includes only goods and services provided by the government. It excludes transfers such as social security and unemployment insurance. All transfers are included in the tax/transfer function (see equation (5)).

[8]Whenever there is no loss of clarity, we use f_x to denote $\partial f(x,y)/\partial x$ for some function $f(x,y)$.

[9]In our environment, assuming consumption-leisure normality also implies that the utility function verifies the Spence-Mirrlees *single-crossing* condition: $d\left(\frac{-u_n(c,y/w)}{wu_c(c,y/w)}\right)/dw < 0$. This is because $d\left(\frac{-u_n(c,y/w)}{wu_c(c,y/w)}\right)/dw = -\frac{u_n}{wu_c}\left[-\frac{1}{w} + \frac{y}{w^2}\left(\frac{u_{cn}}{u_c} - \frac{u_{nn}}{u_n}\right)\right]$, which, using normality, is strictly negative.

inequality.

A large pool of potential immigrants, indexed by i, stands ready to enter the country. Immigrants can be low- and high-skill workers. We denote the mass of entering immigrants with skill a by $\pi_{i,a}$. After entering the country, immigrants choose how much to consume, $c_{i,a}$, and work, $n_{i,a}$, and obtain the following utility:

$$U_{i,a} \equiv u(c_{i,a}, n_{i,a}) + v\left(G\right). \tag{2}$$

Implicitly, we are assuming that there is no exclusion in the consumption of public goods, that is, all workers derive utility from the total provision of public goods. Immigrants with skill a enter the country only if their utility weakly exceeds their reservation utility, $\overline{U}_a$:

$$U_{i,a} \geq \overline{U}_a, \text{ if } \pi_{i,a} > 0. \tag{3}$$

To simplify, we assume that the outside options are exogenous; that is, these options do not change with immigration flows. This assumption, together with the presence of a large pool of potential immigrants, is appropriate if the country is small relative to the world economy.[10]

Goods production combines native and immigrant labor with a fixed factor (land), L, according to the production function $F\left(L, N_l, N_h\right)$, where total labor of skill type a is the sum of native and immigrant labor supplies, $N_a \equiv \pi_{n,a}n_{n,a} + \pi_{i,a}n_{i,a}$. The aggregate endowment of land is $\overline{L}$. We make the standard assumptions that the production function is strictly increasing, $F_L, F_l, F_h > 0$, strictly concave, and homogeneous of degree one. Furthermore, we assume that production is weakly separable in land, so the stock of land does not affect the skill premium.[11]

Assumption 2 (Weak separability in land)**.** *Assume that F_h/F_l is independent of L.*

[10]This assumption could be relaxed in different ways. An interesting way to endogenize the outside options is to extend the analysis to a more complex multi-country model.

[11]Krusell, Ohanian, Rios-Rull, and Violante (2000) estimate an aggregate production function that includes skilled and unskilled labor. In their formulation—the stock of structures, the analogue of land in our model—does not affect the skill premium.

The economy's resource constraint is given by

$$\sum_{b\in\{n,i\}}\sum_{a\in\{l,h\}}\pi_{b,a}c_{b,a}+\sigma\left(\pi_{i}\right)G\leq F\left(L,N_{l},N_{h}\right), \tag{4}$$

where $\pi_i \equiv \pi_{i,l} + \pi_{i,h}$, $\sigma(\pi_i) \geq 1$, $\sigma'(\pi_i) \in [0,1]$, and $\sigma''(\pi_i) \geq 0$. This function is meant to capture the *congestion effects* of immigration on the provision of public goods. To provide a total of G units of public goods per household, the government must spend $\sigma(\pi_i)G$ units of output. If $\sigma(\pi_i) \equiv 1$ then there are no congestion effects. If $\sigma(\pi_i) = 1 + \pi_i$ there is full congestion in the sense that the cost of providing public goods scales with the total population, which is the sum of natives and immigrants. In our numerical analysis, we consider the case $\sigma(\pi_i) = 1 + \kappa\pi_i$, where κ represents the share of public goods subject to congestion.

Native and immigrant households A worker with skill a chooses consumption and hours of work to maximize utility subject to the budget constraint

$$c_{n,a}\leq w_{a}n_{n,a}+rl_{n,a}-\mathcal{T}_{n,a}\left(w_{a}n_{n,a},rl_{n,a}\right). \tag{5}$$

The worker receives a wage rate, w_a, which depends only on the worker's skill, and pays taxes according to the tax/transfer function $\mathcal{T}_{n,a}$. Native workers with ability a own $L_{n,a}$ units of land. Landowners decide to rent $l_{n,a} \in [0, L_{n,a}]$ units of land to firms, earning a rental rate r. We assume that high-skill workers own more land than low-skill workers, $L_{n,l} \leq L_{n,h}$, and $\pi_{n,l}L_{n,l} + \pi_{n,h}L_{n,h} = \overline{L}$. Assuming that taxes are a function of labor and non-labor income is equivalent to assuming that only labor and land incomes are observable. The wage rate, labor, and land endowment are not observable.

Immigrant households own no land, $L_{i,a} = 0$, and choose consumption and hours worked to maximize utility subject to the budget constraint

$$c_{i,a}\leq w_{a}n_{i,a}-\mathcal{T}_{i,a}\left(w_{a}n_{i,a},0\right), \tag{6}$$

where $\mathcal{T}_{i,a}$ denotes taxes on immigrant workers, which are potentially different from those paid by native workers.[12]

A note on landownership The assumption that land is privately owned and heterogeneous across households is not standard in the Mirrleesian taxation literature. The standard assumptions are either that wealth is publicly owned, as in Werning (2007), or that it is equally distributed across households, as in Kocherlakota (2010). In these models, the worker's skill level is private information. The two assumptions described above are useful because they prevent the government from using wealth holdings to learn the agent's skill type. As a result, the agent's type must be inferred using only the worker's labor income, so the logic of the Mirrlees-taxation model is preserved.

We consider a model in which land is privately owned and unequally distributed. We assume that taxes cannot be directly levied on the endowments of land, but rather on the income derived from it. Since agents can choose not to rent all their land endowment, concealing information is still possible.

Government The government sets up a tax/transfer scheme. For the sake of generality, we write the tax/transfer function to allow for potential discrimination between natives and immigrants, as well as between low- and high-skill workers. The notation also allows taxes to be arbitrary functions of labor and capital income. In the next sections, we discuss the consequences of different restrictions on the ability of the government to discriminate between worker types for the design of optimal immigration policies.

[12]To simplify, we assume that immigrants have no landholdings in the host country so they only receive labor income. We abstract from financial wealth or income which agents can carry across borders. Incorporating these financial flows into the analysis would require a more complex multi-country model.

The government must satisfy the budget constraint:

$$\sum_{b\in\{n,i\}} \sum_{a\in\{l,h\}} \pi_{b,a} \mathcal{T}_{b,a} \left(w_a n_{b,a}, r l_{b,a}\right) \geq \sigma\left(\pi_i\right) G. \tag{7}$$

We also assume that the government can choose the number of immigrants of each type that enter the economy, subject to the participation constraint (3), by imposing immigration quotas $\{\pi_{i,a}\}_a$.

Firms and factor prices The production technology is operated by competitive firms, hiring labor and renting land to maximize profits. The firms' first-order conditions imply that factor prices are equal to their marginal productivities:

$$w_a = \frac{\partial F\left(l, N_l, N_h\right)}{\partial N_a} \equiv F_a\left(L, N_l, N_h\right),$$
$$r = \frac{\partial F\left(l, N_l, N_h\right)}{\partial L} \equiv F_L\left(L, N_l, N_h\right).$$

Because the production function has constant returns to scale, equilibrium profits are zero.

Equilibrium and free immigration We start by defining the equilibrium for a fixed number of immigrants.

Definition 1. *(Equilibrium) For a given number of immigrants of each type,* $\{\pi_{i,a}\}_{a=l,h}$*, a competitive equilibrium consists of allocations* $c_{b,a}$*,* $n_{b,a}$*,* $l_{i,a}$*, factor prices* w_l*,* w_h*, and* r*, and taxes* $T_{b,a}$ *for all* b, a*, such that: (i) given taxes and factor prices, native households maximize their utility (1) subject to their budget constraint (5), and immigrant households who enter the country maximize their utility (2) subject to their budget constraint (6); (ii) firms maximize profits, implying that factor prices are equal to the marginal productivities; (iii) the government's budget constraint is satisfied; (iv) the goods market, labor, and land markets clear: (4), and*

$$N_a = \sum_{b\in\{n,i\}} \pi_{b,a} n_{b,a},$$

$$L = \sum_{a \in \{l,h\}} \pi_{n,a} l_{n,a};$$

and (v) the immigrants' participation constraint, (3), is satisfied.

We say that there is *free immigration of skill type a* if in equilibrium the participation constraint of immigrants of that skill type holds with equality ($U_{i,a} = \overline{U}_{i,a}$) if $\pi_{i,a} > 0$, and the following inequality holds: $U_{i,a} \leq \overline{U}_{i,a}$ if $\pi_{i,a} = 0$. The concept of free immigration is related to the idea of open borders. With free immigration, the government imposes no quotas on immigration. It receives as many immigrants of skill type a as those willing to immigrate.

The government can restrict immigration either by directly limiting $\pi_{i,a}$ or by taxing immigrants to discourage them from moving. While, under free immigration, the government does not restrict $\pi_{i,a}$ directly, the government might still use the income tax schedule to indirectly affect the level of immigration.

We define a *free immigration with no taxes equilibrium* as an equilibrium in which there is free immigration for all skill types and the government does not tax or subsidize immigration. Formally, this is an equilibrium in which immigrants solve the problem

$$U_{i,a} = \max u\left(c_{i,a}, n_{i,a}\right) + v\left(G\right), \quad \text{s.to } c_{i,a} \leq w_a n_{i,a},$$

and if $\pi_{i,a} > 0$ then $U_{i,a} = \overline{U}_{i,a}$. The concept of free immigration with no taxes is useful because it relates to Friedman's views about illegal immigration. It corresponds to a situation in which immigrants are fully excluded from the tax system. They do not pay taxes, nor do they receive transfers. In the following sections, we discuss conditions under which this kind of immigration can be optimal.

2.1 The immigration surplus with homogeneous workers

To build our intuition, it is useful to review the *immigration surplus* discussed by Borjas (1995). This surplus is the benefit from a marginal increase in immigration in an economy with homogeneous workers and lump-sum taxes on natives. This benefit results from the rise in the productivity of the fixed factor owned by natives, which is land in our model.

Native households earn the rents from land, so it is optimal for them to use all their land in production.

Borjas (1995) assumes that the labor supply is exogenous, so household income can be easily computed. The immigration surplus results from the rise in aggregate labor supply generated by the increase in the pool of workers. This higher labor supply reduces domestic wages but increases the productivity of land, a benefit that accrues only to natives.

We assume that workers are homogeneous, so we drop the index a. Each worker supplies one unit of labor inelastically. The aggregate labor supply is $N \equiv 1 + \pi_i$. The production function is then given by $F(\overline{L}, N)$.

Borjas (1995) abstracts from the provision of public goods. Here, we assume that there are pure public goods (i.e., that $G > 0$ and there are no congestion effects). The government finances its spending with lump-sum taxes on native workers: $\mathcal{T}_n \equiv G$ and $\mathcal{T}_i \equiv 0$. The natives' budget constraint is $c_n = w + r\overline{L} - G$. The immigrants' budget constraint is $c_i = w$.

To derive the immigration surplus, we replace factor prices in the native workers' budget constraint and differentiate with respect to π_i:

$$\frac{dc_n}{d\pi_i} = -F_{NN}(\overline{L}, N)\pi_i > 0.$$

Native households always benefit from the rise in production associated with further immigration. As a result, the equilibrium that maximizes native utility has free immigration.

We assume that only natives pay taxes to finance government spending. As it turns out, this is the policy that maximizes the welfare of natives in the *first-best* solution.[13] In

[13]The first-best solution in this case solves $\max u(c_n)$, subject to $c_n + \pi_i c_i = F(\overline{L}, 1 + \pi_i)$ and $u(c_i) \geq \overline{U}_i$. The first-order conditions for this problem imply that immigrants should not be taxed.

this solution, distorting the extensive-margin choice of immigrants is not optimal since immigration increases the land income and consumption of the native population.[14]

The existence of a positive immigration surplus requires a number of restrictive assumptions. First, we have assumed that all natives have the same level of skill. A number of empirical studies have shown that low-skill immigration has a significant impact on the relative wage of low-skill versus high-skill workers, e.g., Borjas, Freeman, and Katz (1992), Topel (1994) and Card (2009). We have also assumed that immigrants are excluded from the "welfare state," in the sense that they are not entitled to transfers or obliged to pay taxes, and that there is no congestion of public goods. Friedman (1978) argues that "free-immigration to jobs" and "free-immigration to welfare" have very different consequences for natives. To discuss these issues, Sections 4 and 5 use a heterogeneous-agent model to evaluate the impact of immigration on the welfare of the native population.

3 Policy with unrestricted taxes/transfers: first best

We start by assuming that the government can implement discriminatory transfers between all household types: low- and high-skill, natives and immigrants. The government's objective is to maximize a weighted average of the utility of the native population. The weight placed by the government on a native agent with ability a is $\omega_a \geq 0$, and the weights are normalized so that $\pi_{n,l}\omega_l + \pi_{n,h}\omega_h = 1$. Social welfare is given by[15]

$$\sum_{a\in\{l,h\}} \omega_a \pi_{n,a} \left[u\left(c_{n,a}, n_{n,a}\right) + v(G) \right]. \tag{8}$$

[14] This result is related to the Henry George Theorem discussed in Arnott and Stiglitz (1979). These authors show that a planner that maximizes the welfare of residents, both natives and immigrants, allows immigrants to enter in order to raise the value of land income up to the point where land income is sufficiently high to finance the provision of pure public goods.

[15] The solution to this planning problem can be interpreted as the result of a version of the probabilistic voting game proposed by Lindbeck and Weibull (1987) in which only natives vote.

To evaluate the consequences of immigration, we consider a two-stage problem. The first stage is to find allocations $\{c_{b,a}, n_{b,a}, l_{b,a}\}_{b,a}$ and government spending G, and land used in production L that maximize welfare (8), subject to the participation constraint of immigrant workers, (3), the resource constraint, (4), using market clearing to replace aggregate labor and land in production.[16] We use $\mathcal{W}$ to denote the maximal welfare for a given level of immigration.

The second stage is to find optimal immigration levels using the condition $d\mathcal{W}/d\pi_{i,a} \leq 0$, which must be satisfied with equality if $\pi_{i,a} > 0$.

Because L only enters the production function, it is optimal to set it as high as possible so that $l_{n,a} = L_{n,a}$ is optimal (and $\sum \pi_{n,a} l_{n,a} = \overline{L}$).

The optimal solution for the consumption and labor of immigrant workers has to satisfy the following conditions:

$$-\frac{u_l\left(c_{i,a}, n_{i,a}\right)}{u_c\left(c_{i,a}, n_{i,a}\right)} = F_a\left(\overline{L}, N_l, L_h\right),$$

$$u\left(c_{i,a}, n_{i,a}\right) + v(G) = \overline{U}_a,$$

where N_a denotes the aggregate labor supply of type a workers. For a fixed $\pi_{i,a}$, this allocation can be implemented by setting a lump-sum transfer on immigrants, $\mathcal{T}_{i,a}\left(y_N, y_L\right) = T_{i,a}$ for all $y_N, y_L \in \mathbb{R}$, with

$$T_{i,a} \equiv F_a\left(\overline{L}, N_l, N_h\right) n_{i,a} - c_{i,a}.$$

The envelope condition with respect to π_i is

$$\frac{d\mathcal{W}}{d\pi_{i,a}} = \lambda\left[F_a\left(\overline{L}, N_l, N_h\right) n_{i,a} - c_{i,a} - \sigma'\left(\pi_i\right) G\right] = \lambda\left[T_{i,a} - \sigma'\left(\pi_i\right) G\right],$$

where λ is the Lagrange multiplier on the resource constraint. Clearly, optimality requires $T_{i,a} = \sigma'\left(\pi_i\right) G$ if $\pi_{i,a} > 0$.

[16] In appendix A.2, we show that these are necessary and sufficient conditions for implementability of the allocations as an equilibrium with unrestricted taxes.

Proposition 1. *Suppose that the government can discriminate between all worker types. Then, the optimal policy imposes an equal lump-sum tax on all immigrant workers to correct for congestion effects in the provision of public goods, but imposes no marginal distortions,*

$$\mathcal{T}_{i,a} = \sigma'(\pi_i)\, G, \quad a = l, h.$$

Free immigration of all skill types is optimal.

Corollary 1. *Suppose further that there are no congestion effects on public-goods provision, $\sigma(\pi_i) = 1$. Then, free immigration with no taxes is optimal.*

This proposition shows that the optimal immigration policy in this case satisfies two conditions: (1) any immigrant should be free to enter the country, which leads to the condition $U_{i,a} = \overline{U}_a$ if immigration is interior; and (2) upon entering the country, immigrant workers only pay taxes that correct for congestion effects. Free immigration with no taxes is optimal only if there are no congestion effects on public goods.

4 Mirrleesian policy

In general, the benchmark model discussed above cannot be implemented when the government cannot discriminate between low- and high-skill native workers. In this case, redistributing across agents requires the use of distortionary taxation. We now study the effects of immigration in a second-best economy with Mirrleesian non-linear income taxation. We consider two cases. In the first case, the government can discriminate between natives and immigrants. In the second case, this discrimination is not possible, so immigrants cannot be excluded from the welfare state either as net recipients or as contributors.

4.1 Mirrleesian policy with discrimination: immigration to jobs

Consider first the case in which the government can distinguish between natives and immigrants but is restricted in the way in which it can redistribute resources between low- and high-skill natives. As in Mirrlees (1971), we allow for arbitrary, non-linear income-tax/transfer functions. These assumptions imply that

$$\mathcal{T}_{n,l}(y_N, y_L) = \mathcal{T}_{n,h}(y_N, y_L) \equiv \mathcal{T}_n(y_N, y_L)$$

for all y_N and y_L, but the government is otherwise unrestricted. We continue to assume that the government can use tax schedules that discriminate between immigrants of different skills.

In appendix A.4, we show that under these assumptions on the tax functions, the participation constraints (3), the resource constraint (4), and the incentive constraints for native workers,

$$u\left(c_{n,l}, n_{n,l}\right) \geq u\left(c_{n,h}, \frac{F_h n_{n,h}}{F_l}\right), \text{ if } l_{n,h} \leq L_{n,l}, \tag{9}$$

$$u\left(c_{n,h}, n_{n,h}\right) \geq u\left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right), \tag{10}$$

are necessary and sufficient conditions to characterize an equilibrium for $\{c_{b,a}, n_{b,a}, l_{b,a}\}_{b,a}$, $\pi_{i,l}$, $\pi_{i,h}$, L, and G. The problem of the government is to maximize welfare, (8), subject to the constraints that describe the implementable set.

Given that production is weakly separable in land, the amount of land in use does not affect the skill premium directly. Land use enters the problem through the production function, and it influences the ability of low-skill native workers to imitate high-skill native workers. If $l_{n,h} > L_{n,l}$, then the maximum land income of low-skill native workers is too low to allow them to imitate high-skill workers. Increasing $l_{n,a}$ has benefits in terms of increasing production and either does not affect incentive constraints or helps remove one constraint if $l_{n,h} > L_{n,l}$. Therefore, it is optimal to use all available land, $l_{n,a} = L_{n,a}$ and $L = \sum_a \pi_{n,a} L_{n,a}$.

Lemma 1. *Suppose that the government can distinguish between natives and immigrants and that the production function satisfies weak separability in land. Then, the optimal plan is such that native workers use all their productive land* $l_{n,a} = L_{n,a}$.

In general, the skill premium, F_h/F_l, is endogenous because it depends on aggregate labor supplies. Standard Mirrlees-style models often assume that different skill types are perfect substitutes in production, differing only in the number of efficiency units produced by each unit of labor. In these settings the skill premium is exogenous.[17] Below, we consider both cases: an exogenous skill premium and an endogenous skill premium.

Assumption 3 (Perfect substitution in skill types)**.** *Assume that the production function can be written as* $F(L, N_l, N_h) = F(L, \theta_l N_l + \theta_h N_h)$ *for scalars* $\theta_l, \theta_h \in \mathbb{R}_+$ *such that* $\theta_h > \theta_l$.

This assumption implies that the skill premium is constant and given by

$$\frac{F_h}{F_l} = \frac{\theta_h}{\theta_l}.$$

In this case, the skill premium is exogenous. If this assumption holds, the implications for optimal immigration policy are the same as in the case of unrestricted taxation. The optimal plan implies the following conditions:

$$-\frac{u_n(c_{i,a}, n_{i,a})}{u_c(c_{i,a}, n_{i,a})} = F_N\left(L, N\right)\theta_a, \quad \text{and} \qquad u\left(c_{i,a}, n_{i,a}\right) + v\left(G\right) = \overline{U}_a,$$

and the envelope condition is

$$\frac{\partial W}{\partial \pi_{i,a}} = \lambda\left[F_N(L, N)\theta_a n_{i,a} - c_{i,a} - \sigma'\left(\pi_i\right)G\right].$$

The optimal plan features free immigration with the possibility of a lump-sum tax to correct for the congestion externality. If there are no congestion effects, $\sigma(\pi_i) = 1$, then

[17] See, for example, Golosov, Kocherlakota, and Tsyvinski (2003), Werning (2007), and Kocherlakota (2010).

more immigration is desirable as long as the immigrants' contribution to production is lower than their consumption (i.e., $F_N(L,N)\theta_a n_{i,a} - c_{i,a} > 0$). It turns out that the optimal immigration policy is free immigration with no taxes. We summarize these results in the following proposition and corollary, which are proved in appendix A.6.

Proposition 2. *Suppose that the government can discriminate between natives and immigrants and that there is perfect substitution in skill types. Then, free immigration of all skill types is optimal. The optimal policy imposes an equal lump-sum tax on all immigrant workers to correct for the congestion externality, and no further distortions are imposed:*

$$\mathcal{T}_{i,a}(\cdot) = \sigma'\left(\pi_i\right)G.$$

Corollary 2. *Suppose further that there are no congestion effects on public-goods provision, $\sigma\left(\pi_i\right) = 1$. Then, free immigration with no taxes is optimal.*

These results can be interpreted as optimality of production efficiency.[18] Immigration can be interpreted as a technology, and, in that sense, free immigration with no taxes corresponds to production efficiency provided there are no externalities. If there are congestion effects on public goods, production efficiency requires lump-sum taxation of immigrants.

Atkinson and Stiglitz (1976) show that production efficiency is optimal in a Mirr–leesian setting. Proposition 2 is an application of this principle. This application relies crucially on two assumptions. First, immigration has no general-equilibrium effects on relative wages. This assumption is important because it implies that immigration does not affect the incentive constraints. Second, immigrants can be excluded from the welfare system. In what follows, we analyze the case in which immigration has general-equilibrium effects on wages. The case of no exclusion from the welfare system is analyzed in the next section.

[18]See Diamond and Mirrlees (1971), Atkinson and Stiglitz (1976), and Scheuer and Werning (2018).

Stiglitz (1982) analyzes optimal Mirrleesian taxation in a model with general-equilibrium effects, resulting from imperfect skill substitutability. He shows that the optimal plan involves marginal subsidies to high-skill workers and larger marginal taxes on low-skill workers. This tax configuration induces larger relative high-skill labor supply, which reduces the skill premium. Naito (1999) extends this analysis to optimal commodity taxation and concludes that the same rationale implies deviations from uniform taxation. In our setting, the general-equilibrium effects of immigration on wages can lead the optimal immigration policy to deviate from production efficiency. To analyze the impact of these general-equilibrium effects on the optimal immigration policy, we make the following assumption.

Assumption 4 (Skill-premium monotonicity). *Assume that F_h/F_l is strictly increasing in N_l and strictly decreasing in N_h.*

Under the assumption of skill-premium monotonicity, by increasing the aggregate supply of high-skill labor and decreasing the aggregate supply of low-skill labor, the planner can reduce the skill premium. This result implies that the planner can use immigration policy to affect the composition of the labor force and improve income redistribution in the economy.

Proposition 3. *Suppose that the government can discriminate between natives and immigrants and that the skill premium is endogenous and satisfies the skill-premium monotonicity condition. Suppose further that the incentive constraint of high-skill workers, (10), binds and that of low-skill workers, (9), does not bind. Then, free immigration of all skill types is optimal. However, the taxes paid by high-skill workers are lower than those required to correct for congestion effects:*

$$\mathcal{T}_{i,h}(w_h n_h, 0) < \sigma'(\pi_i)\, G,$$

and taxes paid by low-skill workers are higher than those required to correct for congestion effects:

$$\mathcal{T}_{i,l}(w_l n_l, 0) > \sigma'(\pi_i)\, G.$$

If the planner has a strong incentive to redistribute to low-skill workers, then the incentive constraint of high-skill workers binds and that of low-skill workers does not. In that case, the planner has an incentive to reduce the skill premium to loosen the binding incentive constraint. In order to decrease the skill premium, the optimal immigration policy gives relatively less taxes to high-skill workers than to low-skill workers. By affecting the extensive margin choice in this way, the planner incentivizes more high-skill immigrants and fewer low-skill immigrants to enter the country. The optimal plan delivers a shift in the composition of the labor force toward a bigger share of high-skill workers.

The marginal subsidies for high-skill workers and higher marginal taxes for low-skill workers in Stiglitz (1982) are also part of the optimal tax/transfer system in our model. Both native and immigrant high-skill workers are subsidized on the margin in order to induce higher labor supply and reduce the skill premium. For the same reason, low-skill native and immigrant workers are subject to higher marginal taxes.

Even in the case in which low- and high-skill workers are perfect substitutes, the optimality of production efficiency requires the ability of the tax system to discriminate between immigrants and natives. This discrimination is important because it excludes immigrants from the welfare system and gives the government the ability to redistribute income only toward low-skill native workers. It also allows the government to incentivize high-skill immigrants to enter the country while imposing heavy taxes on high-skill natives. The next subsection studies the case in which immigrants cannot be excluded from the welfare system.

4.2 Mirrleesian policy without discrimination: immigration to welfare

In this section, we assume that the government cannot condition taxes on immigration status. The planner must set the same tax/transfer function for all worker types: $\mathcal{T}_n(y_N, y_L) = \mathcal{T}_i(y_N, y_L) \equiv \mathcal{T}(y_N, y_L)$. To avoid the uninteresting case in which the government would distinguish between native and immigrant workers on the basis of landownership, we assume that $L_{n,l} = 0$. This assumption means that immigrant workers can feasibly imitate low-skill workers. As a result, the present model can capture the trade-offs emphasized by Friedman (1978).

In appendix A.8, we show that the set of implementable allocations is constrained by the participation constraint of immigrants with ability a if $\pi_{i,a} > 0$, (3), the resource constraint, (4), and the following incentive constraints:

$$u(c_{b,a}, n_{b,a}) = \max_{(b',a') \in \Theta_{b,a}} u\left(c_{b',a'}, \frac{F_{a'} n_{b',a'}}{F_a}\right),$$

for all (b, a), where $\Theta_{b,a} \equiv \{(b', a') : \pi_{b',a'} > 0 \text{ and } l_{b',a'} \leq L_{b,a}\}$.

Each worker type has potentially three incentive constraints, resulting in twelve incentive constraints in total. However, the next lemma shows that we can simplify the analysis. Intuitively, it shows that because low-skill workers face the same productivity and tax/transfer function, the optimal plan features the same consumption and labor supply for low-skill natives and immigrants: $c_{n,l} = c_{i,l}$ and $l_{n,l} = l_{i,l}$.

The lemma also shows that, as before, because the production function is weakly separable in land, the optimal plan features full land use.

Lemma 2. *Suppose that the government cannot distinguish between natives and immigrants and low-skill native workers own no land. Suppose further that the production function is weakly separable in land. Then, the optimal plan is such that high-skill native workers use all their productive land, $l_{n,h} = L_{n,h}$, and both native and immigrant low-skill workers receive the same consumption-labor bundle, $c_{n,l} = c_{i,l}$ and $n_{n,l} = n_{i,l}$.*

With the simplification provided by this lemma, the problem can be reduced to maximizing welfare subject to the resource constraint, the participation constraint for the immigrant of skill type a, and one of the following two sets of incentive constraints.

If $\pi_{i,h} > 0$, the following incentive constraints must be satisfied: the high-skill-native incentive constraint,

$$u\left(c_{n,h}, n_{n,h}\right) \geq u\left(c_{i,h}, n_{i,h}\right),$$

the high-skill-immigrant incentive constraint,

$$u(c_{i,h}, n_{i,h}) \geq u\left(c_l, \frac{F_l n_l}{F_h}\right),$$

and the low-skill incentive constraint,

$$u(c_l, n_l) \geq u\left(c_{i,h}, \frac{F_h n_{i,h}}{F_l}\right).$$

The first and second conditions combined also imply that the high-skill native worker does not want to mimic a low-skill worker.

If $\pi_{i,h} = 0$, the only relevant incentive constraint is the one of high-skill natives mimicking low-skill workers:

$$u(c_{n,h}, n_{n,h}) \geq u\left(c_l, \frac{F_l n_l}{F_h}\right).$$

Low-skill workers cannot mimic high-skill natives because they have no land endowment.[19]

Proposition 4. *Suppose that the government cannot distinguish between natives and immigrants and that there is perfect substitution in labor types. Then, in the optimal plan, either:*

1. *Low-skill workers receive no net transfers:* $c_l = w_l n_l - \sigma'(\pi_i)G$*; or*

2. *The government bans low-skill immigration:* $\pi_{i,l} = 0$ *and* $c_l > w_l n_l - \sigma'(\pi_i)G$.

This proposition implies that, depending on the welfare weight attached to low-skill workers, it might be optimal to either not have a welfare system and allow free immigration of low-skill workers or to have a welfare system and ban immigration of low-skill workers.

Unlike in the case with discrimination, when it is optimal to have a welfare system, low-skill immigration is not desirable even when the reservation utility of the potential

[19]See appendix A.10 for a proof of the sufficiency of these constraints.

immigrants is low. The participation constraint of low-skill immigrants may not bind because these immigrants have the same utility as low-skill natives.

Intuitively, if the government wants to redistribute resources toward low-skill workers (the relatively poor), then low-skill immigrants reap the benefits of this redistribution. As a result, the government finds it optimal to ban low-skill immigration.

The proposition pertains to the extreme case in which low- and high-skill workers are perfect substitutes. In this case, when the tax system can discriminate based on immigration status, as in the previous section, there is no reason to deviate from production efficiency. This case is useful because it makes clear that, even when the skill premium is exogenous, production efficiency is not optimal when immigrants cannot be discriminated from natives.

If the planner wants to redistribute to low-skill workers, and the incentive constraint of high-skill workers is binding, an endogenous skill premium reinforces the previous result because it reduces the desirability of low-skill immigration.

5 Optimal immigration policy: a quantitative exercise

5.1 Baseline calibration

In this section, we discuss the features of optimal immigration policy in a calibrated version of the model. For this purpose, we consider the following production function:

$$F\left(L, N_l, L_h\right) = AL^{\gamma}\left[\left(1-\alpha\right)N_l^{\frac{\rho-1}{\rho}} + \alpha\left(SN_h\right)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}(1-\gamma)}.$$

Skill-biased technical change is represented by the parameter S, which increases the productivity of high-skill workers relative to low-skill workers.

For simplicity, our theoretical results assume that immigrant and native workers are perfect substitutes. However, a large literature, including Grossman (1982), Manacorda, Manning, and Wadsworth (2012), and Card (2005, 2009), finds that natives and immigrants are not close substitutes. To bring our results closer to this literature, we assume that aggregate labor supply for skill type a is itself a CES aggregator of immigrant and native labor supplies:

$$N_a = \left[(\pi_{n,a} n_{n,a})^{\frac{\varepsilon-1}{\varepsilon}} + \varphi (\pi_{i,a} n_{i,a})^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}.$$

Here, ε controls the elasticity of substitution between immigrants and natives, while φ is a parameter that controls the native-immigrant relative wage.

Perfectly competitive firms hire each labor types and rent land to maximize profits, implying that the price of each factor equals its marginal productivity.

For given labor supplies, we define the skill premium, $\mathcal{SP}$, as the average hourly wage of high-skill workers relative to that of low-skill workers. Defining aggregate labor supply by natives and immigrants of skill type a, $N_{n,a} \equiv \pi_{n,a} n_{n,a}$ and $N_{i,a} \equiv \pi_{i,a} n_{i,a}$, respectively, we can write

$$\mathcal{SP} \equiv \frac{\left(N_{n,h} w_{n,h} + N_{i,h} w_{i,h}\right)/N_h}{\left(N_{n,l} w_{n,l} + N_{i,l} w_{i,l}\right)/N_l} = \frac{\alpha}{1-\alpha} S^{\frac{\rho-1}{\rho}} \left(\frac{N_l}{N_h}\right)^{\frac{1}{\rho}}. \tag{11}$$

The direct effect of an increase in S is to increase the skill premium. The native skill premium is given by:

$$\frac{w_{n,h}}{w_{n,l}} = \mathcal{SP} \times \left(\frac{N_{n,l}}{N_{n,h}}\right)^{\frac{1}{\varepsilon}}. \tag{12}$$

The immigrant wage gap for skill type a, is given by:

$$\frac{w_{i,a}}{w_{n,a}} = \varphi \left(\frac{N_{n,a}}{N_{i,a}}\right)^{\frac{1}{\varepsilon}}.$$

These expressions make clear that changes in relative labor supplies have an impact on the skill premia. In particular, an increase in the supply of high-skill labor relative to low-skill labor decreases the skill premium. This property means that the general-equilibrium effects that underlie the results in Stiglitz (1982), among others, are also present here. In our model with Mirrleesian taxation, these general-equilibrium effects make low-skill immigration relatively less desirable than high-skill immigration. By restricting low-skill immigration and incentivizing high-skill workers to enter the country, the government can reduce the skill premium and improve redistribution. As we have emphasized, the optimal immigration policy no longer involves production efficiency. How and by how much should production efficiency be distorted becomes a quantitative question to which we now turn.

Calibration of status-quo economy We consider a sequence of static economies to match different features of the period between 1994 and 2008.[20] Our status-quo economy is an equilibrium with taxes and government spending. We summarize the calibration in table 1 below.

Using the IPUMS-Current Population Survey (IPUMS-CPS) database,[21] we compute the shares of native and immigrant, low- and high-skill workers in the total population for this period. In Figure 1, we normalize the native population to one in each period and look at the empirical counterparts of $\pi_{n,a}$ and $\pi_{i,a}$ for each a.[22]

We assume that preferences are separable and isoelastic. Consistent with the findings discussed in Chetty (2006) and Chetty, Guren, Manoli and Weber (2011), we set the consumption elasticity to unity and the Frisch elasticity to 0.75; that is,

$$u\left(c,n\right)=\log\left(c\right)-\zeta\frac{n^{1+\nu}}{1+\nu},$$

with $\nu=4/3$. The labor disutility parameter is set so that low-skill households work, on average, one-third of their time endowment, $\zeta=11.06$.

[20] We focus on the pre-financial crisis period starting in 1994 because of the availability of data about immigrants of different skills in the CPS.

[21] Flood, King, Rodgers, Ruggles, and Warren (2018).

[22] We remove high-frequency variation on changes in population shares by working with the fitted values of a quadratic time trend, which we fit using a least squares procedure.

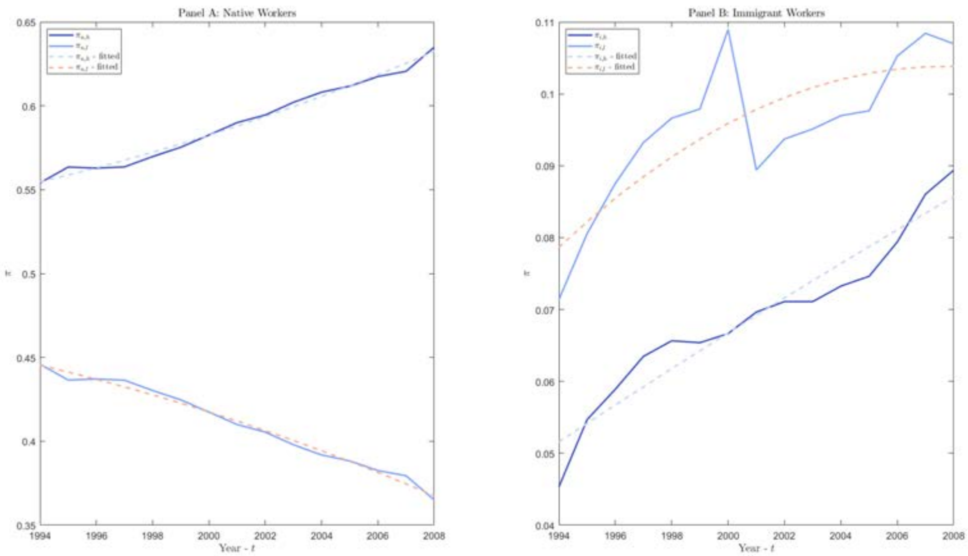

Figure 1: Native and immigrant low- and high-skill worker shares

The tax/transfer function is assumed to be the same for all worker types and to take the same form as in Benabou (2000) and Heathcote, Storesletten, and Violante (2017) for labor and land income:

$$\mathcal{T}(wn, rl) = wn + rl - \lambda(wn + rl)^{1-\tau}.$$

We assume that the government maintains a constant spending-to-GDP ratio. We use the time series in Jaimovich and Rebelo (2017) to calibrate this ratio. We set $v(G) = \chi \log(G)$ and choose χ so that, on average, the marginal utility of spending and the weighted-average marginal utilities of consumption are equated.

We follow the method proposed by Ferriere and Navarro (2019) to obtain estimates of tax progressivity from the NBER TAXSIM data. Finally, we let λ adjust to maintain a balanced budget.

We set $\sigma(\pi_i) = 1 + \kappa\pi_i$. In this formulation $\kappa \in [0, 1]$ is interpreted as the share of government spending subject to congestion effects. In the calibration, we set this parameter equal to the ratio of non-military spending to total spending in each year.

Table 1: Model calibration

Parameters	Description	Value	Source/Target
Time period		1994 – 2008	
Preferences			
ν	Inverse-Frisch elast.	4/3	Chetty (2006)
ζ	Labor disutility	11.06	$n = 1/3$
χ	Preference for G	0.2026	$\chi = (\sum_a \omega_a \pi_{n,a}/c_{n,a})\sigma(\pi_i)G$
Production			
A	TFP	1	Normalization
$\overline{L}$	Land endowment	1	Normalization
γ	Land share	0.05	H&V (2008)
ρ	Skill elasticity	3	A&A (2011), Card (2009)
ε	Nat-Immig. elasticity	20	Card (2009)
φ	Immig. wage gap	0.89	Card (2005)
α	High-skill share	0.66	U.S. skill premium
S	SBTC	*Time Series*	U.S. skill premium
Population			
$\pi_{n,a}$	Share natives skill a	*Time Series*	*CPS*
$\overline{U}_{i,h}$	Outside opt. - h	*Time Series*	Free immigration status quo
$\overline{U}_{i,l}$	Outside opt. - l	*Time Series*	Skill premium *LAC-7*
Government			
κ	Congestion	0.93	Non-military spending
ω_h	High-skill weight	1	Utilitarian planner

Consistent with the estimates in Acemoglu and Autor (2011) and Card (2009), we set ρ, the parameter that controls the elasticity of substitution between low- and high-skill workers, to 3. We choose the skill-biased technical change parameter S so that the baseline economy replicates the skill premium in Acemoglu and Autor (2011).[23,24]

[23]A vast literature (e.g., Katz and Murphy (1992), Autor, Katz and Krueger (1998), and Acemoglu and Autor (2011) shows that the skill premium has been increasing in the United States over the past four decades.

[24]We remove high-frequency variation on the skill premia by working with the fitted values of a

Panel A of figure 2 shows the calibrated time series for skill-biased technical change.

Consistent with the estimates in Card (2009), we set the elasticity of substitution between immigrants and natives to 20. We choose φ so that our status-quo economy replicates an average immigrant wage gap of 0.89 (on average immigrants have 11 percent lower wages relative to natives), which is consistent with the evidence in Card (2005).

We assume that high-skill workers own all the land in the economy, $L_{n,l} = 0$, and normalize the aggregate labor endowment and the total factor productivity parameter to one. Consistent with the findings in Herrendorf and Valentinyi (2008), we set the land share in production to 5 percent (i.e., $\gamma = 0.05$).

To recover the outside options of immigrants, we assume that there is free immigration for high-skill workers. The outside option for these immigrants is their equilibrium utility. To the extent that the actual immigration policy does not feature free immigration of high-skill immigrants, this calibration is an upper bound on the value of the outside option. In a robustness exercise, we consider lowering the outside options for all immigrants.

Because we do not assume that there is free immigration for low-skill occupations, we cannot recover the outside option of low-skill immigrants from their equilibrium utility. Instead, we assume that the outside option scales with the foreign skill premium. We use the time-series data for the skill premium in Latin America from De la Torre, Yeyati, and Pienknagura (2013)[25] and assume that

$$\overline{U}_{i,l} = \overline{U}_{i,h} - \log\left(\frac{w_h^*}{w_l^*}\right).$$

This approximation is consistent with the assumption of logarithmic utility in consumption. Implicitly, we assume that all other benefits abroad (other than labor income)

quadratic time trend, which we fit using a least squares procedure.

[25] De la Torre, Yeyati, and Pienknagura (2013) document that the skill premium started to fall in the late 1990s for a group of seven Latin American countries (LAC-7). This group includes Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Uruguay.

scale with the wage. Panel B of figure 2 shows the calibrated time series for the outside options.

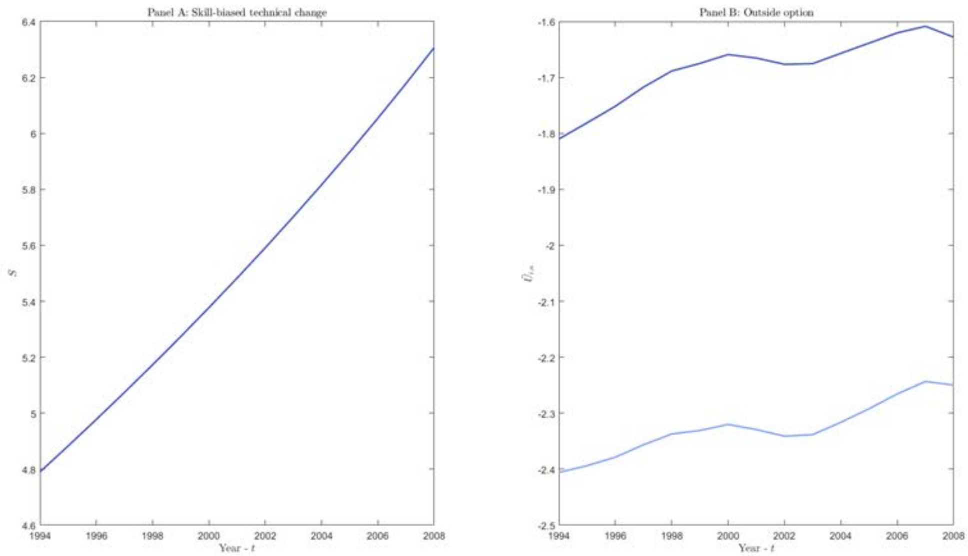

Figure 2: Calibration: outside options and skill-biased technical change

Policy with unrestricted taxes/transfers: first best We now consider the unrestricted taxes benchmark. The government maximizes a utilitarian welfare function with equal weights, $\omega_h = \omega_l = 1$, subject to the resource constraint and participation constraints of immigrant workers.

Figure 3 shows consumption, labor, domestic skill premium, and optimal immigration flows in this case. The optimal level of immigration is such that immigrant workers have the same utility upon entering the country as their outside option; that is, their participation constraint binds. The composition of the immigrant population is such that the behavior of the native-skill premium approximates that of the foreign skill premium.

Immigrant high-skill workers have high outside options, so the government must assign them a good consumption-labor bundle to convince them to immigrate. Indeed, high-skill immigrants consume more and work less than high-skill natives. This property reflects the fact that the status-quo economy, which we use to infer the outside options of high-skill immigrants, features relatively low redistribution and therefore the equilibrium utility for high-skill workers is relatively high. Because preferences are separable and the Pareto weights are symmetric, low- and high-skill native workers consume the same amounts, but high-skill workers work more hours.

We find that the total level of immigration is close to that in the data, but the composition of the immigrant population is very different from the one we observe. Immigration starts at 0.008 for low-skill immigrants and 0.13 for high-skill immigrants per native worker. The planner chooses a high level of high-skill immigration because high-skill workers are more productive than low-skill workers, and they are relatively close substitutes.

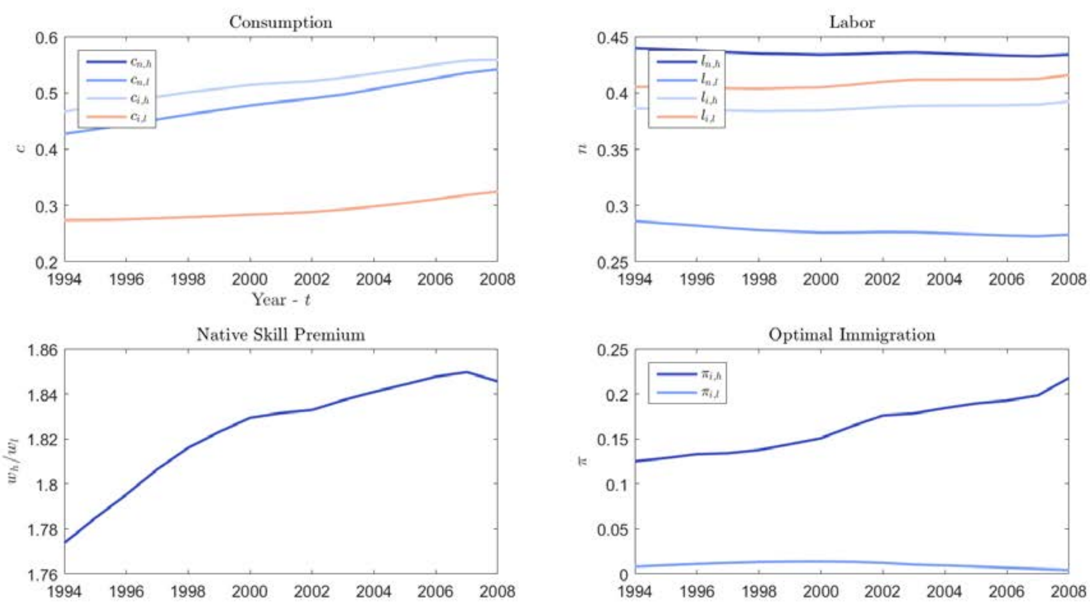

Figure 3: Policy with unrestricted taxes/transfers

Mirrleesian policy with discrimination: immigration to jobs We now consider the case in which taxation can discriminate based on immigration status but not on the native workers' skill.

Figure 4 shows consumption, labor, domestic skill premium, and optimal immigration flows for this case. The Mirrleesian policy with discrimination features higher consumption for high-skill native workers relative to low-skill native workers. This result is a consequence of the need to provide incentives for high-skill workers to work more than low-skill workers.

Relative to the case with unrestricted taxes, there is more high-skill immigration and less low-skill immigration. On average, high-skill immigration increases by 0.02, a 15 percent increase relative to the first best, while low-skill immigration decreases by 0.005, a 47 percent decline relative to the first best. This pattern is consistent with our theoretical results. To reduce the skill premium, the planner uses tax policy to encourage high-skill immigration and discourage low-skill immigration.

In our baseline calibration, the change in the skill premium with respect to the first-best allocation is small. We can decompose the change in the skill premium into three components: the effects of changing native labor supply, the changes in immigrant labor supply, and the consequences of the change in the composition of the labor force.[26] Interestingly, labor supply changes end up having a positive effect on the skill premium when compared to the case without immigration. If the shares of immigrants were the same as in the previous case, the skill premium would have risen by 4.23 percent because of changes in relative labor supplies. The changes in the labor supply of immigrants have a negligible effect on the skill premium. Changes in the composition of the labor force overcome the changes in relative labor supplies, inducing a fall in the skill premium. Indeed, in this decomposition, changes in the composition of the labor force alone account for a 1.3 percent change in the skill premium.

We can also compare this solution with a Mirrleesian optimal policy that does

[26] We discuss this decomposition in appendix A.12.

not allow for immigration. In our current calibration, this comparison is particularly simple because, as we show in the next section, the Mirrleesian optimal plan without discrimination features zero immigration for all skill types. Compared to that allocation, the optimal plan with discrimination delivers an average reduction in the skill premium of 5.2 percent. As in the comparison with the first-best allocation, labor supply changes end up having a positive effect on the skill premium when compared to the case without immigration. If there were no immigration, the skill premium would be 1.3 percent higher because of changes in relative labor supplies. These changes are mostly driven by native workers. The fact that there are more high-skill than low-skill immigrants induces a 6.5 percent fall in the skill premium.

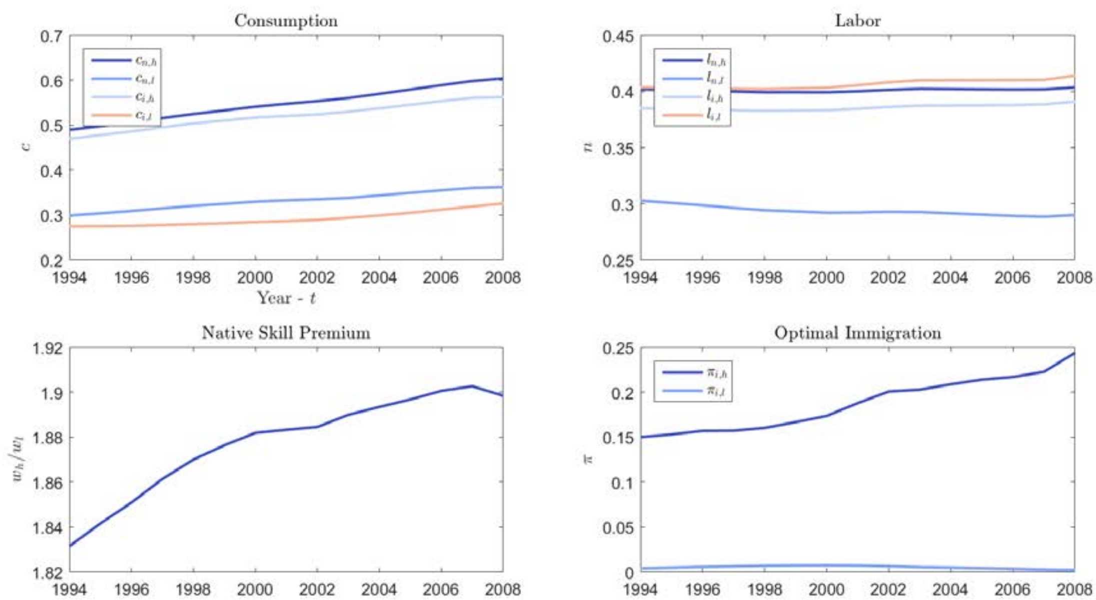

Figure 4: Mirrleesian policy with discrimination

In this case, we assume that the government can still discriminate between immigrants of different skill types. In general, the government may need to use this flexibility to implement the optimal allocations. Interestingly, the optimal policy in our calibrated model does not need to discriminate between low- and high-skill immigrant workers. This is because the optimum is such that neither low-skill immigrants nor high-skill immigrants want to choose the allocations assigned to the other.[27] A single non-linear income tax/transfer function on immigrants can implement the optimal allocation.

Mirrleesian policy without discrimination: immigration to welfare Finally, we consider the case in which immigrants cannot be discriminated from natives. Entering migrants have access to the full benefits and obligations of the welfare state.

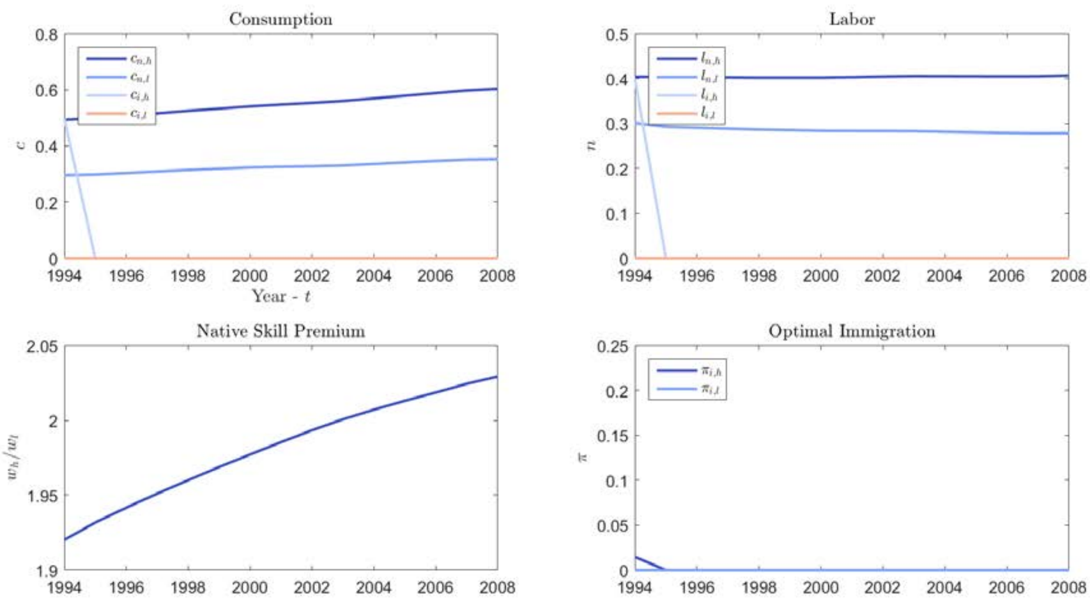

Figure 5: Mirrleesian policy without discrimination

With no discrimination, there is no immigration of low-skill workers under the optimal policy (see Figure 5). Under the baseline calibration, this solution also features no immigration of high-skill workers except in the initial period. The levying of heavy taxes

[27]In this case, the binding constraints for immigrants are their participation constraints. The incentive compatibility constraints do not bind at the optimum because of the presence of binding type-dependent outside options.

on high-skill workers implies that their utility is below the outside option of high-skill immigrants. As a result, high-skill immigrants choose not to immigrate after period 1.

The absence of high-skill immigration is not a robust feature of our model, but depends on the outside option of high-skill immigrants. Figure 6 shows the optimal policy when the outside options for each agent are 5 percent lower in equivalent consumption. The optimal policy responds to this change by letting high-skill immigration jump from close to zero to about 0.2 in 1994 and then rise above 0.4 in 2008. This effect results in a sizable reduction in the tax burden of high-skill workers, native and immigrant alike. The optimal policy also responds to rises in the skill premium by allowing more high-skill immigrants into the country.

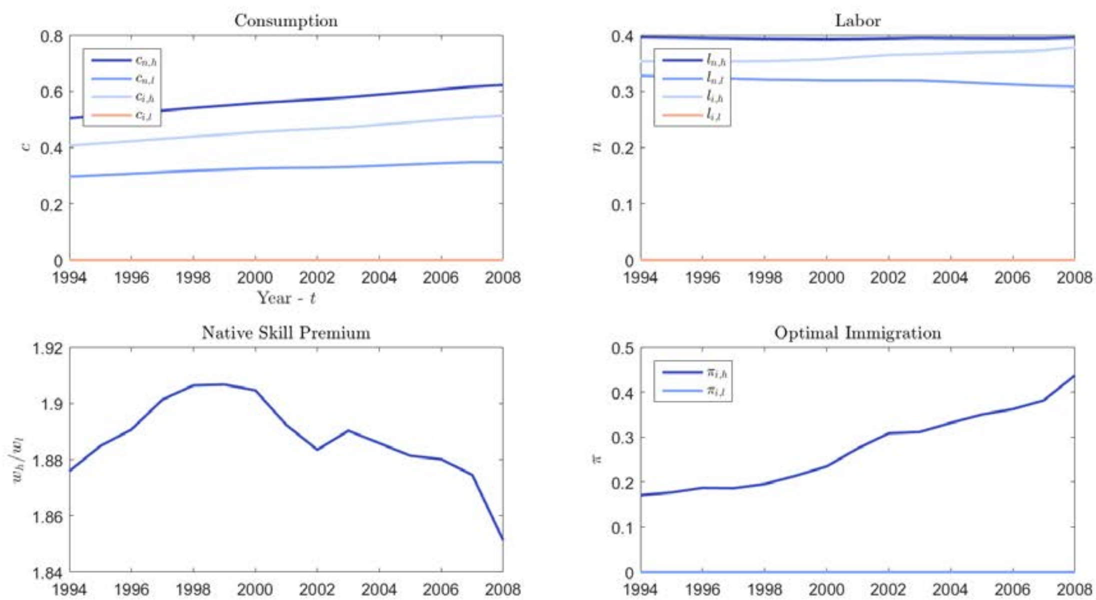

Figure 6: Mirrleesian policy without discrimination: low outside option

In these solutions, high-skill workers are always free to immigrate but they may choose not to. The results discussed above show the existence of two possible regimes: a *domestic redistribution regime* and an *immigration surplus regime*. With high outside options for high-skill immigrants, the domestic redistribution regime is optimal. In this regime, high-skill workers face high taxes, which finance transfers to low-skill workers. As a result of the heavy tax burden imposed on high-skill workers, high-skill immigrants choose not to immigrate.

If high-skill immigrants have lower outside options, the *immigration surplus regime* is optimal. In this regime, high-skill workers pay low taxes. This policy is designed to attract high-skill immigrants and raise land income, which is used to finance transfers to low-skill workers.

Comparing policies Figure 7 shows the utility levels of low- and high-skill natives under the different policies. Overall utility improves over time because of the presence of skill-biased technical change.

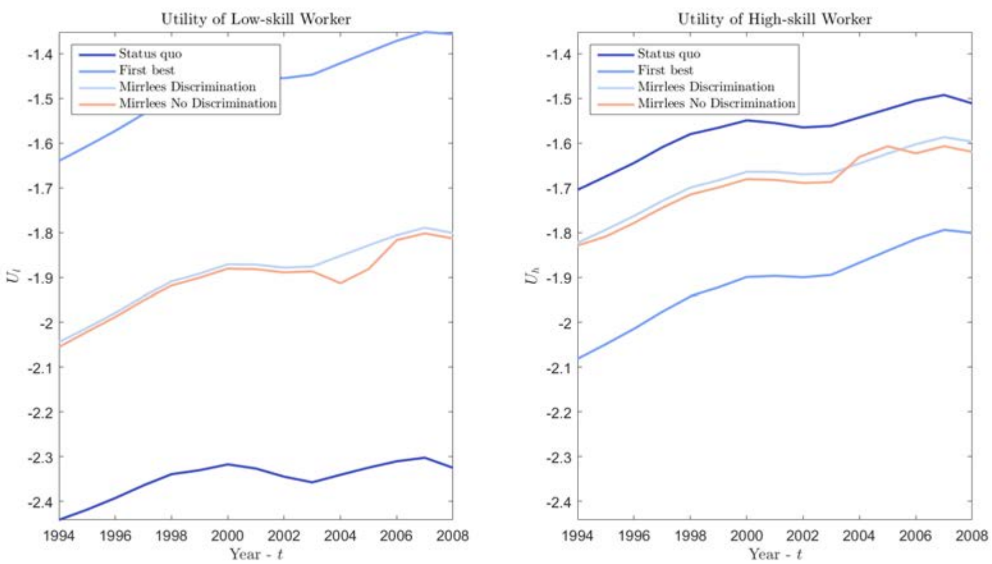

Figure 7: Comparing policies

The left panel shows the utility of low-skill natives. This utility is highest in the first best and lowest in the status quo. The utility ordering reflects the difference in redistribution policies. Low-skill natives prefer the policy with discrimination to that without discrimination since the absence of discrimination limits the scope of redistribution.

The right panel shows the utility of high-skill natives. High-skill natives prefer the status quo to any other policy. This preference reflects the status quo's low level of income redistribution. High-skill natives prefer the policy with discrimination to that with no discrimination since the former allows for the immigration of low-skill workers.

5.2 Robustness

In this subsection, we consider four alternative specifications for the elasticity of substitution in production between immigrants and natives. Table 2 presents results for these specifications for the years 1994 and 2008 for the first-best policy and the Mirrleesian policies with and without discrimination.

Model 1 presents our baseline calibration. In model 2, we consider the case in which the elasticity of substitution between natives and immigrants is lower than that in the benchmark model, $\varepsilon = 10$.[28] We recalibrate φ so that the average immigrant-wage gap is the same as in the baseline calibration, 0.89. Model 3 has a lower degree of substitution between low- and high-skill workers, $\rho = 1.69$; the elasticity of substitution between immigrants and natives is the same as in the baseline calibration. We recalibrate φ to target and an average immigrant wage gap of 0.89. Model 4 is the one we consider in the theory sections where immigrant and native workers are perfect substitutes. We set $\varepsilon = \infty$ and $\varphi = 1$.

We find the robust prediction that, relative to the first best, high-skill immigration in the Mirrleesian optimal plan with discrimination is higher and low-skill immigra-

[28]Estimates of the degree of substitution between natives and immigrants vary widely. Card (2009) estimates an elasticity of substitution of around 20. We use this value in our baseline calibration. However, different studies estimate lower degrees of substitution. Cortes (2008) and Burstein, Hanson, Tian, and Vogel (2019) estimate this elasticity to be closer to 4. Ottaviano and Peri (2012) estimate an elasticity of substitution between 5 and 10.

tion is lower. These results reinforce the importance of general-equilibrium effects for the design of optimal tax policies. Across these specifications, both types of native workers have higher utility in the case with discrimination than in the case without discrimination.

The levels of immigration are particularly sensitive to the elasticity of substitution in production between immigrants and natives. In the case with perfect substitution between natives and immigrants, the numbers for total immigration increase to two immigrants per native worker.[29]

In model 5, we consider the case with less congestion in the provision of public goods. We set κ equal to 0.63. This value was obtained as the ratio of government outlays associated with immigrants and natives in National Academies of Sciences, Engineering, and Medicine (2017).[30]

[29] Benhabib and Jovanovic (2012) use a worldwide welfare function and argue that immigration flows should be much higher than those observed in the data.

[30] We obtain these numbers from Table 8-2, scenario 5, using the third generation of immigrants as a proxy for natives.

Table 2: Robustness checks

			1994						2008					
			First best		Mirrleesian policy				First best		Mirrleesian policy			
					w/ disc.		w/o disc.				w/ disc.		w/o disc.	
			$\pi_{i,a}$	$\Delta U_{n,a}$	$\pi_{i,a}$	$\Delta U_{n,a}$	$\pi_{i,a}$	$\Delta U_{n,a}$	$\pi_{i,a}$	$\Delta U_{n,a}$	$\pi_{i,a}$	$\Delta U_{n,a}$	$\pi_{i,a}$	$\Delta U_{n,a}$
Model 1 *Baseline*	$\rho = 3$ $\varepsilon = 20$	High skill	0.13	−0.31	0.15	−0.11	0.01	−0.12	0.22	−0.25	0.24	−0.08	0	−0.10
	$\varphi = 0.85$	Low skill	0.01	1.23	0	0.49	0	0.47	0	1.63	0	0.69	0	0.67
Model 2 *Low subst. n/i*	$\rho = 3$ $\varepsilon = 10$	High skill	0.11	−0.32	0.11	−0.12	0	−0.13	0.17	−0.26	0.18	−0.09	0	−0.12
	$\varphi = 0.84$	Low skill	0.02	1.22	0.01	0.48	0	0.45	0.01	1.62	0.01	0.67	0	0.63
Model 3 *Low subst. h/l*	$\rho = 1.69$ $\varepsilon = 20$	High skill	0.08	−0.28	0.10	−0.10	0	−0.11	0.13	−0.23	0.15	−0.08	0	−0.09
	$\varphi = 0.85$	Low skill	0.07	0.95	0.03	0.38	0	0.36	0.05	1.26	0.02	0.53	0	0.51
Model 4 *Perfect subst. n/i*	$\rho = 3$ $\varepsilon = \infty$	High skill	1.40	−0.20	1.46	−0.01	0.53	−0.05	1.49	−0.16	1.55	0.00	0.88	−0.03
	$\varphi = 1$	Low skill	0.93	0.93	0.88	0.40	0	0.34	0.52	1.19	0.48	0.52	0	0.47
Model 5		High skill	0.35	−0.31	0.38	−0.10	0.01	−0.12	0.71	−0.23	0.39	−0.06	0	−0.11
Low congestion	$\kappa = 0.63$	Low skill	0.06	1.23	0.04	0.49	0	0.46	0.06	1.64	0.04	0.71	0	0.65

Notes: This table shows numerical results for levels of immigration $\pi_{i,a}$ and changes in utility of the native population $\Delta U_{n,a}$ for each skill type, $a = l, h$. These numbers are computed for the initial year, 1994, and the final year, 2008. For each of these years, we compute the first-best allocation, the Mirrleesian optimal plan with discrimination, and finally the Mirrleesian optimal plan without discrimination. The change in utility is computed as a consumption equivalent measure, which corresponds to the change in consumption in the status-quo equilibrium that would be equivalent to the plan in consideration. The robustness checks involve the elasticity parameters, ρ, ε, and κ. All other parameters are recalibrated to match the same targets as the baseline calibration, except in the case of model 4 in which we let immigrants and native workers be perfect substitutes and fix $\varphi = 1$.

6 Conclusions

We study the immigration policy that maximizes the welfare of the native population in a model where the government designs an optimal redistributive welfare system and supplies public goods.

We show that when the government can design an income tax system that discriminates between native and immigrant workers, it is always optimal to have no quotas and allow free immigration. Abstracting from general-equilibrium effects on the skill premium, the optimal policy is for immigrants to pay no taxes other than a levy that internalizes the congestion they create in the provision of public goods.

Since immigration affects the skill premium, it is optimal to use the tax system to encourage the immigration of high-skill workers and discourage the immigration of low-skill workers. This policy reduces the skill premium, allowing the planner to redistribute more income toward low-skill workers.

When immigrant and native workers must be treated alike in their access to the welfare system, the optimal immigration policy bans the immigration of low-skill workers. High-skill workers are still free to immigrate, but they may choose not to. They may be discouraged by heavy taxes levied on high-skill natives and immigrants. This scenario resembles the experience of the Scandinavian countries. Despite having liberal immigration policies for high-skill workers, the heavy taxes levied on both native and foreign high-skill workers result in very little high-skill immigration to these countries.

In our analysis, we make two simplifying assumptions. We abstract from the potential human capital externalities of immigration discussed by Borjas (2014) and assume that the social planner maximizes the welfare of the native population. Investigating how these assumptions shape the properties of the optimal immigration policy is an interesting avenue for future research.

7 References

Acemoglu, D., Autor, D.H., 2011. Skills, tasks and technologies: implications for employment and earnings. in Handbook Labor Econ. 4. 1043-1171. Elsevier.

Alesina, A., Miano, A., Stantcheva, S., 2018. Immigration and redistribution. NBER Working Paper 24733.

Arnott, R.J., Stiglitz, J.E., 1979. Aggregate land rents, expenditure on public goods, and optimal city size. Q. J. Econ. 93 (4), 471-500.

Atkinson, A.B., Stiglitz, J.E., 1976. The design of tax structure: direct versus indirect taxation. J. Public Econ. 6 (1-2), 55-75.

Autor, D.H., Katz, L.F., Krueger, A.B., 1998. Computing inequality: have computers changed the labor market? Q. J. Econ. 133 (4), 1169-1213.

Becker, G.S., Posner, R.A., 2009. Uncommon sense: economic insights, from marriage to terrorism. University of Chicago Press.

Benabou, R., 2000. Unequal societies: income distribution and the social contract. Am. Econ. Rev. 90 (1), 96-129.

Benhabib, J., Jovanovic, B., 2012. Optimal migration: A world perspective. Intern. Econ. Rev. 53 (2), 321-348.

Borjas, G.J., Freeman, B.R., Katz, L.F., 1992. On the labor market effects of immigration and trade. in Immigration and the Work Force: Economic Consequences for the United States and Source Areas, by Borjas, G.J., and Freeman, R.B., Chicago: University of Chicago Press, 213–44.

Borjas, G.J., 1995. The economic benefits from immigration. J. of Econ. Persp. 9, 3-22.

Borjas, G.J., 1995. *Immigration Economics*, Harvard University Press, June, 2014).

Burstein, A., Hanson, G., Tian, L., Vogel, J., 2019. Tradability and the labor market impact of immigration: theory and evidence for the U.S. Working paper no. 23330, National Bureau of Economic Research.

Card, D., 2005. Is the new immigration really so bad? The Econ. Journal 115, 300-323.

Card, D., 2009. Immigration and inequality. Am. Econ. Rev. P&P 99 (2), 1-21.

Chetty, R., 2006. A new method of estimating risk aversion. Am. Econ. Rev. 96 (5), 1821–1834.

Chetty, R., Guren, A., Manoli, D., Weber, A., 2011. Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins. Am. Econ. Rev. P&P 101 (3), 471-475.

Cortes, P., 2008. The effect of low-skilled immigration on U.S. prices: evidence from CPI data. J. Polit. Economy116 (3), 381-422.

De la Torre, A., Yeyati, E.L., Pienknagura, S., 2013. Latin America and the Caribbean as tailwinds recede : in search of higher growth, LAC semiannual report. World Bank, Washington, DC, April.

Diamond, P.A., Mirrlees, J.A., 1971. Optimal taxation and public production I: production efficiency. Am. Econ. Rev. 61 (1), 8-27.

Ferriere, A., Navarro, G., 2019. The heterogeneous effects of government spending: it's all about taxes. Working paper.

Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, J.R., 2018. Integrated public use microdata series, current population survey: version 6.0 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D030.V6.0

Friedman, M., 1978. What is America? in The Economics of Freedom. Cleveland: Standard Oil Company of Ohio.

Golosov, M., Kocherlakota, N., Tsyvinski, A., 2003. Optimal indirect and capital taxation. Rev. Econ. Stud. 70 (3), 569–587.

Grossman, J.B., 1982. The substitutability of natives and immigrants in production. Rev. Econ. and Stat. 64 (4), 596–603.

Heathcote, J., Storesletten, K., Violante, G.L., 2017. Optimal tax progressivity: An analytical framework. Q. J. Econ. 132 (4), 1693-1754.

Herrendorf, B., Valentinyi, A., 2008. Measuring factor income shares at the sectoral level. Rev. Econ. Dyn. 11 (4), 820-835.

Jaimovich, N., Rebelo, S., 2017. Nonlinear effects of taxation on growth. J. Polit. Economy 125 (1), 265-291.

Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963-1987: supply and demand factors. Q. J. Econ., 107 (1) 35-78.

Kocherlakota, N., 2010. The new dynamic public finance. Princeton, NJ: Princeton University Press.

Krusell, P., Ohanian, L.E., Rios-Rull, J., Violante, G.L., 2000. Capital-skill complementarity and inequality: a macroeconomic analysis, Econometrica 68 (5), 1029-1053.

Lindbeck, A., Weibull, J.W., 1987. Balanced-budget redistribution as the outcome of political competition. Pub. Choice 52 (3), 273-297.

Manacorda, M., Manning, A., Wadsworth, J., 2012. The impact of immigration on the structure of wages: theory and evidence from Britain. J. of the European Econ. A. 10, 120-151.

Mirrlees, J.A., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38 (2), 175-208.

Mirrlees, J.A., 1982. Migration and optimal income taxes. J. Pub. Econ. 18 (3), 319-341.

Naito, H., 1999. Re-examination of uniform commodity taxes under a non-linear income tax system and its implications for production efficiency. J. Pub. Econ. 71, 165-188.

National Academies of Sciences, Engineering, and Medicine. 2017. The economic and fiscal consequences of immigration. Washington, DC: The National Academies Press.

Ottaviano, G.I.P., Peri, G., 2012. Rethinking the effect of immigration on wages. J. of the European Econ. A. 10 (1), 152-97.

Scheuer, F., Werning, I., 2018. Mirrlees meets Diamond-Mirrlees. Working paper

No. w22076, National Bureau of Economic Research.

Stiglitz, J., 1982. Self-selection and Pareto efficient taxation. J. Pub. Econ. 17 (2), 213-240.

Topel, R.H., 1994. Regional trends in wage inequality. Am. Econ. Rev. 84, 17-22.

Werning, I., 2007. Optimal fiscal policy with redistribution. Q. J. Econ. 122 (3), 925-967.

A Appendix

A.1 Model equilibrium

An equilibrium, as in definition 1, is composed of allocations $c_{b,a}, n_{b,a}$, and $l_{b,a}$ for all b, a, prices w_l, w_h, and r, and policies $\pi_{i,l}, \pi_{i,h}, G$, and $\mathcal{T}_{b,a}$ for all b, a that satisfy the following conditions:

$$(c_{b,a}, n_{b,a}, l_{b,a}) \in \arg \max_{(c,n,l) \in \mathcal{B}_{b,a}} u(c, n), \tag{1}$$

where $\mathcal{B}_{b,a} \equiv \left\{(c, n, l) \in \mathbb{R}^3_+ : l \leq L_{b,a} \;\&\; c \leq w_a n + rl - \mathcal{T}_{b,a}(w_a n, rl)\right\}$,

$$w_a = F_a(L, N_l, N_h), \tag{2}$$

$$r = F_L(L, N_l, N_h), \tag{3}$$

$$u\left(c_{i,a}, n_{i,a}\right) + v\left(G\right) \geq \overline{U}_a, \text{ if } \pi_{i,a} > 0 \tag{4}$$

$$\sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma\left(\pi_{i,l} + \pi_{i,h}\right) G \leq F\left(L, N_l, N_h\right) \tag{5}$$

and the factor market clearing conditions

$$L = \sum_b \sum_a \pi_{b,a} l_{b,a}, \tag{6}$$

$$N_a = \sum_b \pi_{b,a} n_{b,a}. \tag{7}$$

To simplify, we often write F_a and F_L instead of $F_a(L, N_l, N_h)$ and $F_L(L, N_l, N_h)$.

A.2 Policy with unrestricted taxes: implementability constraints

We look for a set of equilibrium conditions that are necessary and sufficient for the implementability of the allocations

$$\mathcal{A} \equiv \{\{c_{b,a}, n_{b,a}, l_{b,a}\}_{b,a}, \pi_{i,l}, \pi_{i,h}, G\},$$

where $c_{b,a}, n_{b,a}, \pi_{i,a}, G \in \mathbb{R}_+$, and $l_{b,a} \in [0, L_{b,a}]$.

Lemma 3. *Suppose that the government has access to unrestricted taxation. Then, the allocations $\mathcal{A}$ can be implemented as an equilibrium if and only if the resource constraint, (5), and the participation constraints, (4), are satisfied.*

The necessity of (5) and (4) follows trivially from the fact that these are equilibrium conditions.

For sufficiency, note that we can construct aggregate labor and land, N_a and L, from their definitions, and prices

$$w_a = F_a(L, N_l, N_h) \text{ and } r = F_L(L, N_l, N_h).$$

This result means that (2)-(7) are satisfied.

Now we only need to find tax/transfer functions $\mathcal{T}_{b,a}$ such that the choices $c_{b,a}, n_{b,a}$, and $l_{b,a}$ are optimal. For each $b = n, i$ and $a = l, h$, this problem can be solved by setting

$$\mathcal{T}_{b,a}(y_N, y_L) \equiv y_N + y_L - \max\left\{c : u\left(c, \frac{y_N}{w_a}\right) \leq u(c_{b,a}, n_{b,a})\right\}$$

for all $y_N, y_L \in \mathbb{R}_+$. Since this tax/transfer function implies that $(c_{b,a}, n_{b,a}, l_{b,a}) \in \mathcal{B}_{b,a}$, and for all $(c, n, l) \in \mathcal{B}_{b,a}$ it implies that $u(c, n) \leq u(c_{b,a}, n_{b,a})$.

A.3 Proof of proposition 1

We write the value function for given $\pi_{i,l}, \pi_{i,h}$ as

$$\begin{aligned}
\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \sum_a \omega_a \pi_{n,a}[u(c_{n,a}, n_{n,a}) + v(G)] \quad \text{s.to.}& \\
[\eta_a] \quad u(c_{i,a}, n_{i,a}) + v(G) \geq \overline{U}_{i,a}& \\
[\lambda] \quad F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i) G,&
\end{aligned}$$

where η_l,η_h, and λ denote the Lagrange multipliers of each constraint.

Clearly the participation constraint of immigrants must bind, or else it would be possible to increase native welfare by decreasing immigrant consumption and increasing that of natives.

The first-order conditions with respect to $c_{i,a}$ and $n_{i,a}$ are given by

$$\begin{aligned}
\eta_a u_c(c_{i,a}, n_{i,a}) &= \lambda \pi_{i,a} \\
-\eta_a u_n(c_{i,a}, n_{i,a}) &= \lambda F_a \pi_{i,a},
\end{aligned}$$

which together imply that

$$\frac{u_n(c_{i,a}, n_{i,a})}{u_c(c_{i,a}, n_{i,a})} = F_a.$$

Finally, the envelope condition with respect to $\pi_{i,a}$ is

$$\frac{d\mathcal{W}}{d\pi_{i,a}} = \lambda[F_a n_{i,a} - c_{i,a} - \sigma'(\pi_i)G],$$

and note that $\lambda > 0$. Optimality with interior immigration requires that

$$\frac{d\mathcal{W}}{d\pi_{i,a}} = 0 \Rightarrow c_{i,a} = F_a n_{i,a} - \sigma'(\pi_i)G.$$

A possible implementation of this allocation sets

$$\mathcal{T}_{i,a}(y_N, y_L) = \sigma'(\pi_i)G,$$

for all y_N, y_L.

When there are no congestion effects, then $\sigma'(\pi_i) = 0$ and the optimal plan is such that

$$\mathcal{T}_{i,a}(y_N, y_L) = 0.$$

A.4 Mirrleesian policy with discrimination: implementability constraints

Suppose that

$$\mathcal{T}_{n,l}(y_N, y_L) = \mathcal{T}_{n,h}(y_N, y_L) \equiv \mathcal{T}_n(y_N, y_L)$$

for all $y_N, y_L \in \mathbb{R}_+$ but immigrants can still be discriminated from natives:

$$\mathcal{T}_{i,l}(y_N, y_L) \neq \mathcal{T}_{i,h}(y_N, y_L) \neq \mathcal{T}_n(y_N, y_L).$$

Lemma 4. *Suppose that the government cannot discriminate natives from immigrants based on their skill type but can perfectly discriminate between immigrants and natives. Then, the allocations $\mathcal{A}$ can be implemented as an equilibrium if and only if the resource constraint, (5), the participation constraints, (4), and the incentive constraints,*

$$u(c_{n,h}, n_{n,h}) \geq u\left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right), \tag{8}$$

$$u(c_{n,l}, n_{n,l}) \geq u\left(c_{n,h}, \frac{F_h n_{n,h}}{F_h}\right), \text{ if } l_{n,h} \leq L_{n,l}, \tag{9}$$

are satisfied.

The necessity of (5) and (4) is again trivial. To show the necessity of (8) and (9), note first that

$$(c_{n,a}, n_{n,a}, l_{n,a}) \in \arg\max_{(c,n,l)\in\mathcal{B}_{n,a}} u(c,n),$$

which implies that $(c_{n,a}, n_{n,a}, l_{n,a}) \in \mathcal{B}_{n,a}$. Furthermore, we can now see that because $L_{n,h} \geq L_{n,l}$, then $(c_{n,l}, w_l n_{n,l}/w_h, l_{n,l}) \in \mathcal{B}_{b,h}$ because

$$l_{n,l} \leq L_{n,l} \Rightarrow l_{n,l} \leq L_{n,h}$$

$$c_{n,l} \leq w_l n_{n,l} + r l_{n,l} - \mathcal{T}_n\left(w_l n_{n,l}, r l_{n,l}\right) \Rightarrow c_{n,l} \leq w_h \frac{w_l n_{n,l}}{w_h} + r l_{n,l} - \mathcal{T}_n\left(w_l \frac{w_h n_{n,l}}{w_h}, r l_{n,l}\right).$$

This result implies that

$$u(c_{n,h}, n_{n,h}) \geq u\left(c_{n,l}, \frac{w_l n_{n,l}}{w_h}\right),$$

or, replacing $w_a = F_a$, we obtain (8). Instead, $(c_{n,h}, w_h n_{n,h}/w_l, l_{n,h}) \in \mathcal{B}_{b,l}$ only if $l_{n,h} \leq L_{n,l}$; that is, only the high-skill worker's land use is low enough. As a result, we obtain the necessary condition:

$$u(c_{n,l}, n_{n,l}) \geq u\left(c_{n,h}, \frac{F_h n_{n,h}}{F_l}\right), \text{ if } l_{n,h} \leq L_{n,l}.$$

To show sufficiency, suppose that $\mathcal{A}$ satisfies (5), (4), (8), and (9). We can construct prices, aggregate labor endowment for each skill, and aggregate land use using equations (2), (3), (6), and (7).

As before, define

$$\mathcal{T}_{i,a}(y_N, y_L) \equiv y_N + y_L - \max\left\{c : u\left(c, \frac{y_N}{w_a}\right) \leq u(c_{b,a}, n_{b,a})\right\}$$

for $a = l, h$. This choice of $\mathcal{T}_{i,a}(y_N, y_L)$ guarantees that

$$(c_{i,a}, n_{i,a}, 0) \in \arg\max_{(c,n,l)\in\mathcal{B}_{i,a}} u(c,n).$$

If $l_{n,h} \leq L_{n,l}$, we set

$$\mathcal{T}_n(y_N, y_L) = y_N + y_L - \max\left\{c : u\left(c, \frac{y_N}{w_a}\right) \leq u(c_{b,a}, n_{b,a}),\ \forall a\right\},$$

and if $l_{n,h} > L_{n,l}$, then

$$\mathcal{T}_n(y_N, y_L) = \begin{cases} y_N + y_L - c_{n,l}, & \text{if } y_N = w_l n_{n,l} \text{ and } y_L = r l_{n,l} \\ y_N + y_L - c_{n,h}, & \text{if } y_N = w_h n_{n,h} \text{ and } y_L = r l_{n,h} \\ y_N + y_L, & \text{otherwise.} \end{cases}$$

A.5 Proof of lemma 1

We write the value function for given $\pi_{i,l}, \pi_{i,h}$ as

$$\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \sum_a \omega_a \pi_{n,a}[u(c_{n,a}, n_{n,a}) + v(G)] \quad \text{s.to.}$$

$$[\eta_{i,a}] \quad u(c_{i,a}, n_{i,a}) + v(G) \geq \overline{U}_{i,a}$$

$$[\lambda] \quad F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i) G,$$

$$[\eta_{n,h}] \quad u(c_{n,h}, n_{n,h}) \geq u\left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right)$$

$$[\eta_{n,l}] \quad u(c_{n,l}, n_{n,l}) \geq u\left(c_{n,h}, \frac{F_h n_{n,h}}{F_l}\right), \text{ if } l_{n,h} \leq L_{n,l},$$

where the variables in square brackets denote the Lagrange multipliers of each constraint.

To work toward a contradiction, suppose that the optimum $\mathcal{A}$ is such that $l_{n,a} < L_{n,a}$ for some a. Consider the perturbation $\mathcal{A}'$, which is such that it keeps the following allocations constant: $c'_{b,a} = c_{i,a}$ for $a = l, h$, $n'_{b,a} = n_{b,a}$ for $b = n, i$ and $a = l, h$; but in which all land is used, $l'_{n,a} = L_{n,a}$, and government spending is increased,

$$G' = G + F(L', N_l, N_h) - F(L, N_l, N_h).$$

First, note that $G' > G$ because $L' > L \Rightarrow F(L', N_l, N_h) > F(L, N_l, N_h)$. To see that this allocation is still feasible, note that: (1) because $v'(G) > 0$, the participation constraint of immigrants is still satisfied; (2) because G is strictly separable in the utility function, and $d(F_l/F_h)/dL = 0$, the incentive compatibility constraint of high-skill natives is still satisfied; (3) the low-skill native incentive compatibility either is still satisfied for the same reason or, if $L_{n,h} > L_{n,l}$, does not require being satisfied anymore; and, finally, (4) the resource constraint is still satisfied.

Finally, note that $\mathcal{A}'$ yields strictly higher welfare than $\mathcal{A}$, because $v'(G) > 0$.

A.6 Proof of proposition 2

We write the value function for given $\pi_{i,l}, \pi_{i,h}$ as

$$\begin{aligned}
\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \sum_a \omega_a \pi_{n,a}[u(c_{n,a}, n_{n,a}) + v(G)] \quad \text{s.to.} \\
[\eta_{i,a}] \quad & u(c_{i,a}, n_{i,a}) + v(G) \geq \overline{U}_{i,a} \\
[\lambda] \quad & F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i) G, \\
[\eta_{n,h}] \quad & u(c_{n,h}, n_{n,h}) \geq u\left(c_{n,l}, \frac{\theta_l n_{n,l}}{\theta_h}\right) \\
[\eta_{n,l}] \quad & u(c_{n,l}, n_{n,l}) \geq u\left(c_{n,h}, \frac{\theta_h n_{n,h}}{\theta_l}\right), \text{if } L_{n,h} = L_{n,l},
\end{aligned}$$

where the variables in square brackets denote the Lagrange multipliers of each constraint.

Clearly the participation constraint of immigrants must bind, or else it would be possible to increase native welfare by decreasing immigrant consumption and increasing that of natives.

The first-order conditions with respect to $c_{i,a}$ and $n_{i,a}$ are given by

$$\begin{aligned}
\eta_a u_c(c_{i,a}, n_{i,a}) &= \lambda \pi_{i,a} \\
-\eta_a u_n(c_{i,a}, n_{i,a}) &= \lambda F_a \pi_{i,a},
\end{aligned}$$

which together imply that

$$\frac{u_n(c_{i,a}, n_{i,a})}{u_c(c_{i,a}, n_{i,a})} = F_a.$$

Finally, the envelope condition with respect to $\pi_{i,a}$ is

$$\frac{d\mathcal{W}}{d\pi_{i,a}} = \lambda[F_a n_{i,a} - c_{i,a} - \sigma'(\pi_i)G]$$

with $\lambda > 0$. Optimality with interior immigration requires that

$$\frac{d\mathcal{W}}{d\pi_{i,a}} = 0 \Rightarrow c_{i,a} = F_a n_{i,a} - \sigma'(\pi_i)G.$$

A possible implementation of this allocation sets

$$\mathcal{T}_{i,a}(y_N, y_L) = \sigma'(\pi_i)G$$

for all y_N, y_L.

When there are no congestion effects ($\sigma(\pi_i) = 1$), then $\sigma'(\pi_i) = 0$ and the optimal plan is such that

$$\mathcal{T}_{i,a}(y_N, y_L) = 0.$$

A.7 Proof of proposition 3

We write the value function for given $\pi_{i,l}, \pi_{i,h}$ as

$$\begin{aligned}
\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \sum_a \omega_a \pi_{n,a}[u(c_{n,a}, n_{n,a}) + v(G)] \quad \text{s.to.}& \\
[\eta_{i,a}] \quad u(c_{i,a}, n_{i,a}) + v(G) \geq \overline{U}_{i,a}& \\
[\lambda] \quad F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i)G,& \\
[\eta_{n,h}] \quad u(c_{n,h}, n_{n,h}) \geq u\left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right)& \\
[\eta_{n,l}] \quad u(c_{n,l}, n_{n,l}) \geq u\left(c_{n,h}, \frac{F_h n_{n,h}}{F_l}\right), \text{if } L_{n,h} = L_{n,l},&
\end{aligned}$$

where the variables in square brackets denote the Lagrange multipliers of each constraint.

Clearly the participation constraint of immigrants must bind, or else it would be possible to increase native welfare by decreasing immigrant consumption and increasing native consumption.

The envelope conditions are given by

$$\frac{d\mathcal{W}}{d\pi_{i,a}} =\lambda \left[F_a n_{i,a} - c_{i,a} - \sigma'(\pi_i)G\right] - \eta_{n,h} u_n \left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right) n_{n,l} \frac{d(F_l/F_h)}{dN_a} n_{n,a}$$
$$- \eta_{n,l} u_n \left(c_{n,h}, \frac{F_h n_{n,h}}{F_l}\right) n_{n,l} \frac{d(F_h/F_l)}{dN_a} n_{n,a}.$$

If the incentive constraint of high-skill workers binds and that of low-skill workers does not, then $\eta_{n,h} > 0$ and $\eta_{n,l} = 0$.

As a result, the total tax paid by an immigrant worker of skill a is

$$\mathcal{T}_{i,a}(w_a n_{i,a}, 0) = \sigma'(\pi_i)G - \eta_{n,h} \left[-u_n \left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right) n_{n,l} n_{n,a}\right] \frac{d(F_l/F_h)}{dN_a}.$$

By the skill-premium monotonicity assumption, we have that $\frac{d(F_l/F_h)}{dN_l} < 0$ and that $\frac{d(F_l/F_h)}{dN_l} > 0$, which implies that

$$\mathcal{T}_{i,h}(w_h n_{i,h}, 0) < \sigma'(\pi_i)G$$
$$\mathcal{T}_{i,l}(w_l n_{i,l}, 0) > \sigma'(\pi_i)G.$$

When there are no congestion effects ($\sigma(\pi_i) = 1$), then $\sigma'(\pi_i) = 0$ and the optimal plan is such that

$$\mathcal{T}_{i,h}(w_h n_{i,h}, 0) < 0$$
$$\mathcal{T}_{i,l}(w_l n_{i,l}, 0) > 0.$$

A.8 Mirrleesian policy without discrimination: implementability constraints

Suppose that the government cannot discriminate between immigrants and natives:

$$\mathcal{T}_{n,l}(y_N, y_L) = \mathcal{T}_{n,h}(y_N, y_L) = \mathcal{T}_{i,l}(y_N, y_L) = \mathcal{T}_{i,h}(y_N, y_L) \equiv \mathcal{T}(y_N, y_L)$$

for all $y_N, y_L \in \mathbb{R}_+$.

Lemma 5. *Suppose that the government cannot discriminate between households based on skill or immigration status. Then the allocations $\mathcal{A}$ can be implemented as an equilibrium if and only if the participation constraints, (4), the resource constraint, (5), and the following incentive constraints,*

$$u(c_{b,a}, n_{b,a}) = \max_{(b',a')\in\Theta_{b,a}} u\left(c_{b',a'}, \frac{w_{a'} n_{b',a'}}{w_a}\right), \tag{10}$$

are satisfied, where $\Theta_{b,a} \equiv \{(b', a') : \pi_{b',a'} > 0 \;\&\; l_{b',a'} \leq L_{b,a}\}$.

The necessity of (5) and (4) is again trivial. To show necessity of (10), note that $(c_{b',a'}, w_{a'} n_{b',a'}/w_a, l_{b',a'}) \in \mathcal{B}_{b,a}$ only if $l_{b',a'} \in L_{b,a}$.

To show sufficiency, suppose that $\mathcal{A}$ satisfies (5), (4), and (10). We can construct prices, aggregate labor endowment for each skill, and aggregate land use using equations (2), (3), (6), and (7).

Furthermore, we construct the following tax system. If $\pi_{i,l}, \pi_{i,h} > 0$, then

$$\mathcal{T}(y_N, y_L) = \begin{cases} y_N + y_L - c_{n,l}, & \text{if } y_N = w_l n_{n,l} \text{ and } y_L = r l_{n,l} \\ y_N + y_L - c_{n,h}, & \text{if } y_N = w_h n_{n,h} \text{ and } y_L = r l_{n,h} \\ y_N + y_L - c_{i,l}, & \text{if } y_N = w_l n_{i,l} \text{ and } y_L = r l_{i,l} \\ y_N + y_L - c_{i,h}, & \text{if } y_N = w_h n_{i,h} \text{ and } y_L = r l_{i,h} \\ y_N + y_L, & \text{otherwise.} \end{cases}$$

If $\pi_{i,l} = 0$ and $\pi_{i,h} > 0$, then

$$\mathcal{T}(y_N, y_L) = \begin{cases} y_N + y_L - c_{n,l}, & \text{if } y_N = w_l n_{n,l} \text{ and } y_L = r l_{n,l} \\ y_N + y_L - c_{n,h}, & \text{if } y_N = w_h n_{n,h} \text{ and } y_L = r l_{n,h} \\ y_N + y_L - c_{i,h}, & \text{if } y_N = w_h n_{i,h} \text{ and } y_L = r l_{i,h} \\ y_N + y_L, & \text{otherwise.} \end{cases}$$

If $\pi_{i,l} > 0$ and $\pi_{i,h} = 0$, then

$$\mathcal{T}(y_N, y_L) = \begin{cases} y_N + y_L - c_{n,l}, & \text{if } y_N = w_l n_{n,l} \text{ and } y_L = r l_{n,l} \\ y_N + y_L - c_{n,h}, & \text{if } y_N = w_h n_{n,h} \text{ and } y_L = r l_{n,h} \\ y_N + y_L - c_{i,l}, & \text{if } y_N = w_h n_{i,l} \text{ and } y_L = r l_{i,l} \\ y_N + y_L, & \text{otherwise.} \end{cases}$$

Finally, if $\pi_{i,l} = \pi_{i,h} = 0$, then

$$\mathcal{T}(y_N, y_L) = \begin{cases} y_N + y_L - c_{n,l}, & \text{if } y_N = w_l n_{n,l} \text{ and } y_L = r l_{n,l} \\ y_N + y_L - c_{n,h}, & \text{if } y_N = w_h n_{n,h} \text{ and } y_L = r l_{n,h} \\ y_N + y_L, & \text{otherwise.} \end{cases}$$

A.9 Proof of lemma 2

Full land use We write the value function for given $\pi_{i,l}, \pi_{i,h}$ as

$$\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \sum_a \omega_a \pi_{n,a} [u(c_{n,a}, n_{n,a}) + v(G)] \quad \text{s.to.}$$

$$[\eta_{i,a}] \quad u(c_{i,a}, n_{i,a}) + v(G) \geq \overline{U}_{i,a}$$

$$[\lambda] \quad F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i) G,$$

$$[\chi_{b,a}] \quad u(c_{b,a}, n_{b,a}) \geq \max_{(a',b') \in \Theta_{b,a}} u\left(c_{a',b'}, \frac{F_{a'} n_{a',b'}}{F_a}\right).$$

To work toward a contradiction, suppose that the optimum $\mathcal{A}$ is such that $l_{n,a} < L_{n,a}$ for some a. Consider the perturbation $\mathcal{A}'$, which keeps the following allocations constant: $c'_{b,a} = c_{i,a}$ for $a = l, h$, $n'_{b,a} = n_{b,a}$ for $b = n, i$ and $a = l, h$; but in which all land is used $l'_{n,a} = L_{n,a}$ and government spending is increased:

$$G' = G + F(L', N_l, N_h) - F(L, N_l, N_h).$$

First note that $G' > G$ because $L' > L \Rightarrow F(L', N_l, N_h) > F(L, N_l, N_h)$.

To see that this allocation is still feasible, note that: (1) because $v'(G) > 0$, the participation constraint of immigrants is still satisfied; (2) because G is strictly separable in the utility function, and $d(F_l/F_h)/dL = 0$, the incentive compatibility constraints of low- and high-skill natives are still satisfied; (3) the resource constraint is still satisfied. Finally, note that $\mathcal{A}'$ yields strictly higher welfare than $\mathcal{A}$ because $v'(G) > 0$.

Low-skill allocation If $\pi_{i,l} = 0$, we can set $x_{i,l} = x_{n,l}$ for $x = c, n, l$ without loss of generality. Consider instead the case where $\pi_{i,l} > 0$. Define $U_{b,a} = u(c_{b,a}, n_{b,a}) + v(G)$

and note that the incentive constraints imply that $U_{n,l} \geq U_{i,l}$ and $U_{i,l} \geq U_{n,l}$. Then, $U_{n,l} = U_{i,l} \equiv U_l$.

The social planner's problem can be written as follows:

$$\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \sum_a \omega_a \pi_{n,a} U_{n,a} \quad \text{s.to.}$$

$$\begin{aligned}
[\eta_{i,a}] &\quad U_{i,a} \geq \overline{U}_{i,a} \\
[\lambda] &\quad F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i) G, \\
[\chi_{n,h}] &\quad U_{n,h} \geq U_{i,h} \\
[\chi^n_{n,h}] &\quad U_{i,h} \geq U_l + u\left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right) - u(c_{n,l}, n_{n,l}) \\
[\chi^i_{n,h}] &\quad U_{i,h} \geq U_l + u\left(c_{i,l}, \frac{F_l n_{i,l}}{F_h}\right) - u(c_{i,l}, n_{i,l}) \\
[\phi_{b,h}] &\quad u(c_{b,h}, n_{b,h}) + v(G) = U_{b,h} \\
[\phi_{b,l}] &\quad u(c_{b,l}, n_{b,l}) + v(G) = U_l.
\end{aligned}$$

Suppose, to work toward a contradiction, that $n_{b,l} < n_{b',l}$. This inequality implies that $c_{b,l} < c_{b',l}$, since both bundles must achieve the same utility. Furthermore, using the single-crossing condition,

$$u(c_{b,l}, n_{b,l}) = u(c_{b',l}, n_{b',l}) \Rightarrow u\left(c_{b,l}, \frac{F_l n_{b,l}}{F_h}\right) < u\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right).$$

As a result, the incentive constraint of high-skill workers mimicking b, l does not bind (i.e., $\chi^b_{n,h} = 0$ and $\chi^{b'}_{n,h} \geq 0$).

The first-order necessary conditions with respect to $c_{b,l}$, $n_{b,l}$, $c_{b',l}$, $n_{b',l}$ are

$$\lambda\pi_{b,l} = \phi_{b,l}u_c(c_{b,l}, n_{b,l})$$
$$\lambda F_l\pi_{b,l} = -\phi_{b,l}u_n(c_{b,l}, n_{b,l})$$
$$\lambda\pi_{b',l} = \left(\phi_{b',l} + \chi^{b'}_{n,h} - \chi^{b'}_{n,h}\frac{u_c\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_c\left(c_{b',l}, n_{b',l}\right)}\right) u_c(c_{b',l}, n_{b',l})$$
$$\lambda F_l\pi_{b',l} = -\left(\phi_{b',l} + \chi^{b'}_{n,h} - \chi^{b'}_{n,h}\frac{u_n\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_n\left(c_{b',l}, n_{b',l}\right)}\right) u_n(c_{b',l}, n_{b',l}).$$

This condition implies that

$$\frac{-u_n(c_{b,l}, n_{b,l})}{u_c(c_{b,l}, n_{b,l})} = F_l,$$
$$\frac{-u_n(c_{b',l}, n_{b',l})}{u_c(c_{b',l}, n_{b',l})} = \frac{\left(\phi_{b',l} + \chi^{b'}_{n,h} - \chi^{b'}_{n,h}\frac{u_c\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_c\left(c_{b',l}, n_{b',l}\right)}\right)}{\left(\phi_{b',l} + \chi^{b'}_{n,h} - \chi^{b'}_{n,h}\frac{u_n\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_n\left(c_{b',l}, n_{b',l}\right)}\right)} F_l.$$

Furthermore, the single-crossing condition also implies that

$$\frac{\left(\phi_{b',l} + \chi^{b'}_{n,h} - \chi^{b'}_{n,h}\frac{u_c\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_c\left(c_{b',l}, n_{b',l}\right)}\right)}{\left(\phi_{b',l} + \chi^{b'}_{n,h} - \chi^{b'}_{n,h}\frac{u_n\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_n\left(c_{b',l}, n_{b',l}\right)}\right)} \leq 1 \Leftrightarrow \frac{-u_n\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{-u_n\left(c_{b',l}, n_{b',l}\right)} \leq \frac{u_c\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_c\left(c_{b',l}, n_{b',l}\right)}$$
$$\Leftrightarrow \frac{-u_n\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)}{u_c\left(c_{b',l}, \frac{F_l n_{b',l}}{F_h}\right)} \leq \frac{-u_n\left(c_{b',l}, n_{b',l}\right)}{u_c\left(c_{b',l}, n_{b',l}\right)}.$$

Finally, using this observation and the first marginal rates of substitution derived above, we note that

$$\frac{-u_n(c_{b,l}, n_{b,l})}{u_c(c_{b,l}, n_{b,l})} \geq \frac{-u_n(c_{b',l}, n_{b',l})}{u_c(c_{b',l}, n_{b',l})},$$

which is a contradiction of the fact that $n_{b',l} > n_{b,l}$ and $c_{b',l} > c_{b,l}$, provided the utility function satisfies the consumption-leisure normality assumption. Therefore, it must be that $n_{n,l} = n_{i,l}$ and $c_{n,l} = c_{i,l}$.

A.10 Mirrleesian policy without discrimination: Simplified implementability constraints

The incentive constraints of the original problem are

$$u(c_{b,a}, n_{b,a}) = \max_{(b',a') \in \Theta_{b,a}} u\left(c_{b',a'}, \frac{F_{a'} n_{b',a'}}{F_a}\right) \tag{11}$$

for all $b = n, i$ and $a = l, h$.

Lemma 6. *Suppose that the allocations $\mathcal{A}$ satisfy $c_{n,l} = c_{i,l} \equiv c_l$ and $n_{n,l} = n_{i,l} \equiv n_l$, $l_{n,h} = L_{n,h}$, and if $\pi_{i,h} > 0$*

$$u(c_{n,h}, n_{n,h}) \geq u(c_{i,h}, n_{i,h}) \tag{12}$$

$$u(c_{i,h}, n_{i,h}) \geq u\left(c_l, \frac{F_l n_l}{F_h}\right) \tag{13}$$

$$u(c_l, n_l) \geq u\left(c_{i,h}, \frac{F_l n_{i,h}}{F_h}\right), \tag{14}$$

or, if $\pi_{i,h} = 0$,

$$u(c_{i,h}, n_{i,h}) \geq u\left(c_l, \frac{F_l n_l}{F_h}\right). \tag{15}$$

Then, the allocations $\mathcal{A}$ satisfy (11).

Note that because $L_{n,h} > 0 = L_{b,a}$ then $(n,h) \notin \Theta_{b,a}$ for $(b,a) = (n,l), (i,h), (i,l)$.

Suppose first that $\pi_{i,h} > 0$. Note that (12), combined with (13) and $x_{n,l} = x_{i,l}$ for $x = c, n$, implies that

$$u(c_{n,h}, n_{n,h}) \geq \max\left\{u(c_{i,h}, n_{i,h}), u\left(c_{n,l}, \frac{F_l n_{n,l}}{F_h}\right), u\left(c_{i,l}, \frac{F_l n_{i,l}}{F_h}\right)\right\};$$

that is, $u(c_{n,h}, n_{n,h}) = \max_{(b',a') \in \Theta_{n,h}} u\left(c_{b',a'}, \frac{F_{a'} n_{b',a'}}{F_h}\right)$ irrespective of $\pi_{i,h}$.

High-skill immigrants can only mimic low-skill workers $\Theta_{i,h} \subset \{(n,l),(i,l)\}$, and then (13) implies that $u(c_{i,h}, n_{i,h}) = \max_{(b',a')\in\Theta_{i,h}} u\left(c_{b',a'}, \frac{F_{a'} n_{b',a'}}{F_h}\right)$. In a similar way, $\Theta_{n,l} \subset \{(i,h),(i,l)\}$ and $\Theta_{i,l} = \{(n,l),(i,h)\}$. Note that because low-skill natives and immigrants have the same allocation, that incentive compatibility is satisfied. Furthermore, (14) implies that $u(c_l, n_l) = \max_{(b',a')\in\Theta_{b,l}} u\left(c_{b',a'}, \frac{F_{a'} n_{b',a'}}{F_l}\right)$. Because low-skill natives and immigrants have the same consumption bundle, whether $\pi_{i,l} > 0$ or $\pi_{i,l} = 0$ is irrelevant.

Finally, suppose that $\pi_{i,h} = 0$. Then, $\Theta_{n,h} \subset \{(n,l),(i,l)\}$, and (15) guarantees incentive compatibility; that is, $u(c_{n,h}, n_{n,h}) = \max_{(b',a')\in\Theta_{n,h}} u\left(c_{b',a'}, \frac{F_{a'} n_{b',a'}}{F_h}\right)$. No low-skill workers can mimic high-skill native workers, so then $\Theta_{n,l} \subset \{(i,l)\}$ and $\Theta_{i,l} = \{(n,l)\}$. Since all low-skill workers obtain the same consumption bundle, incentive compatibility is trivially satisfied.

A.11 Proof of proposition 4

We write the value function for given $\pi_{i,l}, \pi_{i,h}$ as

$$
\begin{aligned}
\mathcal{W}(\pi_{i,l}, \pi_{i,h}) \equiv \max \omega_h \pi_{n,h}[u(c_{n,h}, n_{n,h}) + v(G)] + \omega_l \pi_{n,l}[u(c_l, n_l) + v(G)] \quad \text{s.to.} \\
[\eta_{i,h}] \quad & u(c_{i,h}, n_{i,h}) + v(G) \geq \overline{U}_{i,a} \\
[\eta_{i,l}] \quad & u(c_l, n_l) + v(G) \geq \overline{U}_{i,a} \\
[\lambda] \quad & F(L, N_l, N_h) \geq \sum_b \sum_a \pi_{b,a} c_{b,a} + \sigma(\pi_i) G, \\
[\chi_{n,h}] \quad & u(c_{n,h}, n_{n,h}) \geq u\left(c_{i,h}, \frac{\theta_l n_{n,l}}{\theta_h}\right) \\
[\chi_{i,h}] \quad & u(c_{i,h}, n_{i,h}) \geq u\left(c_l, \frac{\theta_l n_l}{\theta_h}\right) \\
[\chi_l] \quad & u(c_l, n_l) \geq u\left(c_{i,h}, \frac{\theta_h n_{i,h}}{\theta_l}\right),
\end{aligned}
$$

where the variables in square brackets denote the Lagrange multipliers of each constraint.

The envelope condition is

$$\frac{d\mathcal{W}}{d\pi_{i,l}} = \lambda \left[F_l n_l - c_l - \sigma'(\pi_i) G \right].$$

If $\pi_{i,l} > 0$, the following condition must hold:

$$\frac{d\mathcal{W}}{d\pi_{i,l}} = 0 \Leftrightarrow c_l = F_l n_l - \sigma'(\pi_i) G,$$

otherwise $\pi_{i,l} = 0$.

A.12 Skill premium change decomposition

In an equilibrium with allocations $\mathcal{A}$, the skill premium is computed as follows:

$$SP\left(N_l, N_h\right) \equiv \frac{\alpha}{1-\alpha} S^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{N_l}{N_h} \right)^{\frac{1}{\varepsilon}}. \tag{16}$$

The skill premium depends on aggregate labor supplies N_l and N_h, which in turn depend on individual labor supplies and the composition of the labor force :

$$N_a = \pi_{n,a} n_{n,a} + \pi_{i,a} n_{i,a}.$$

Consider two equilibria $\mathcal{A}$ and $\mathcal{A}'$. We can decompose the change in the skill premium between these equilibria as follows:

$$\begin{aligned}
&SP(N_l', N_h') - SP(N_l, N_h) \\
&= \underbrace{SP(\pi_{n,l} n_{n,l}' + \pi_{i,l} n_{i,l}, \pi_{n,h} n_{n,h}' + \pi_{i,h} n_{i,h}) - SP(\pi_{n,l} n_{n,l} + \pi_{i,l} n_{i,l}, \pi_{n,h} n_{n,h} + \pi_{i,h} n_{i,h})}_{(1)} \\
&+ \underbrace{SP(\pi_{n,l} n_{n,l}' + \pi_{i,l} n_{i,l}', \pi_{n,h} n_{n,h}' + \pi_{i,h} n_{i,h}') - SP(\pi_{n,l} n_{n,l}' + \pi_{i,l} n_{i,l}, \pi_{n,h} n_{n,h}' + \pi_{i,h} n_{i,h})}_{(2)} \\
&+ \underbrace{SP(\pi_{n,l} n_{n,l}' + \pi_{i,l}' n_{i,l}', \pi_{n,h} n_{n,h}' + \pi_{i,h}' n_{i,h}') - SP(\pi_{n,l} n_{n,l}' + \pi_{i,l} n_{i,l}', \pi_{n,h} n_{n,h}' + \pi_{i,h} n_{i,h}')}_{(3)}.
\end{aligned}$$

Part (1) captures the effects of changing native labor supply, (2) captures the effects of changing immigrant labor supply, and finally (3) captures the effects of a change in the composition of the labor force.

In this decomposition, we start by changing labor supplies and only then change the composition of the labor force. A word of caution is in order, since the effects are not invariant to the order of the decomposition.