

GEORGE J. BORJAS

Harvard University

Does Immigration Grease the Wheels of the Labor Market?

MOST STUDIES OF the economic impact of immigration are motivated by the desire to understand how immigrants affect various dimensions of economic status in the population of the host country. This motivation explains the persistent interest in determining whether immigrants “take jobs away” from native workers, as well as the attention paid to measuring the fiscal impact that immigration inevitably has on host countries that offer generous welfare benefits.¹

For the most part, the existing literature overlooks the factor that places immigration issues and the study of labor mobility in general at the core of modern labor economics. The analysis of labor flows, whether within or across countries, is a central ingredient in any discussion of labor market equilibrium. Presumably, workers respond to regional differences in economic opportunities by voting with their feet, and these labor flows improve labor market efficiency.

In this paper I emphasize this different perspective to analyzing the economic impact of immigration: immigration as grease on the wheels of the

I am grateful to Donald Davis, Richard Freeman, Edward Glaeser, Daniel Hamermesh, Lawrence Katz, Dani Rodrik, Mark Rosenzweig, Robert Shimer, Robert Topel, Steven Trejo, and Andrew Weiss for helpful comments, and to the Smith Richardson Foundation and the National Science Foundation for research support.

1. Borjas (1999b), Friedberg and Hunt (1995), and LaLonde and Topel (1997) survey this voluminous literature. Recent studies of the impact of immigration on native labor market opportunities include Borjas, Freeman, and Katz (1997), Card (2001), and Schoeni (1997); recent studies of the fiscal impact of immigration include Borjas and Hilton (1996), Smith and Edmonston (1997), and Storesletten (2000).

labor market. Labor market efficiency requires that the value of the marginal product of workers be equalized across labor markets, such as U.S. metropolitan areas, states, or regions. Although workers in the United States are quite mobile, particularly when compared with workers in other countries, this mobility is insufficient to eliminate geographic wage differentials quickly. The available evidence suggests that it takes around thirty years for the equilibrating flows to cut interstate income differentials by half.²

I argue that immigration greases the wheels of the labor market by injecting into the economy a group of persons who are very responsive to regional differences in economic opportunities.³ My empirical analysis uses data drawn from the 1950–90 U.S. censuses to analyze the link between interstate wage differences for a particular skill group and the geographic sorting of immigrant and native workers in the United States. The evidence shows that interstate dispersion of economic opportunities generates substantial behavioral differences in the location decisions of immigrant and native workers. New immigrant arrivals are much more likely to be clustered in those states that offer the highest wages for the types of skills that they have to offer. In other words, new immigrants make up a disproportionately large fraction of the “marginal” workers who chase better economic opportunities and help equalize opportunities across areas. The data also suggest that wage convergence across geographic regions is faster during high-immigration periods. As a result, immigrant flows into the United States may play an important role in improving labor market efficiency.

The paper presents a simple theoretical framework for calculating this efficiency gain from immigration. Simulation of this model suggests that the efficiency gain accruing to natives in the United States—between \$5 billion and \$10 billion annually—is small relative to the overall economy, but not relative to earlier estimates of the gains from immigration (which are typically below \$10 billion). It seems, therefore, that the measurable benefits from immigration are significantly magnified when estimated in the context of an economy with regional differences in marginal product, rather than in the context of a one-region aggregate labor market.

2. Barro and Sala-i-Martin (1991, 1992); Blanchard and Katz (1992).

3. The analysis is similar in spirit to Card and Hyslop’s (1997) investigation of the hypothesis that inflation greases the wheels of the labor market by making it easier for employers to adjust real wages downward; see also Tobin (1972).

Framework

The intuition underlying the hypothesis developed in this paper is easy to explain.⁴ There exist sizable wage differences across regions or states in the United States, even for workers with particular skills looking for similar jobs.⁵ Persons born and living in the United States often find it difficult (that is, expensive) to move from one state to another. Suppose that migration costs are, for the most part, fixed costs, and that these are relatively high. The existing wage differentials across states may then fail to motivate large numbers of native workers to move, because the migration costs swamp the interstate differences in income opportunities. As a result, native internal migration will not arbitrage interstate wage differentials away.

In contrast, newly arrived immigrants in the United States are a self-selected sample of persons who have chosen to bear the fixed cost of the geographic move. Suppose that once this fixed cost is incurred, it costs little more to choose one state as the destination over another. Income-maximizing immigrants will obviously choose the destination that offers the best income opportunities. Newly arrived immigrants will then tend to live in the “right” states, in the sense that they are clustered in the states that offer them the highest wages.

In short, the location decisions of immigrant workers should be much more responsive to interstate wage differentials than those of natives. As a result, immigrants may play a crucial—and neglected—role in a host country’s labor market: they are “marginal” workers whose location decisions arbitrage wage differences across regions. The immigrant population may therefore play a disproportionate role in helping the national labor market attain an efficient allocation of resources.

The Location Decisions of Native Workers and Immigrants

This hypothesis can be formalized as follows. Consider initially the interstate migration decision faced by workers born in the United States. Let w_{jk} be the wage paid in state j to a native worker with skills k (for

4. Borjas (1999a) first developed some of the implications of this argument in the context of immigrant and native responses to interstate differences in welfare benefits.

5. Karoly and Klerman (1994) investigate the contribution of regional wage differentials to overall trends in U.S. wage inequality.

example, a worker with a high school diploma). The worker currently lives in state b . The sign of the index function determines the worker's internal migration decision:

$$(1) \quad I = \max_j \{w_{jk}\} - w_{bk} - C,$$

where C gives the migration costs. Although these include both variable and fixed costs, I assume that they are mostly fixed. Perhaps the most important fixed cost is the disutility suffered by the migrant who leaves family and friends behind and begins life in a new and uncertain environment. The native worker migrates if $I > 0$.⁶

What does the index function in equation 1 imply about the equilibrium sorting of native workers across states? Suppose that the fixed costs of moving are very high, so that the wage gap between the current state of residence and the state offering the highest wage cannot cover the migration costs. In this extreme case, the geographic distribution of native workers is determined solely by the random allocation that occurs at birth and has little to do with interstate differences in economic opportunities. Because native workers do not respond to interstate wage differentials, these differences will persist (in the absence of other equilibrating flows).

Of course, native workers do in fact move from state to state. Some natives will find that the wage differential between the highest-paying state and the current state of residence is sufficient to cover the fixed migration costs. But many others will find that these migration costs act as a wedge, preventing them from taking full advantage of interstate differences in economic opportunities. As a result, the native working population will not be sorted efficiently, and many native workers end up living in states where their marginal product is not maximized.

Capital flows across localities could help to equilibrate the national economy. In the short run, however, moving physical capital—whether

6. Equation 1 implicitly assumes that the national labor market is in disequilibrium, in the sense that different regions offer different opportunities to the same worker. However, regional wage differences may partly reflect compensating factors that penalize or reward workers for various amenities or disamenities in the region where they live (Roback, 1982; Topel, 1986). Even though a particular worker might face different wages in different labor markets, that worker's utility would then be constant across labor markets. The wage differentials that determine the migration decision summarized by equation 1 are those that persist after the analysis has controlled for regional differences in the value of amenities and disamenities.

plant or equipment—across localities is expensive. As a result, the adjustment of capital stocks will depend largely on new investment, a process that is gradual and can take many years. In what follows I simplify the exposition by assuming that the capital stock is fixed.

Immigrants are born in country 0 and are income maximizers. Their index function is

$$(2) \quad I = \max_j \{w_{jk}\} - w_{0k} - C.$$

Since the wage differential between the United States and many other countries far exceeds the differences that exist between regions in the United States, it is likely that many residents of other countries will find it optimal to move to the United States.⁷ More important, the self-selected sample of foreign-born workers observed in survey data collected in the United States is composed of persons for whom the index I defined in equation 2 is positive. Suppose then that a particular immigrant worker chooses to live in state ℓ . For immigrants in the United States, this residential choice *must* satisfy the condition

$$(3) \quad w_{\ell k} = \max_j \{w_{jk}\}.$$

Put differently, immigrants in the United States will reside in the state that pays the highest wage for the skills they possess. Note that the condition in equation 3 holds regardless of the level of fixed costs, the magnitude of interstate dispersion in wages, or the size of the wage differential between the United States and the source country. Relatively high fixed costs (or a relatively high wage in the source country) simply imply that there will be fewer immigrants. But the sample of foreign-born workers who choose to move will still end up in the right state.

This hypothesis has a number of interesting implications. First, because many native workers are “stuck” in the state where they were born, and immigrant workers are clustered in the states that offer the best economic opportunities, immigrants and natives will be observed living in different states. Moreover, different types of immigrants—depending on their

7. Of course, the provisions of immigration policy allow the United States to pick and choose among the many persons who demand entry. But these policy restrictions are binding only on the subsample of foreign-born persons who find it optimal to move to the United States in the first place.

skills—will also be living in different states. In short, the labor supply of immigrant workers to a particular regional labor market should exhibit greater sensitivity to interstate wage differentials than the labor supply of natives.

Second, the group of immigrants whose location decisions are most responsive to regional differences in economic opportunities should be the sample of newly arrived immigrants. Over time, economic opportunities will probably change differently in different states, and the sample of new immigrants will become like the sample of natives in one very specific way: they all get trapped in the state where they reside. As a result, earlier immigrant waves should be found living in different states than the newest immigrants.

Third, the insight that the location decisions of a particular group of workers—recent movers—are most sensitive to interstate wage differences is not specific to immigrants. It applies to *any* group of movers, whether foreign-born or native-born. As a result, the location decisions of the self-selected sample of native workers who have chosen to move across states should also be quite sensitive to interstate wage differentials.

Finally, the clustering effect implicit in equation 3 has important implications for studies of labor market equilibrium and for estimates of the benefits from immigration. Native migration flows, perhaps because of relatively high fixed migration costs, cannot fully arbitrage away the regional wage differences. The immigrant flow, in contrast, is self-targeted to those regions of the country where their productivity is highest. As I will show shortly, this clustering effect greases the wheels of the labor market, by speeding up the process of wage convergence, and improves economic efficiency. It is important to emphasize that these gains from immigration differ conceptually from the productivity gains typically stressed in the literature.⁸ The productivity gains arise because immigrants and natives complement each other in the production process, and estimates of these gains explicitly assume that the national labor market is in a “single-wage” equilibrium.

Obviously, these strong theoretical implications follow from a framework that uses very restrictive assumptions. In particular, I ignore the many factors other than wage differentials that determine the location

8. Borjas (1995) and Johnson (1998) provide extended discussions of the economic benefits from immigration in a one-sector framework.

decisions of both immigrants and natives. For example, the resurgence of immigrant flows into the United States since 1965 has led to the creation of large ethnic enclaves in many American cities, but in the context of this model it is unclear that these ethnic enclaves arise exogenously. For instance, the first immigrant arriving in the United States from country n may have chosen to live in region j because that region maximized his or her income opportunities.⁹ If most workers in a particular national origin group have roughly similar skills, it would not be too surprising if most new immigrants from that source country also settle in region j . But the ethnic networks that link immigrants in the United States with their source countries also help transmit valuable information about income opportunities to potential migrants. These information flows reduce the costs of migration to specific regions for particular ethnic groups and could lead to a different geographic sorting than that predicted by the income maximization model with fixed migration costs. Any empirical analysis of the magnetic effects generated by interstate differences in labor market opportunities, therefore, must incorporate relevant information about these ethnic networks.

Welfare Implications

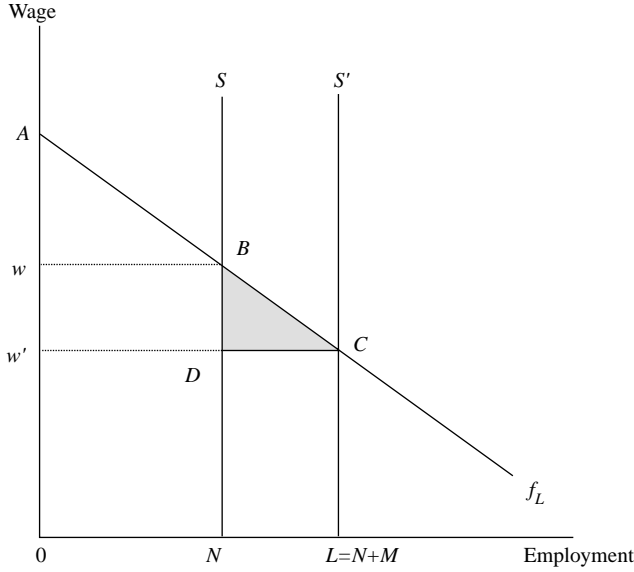
Why does the greater sensitivity of immigrants than natives to regional wage differentials generate economic gains? How large are those gains? And do they accrue to immigrants or to the native population?

Before addressing these questions, it is instructive to review how the benefits from immigration arise in the traditional, one-sector model. Suppose the production technology in the host country can be described by a linear homogeneous aggregate production function with two inputs, capital and labor (L), the price of the output being the numéraire. Suppose further that all workers, whether native or foreign-born, are perfect substitutes in production. Finally, assume that natives own the entire capital stock in the host country and that the supply of all factors of production is perfectly inelastic.

In a competitive equilibrium, the price of each factor equals its marginal product. Figure 1 illustrates the initial preimmigration equilibrium, with N

9. Suro (1998) describes how the migration of a single person from the Guatemalan region of Totonicapan to Houston developed into a flow over the subsequent years, with many of the workers in this immigrant flow ending up in related jobs.

Figure 1. Immigration Surplus in a Single-Region Economy with Homogeneous Labor and Fixed Capital



Source: Author's model as described in the text.

native workers employed at a wage of w . Because the supply of capital is fixed, the area under the curve representing the marginal product of labor (f_L) gives the economy's total output. National income, all of it accruing to natives, is then given by the trapezoid $ABNO$.

The entry of M immigrants shifts the supply curve to S' and lowers the market wage to w' . The area in the trapezoid $ACL0$ now gives national income. Part of the increase in national income is distributed directly to immigrants (who get $w'M$ in labor earnings). The area in the triangle BCD is the increase in national income that accrues to natives, or the "immigration surplus." Note that the immigration surplus arises because natives own all of the capital, and the additional labor raises the return to this fixed capital stock. The immigration surplus, as a fraction of GDP, is¹⁰

$$(4) \quad \text{surplus} = \frac{1}{2} s \delta m^2,$$

10. Borjas (1995, p. 7).

where s is labor's share of national income, δ is the absolute value of the factor price elasticity (or $-d \ln w/d \ln L$), and m is the fraction of the work force that is foreign-born. To illustrate, suppose that labor's share of income is 0.7, that the factor price elasticity is 0.3 (so that a 10 percent increase in labor supply lowers wages by 3 percent), and that immigrants make up 10 percent of the work force (as in the United States today). Equation 4 then implies that the immigration surplus is on the order of 0.1 percent of GDP, or roughly \$10 billion annually.

Now consider the nature of the gains from immigration in a multi-region economy where there are wage differences across regions in the initial equilibrium.¹¹ Suppose the United States has two regions and that the *same* linear marginal product schedule, f_L , gives the labor demand curve in each. The total (and fixed) number of natives in the economy is N , with a fraction λ of the natives living in region 1. For concreteness, assume that $\lambda < 0.5$. Further suppose that labor is supplied inelastically in each region, with supply curves S_1 and S_2 , respectively. As before, natives own the entire capital stock, which is fixed within each region. Figure 2 illustrates the initial equilibrium. The supply imbalance between the two regions implies that w_1 , the wage in region 1, exceeds w_2 , the wage in region 2.

Since capital is fixed in each region, one can write the quadratic production function in region j ($j = 1, 2$) as

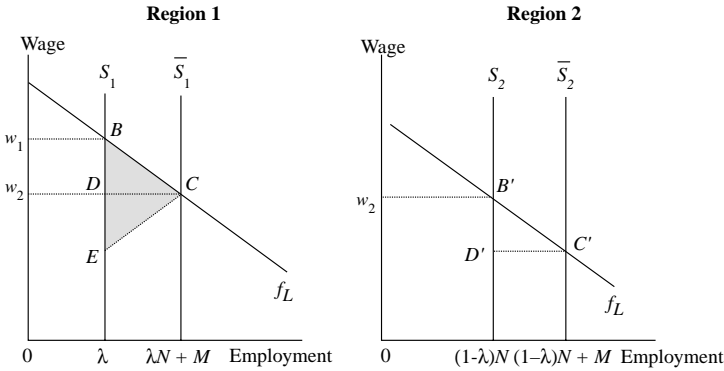
$$(5) \quad Q_j = \alpha L_j - \beta L_j^2,$$

where L_j gives the number of workers in region j , and $\beta > 0$. This quadratic production function generates the linear marginal product curves in figure 2. The initial wage of workers in region 1 equals $w_1 = \alpha - 2\beta\lambda N$, and the wage in region 2 is $w_2 = \alpha - 2\beta(1 - \lambda)N$. These wages are assumed to be positive over the relevant range of employment. I assume initially that natives are immobile, so that the regional wage differential is not arbitrated away by internal migration.

Suppose the United States decides to admit M immigrants. It is useful to write M in terms of the number and geographic distribution of natives in the labor markets. In particular, the difference in the number of natives

11. I am grateful to Robert Topel for raising a number of questions that helped to clarify some of the conceptual issues that arise in this type of framework.

Figure 2. Total Efficiency Gain From Complete Immigration in a Two-Region Economy^a



Source: Author's model as described in text.

a. Assumes homogeneous labor and fixed capital. Under complete immigration, exactly enough immigrants enter region 1 to equalize wages between the two regions.

residing in the two regions is $N_2 - N_1 = (1 - 2\lambda)N$. The number of immigrants can then be written as

$$(6) \quad M = k(1 - 2\lambda) N.$$

The parameter $k = 1$ when the number of immigrants exactly equals the supply imbalance between the two regions. If all of these immigrants were to enter region 1 (as income-maximizing behavior on the part of immigrants would imply), immigration would completely equalize wages between the two regions. In terms of figure 2, this case of “complete immigration” would shift the supply curve in region 1 to \bar{S} , and the single wage in the national economy would be w_2 . For simplicity, I will assume that $0 \leq k \leq 1$ throughout the analysis.

Let θ be the fraction of immigrants who choose to live in region 1. The total number of workers in each region can then be written as

$$(7) \quad L_1 = \lambda N + k\theta(1 - 2\lambda)N,$$

$$(8) \quad L_2 = (1 - \lambda)N + k(1 - \theta)(1 - 2\lambda)N,$$

and GDP in this two-region economy with immobile native workers is given by

$$(9) \quad Q = \alpha[N + k(1 - 2\lambda)N] - \beta[\lambda N + k\theta(1 - 2\lambda)N]^2 - \beta[(1 - \lambda)N + k(1 - \theta)(1 - 2\lambda)N]^2.$$

The parameter θ equals 1 when the geographic sorting of immigrants in the United States is the sorting that maximizes immigrant income. Not surprisingly, this type of immigrant behavior also maximizes GDP for the *entire* U.S. population (which now includes both natives and immigrants). Put differently, Q is maximized at $\theta = 1$ for a given volume of immigration. Figure 2 illustrates the nature of this result for the special case where $k = 1$. The increase in GDP to the entire country if all immigrants were to migrate to region 1 equals the area under the demand curve between points B and C . In contrast, the increase in GDP if all immigrants were to migrate to region 2 equals only the area under the demand curve between the points B' and C' . Comparing these two polar cases makes it clear that the net increase in GDP attributable to optimizing behavior on the part of immigrants is given by the shaded triangle BCD .

In an important sense, this result summarizes the economic content of the statement that immigration greases the wheels of the labor market: income-maximizing behavior leads to a more efficient allocation of resources and maximizes GDP per capita in the host country. This type of immigrant behavior speeds up the process of adjustment to long-run equilibrium, and the larger national output may impart benefits to some sectors of the economy. In the absence of any redistribution mechanism, however, it turns out that the immigrants get to keep much of the increase in GDP that can be attributed to their locating in the high-wage region. As a result, it is important to examine to what extent natives benefit from the fact that income-maximizing immigrants cluster in high-wage regions and thereby improve market efficiency. Consider again the case where natives are immobile. The income accruing to natives is then given by

$$(10) \quad Q_N = Q - w_1M_1 - w_2M_2.$$

The maximization of equation with respect to θ indicates that the relation between Q_N and θ is U-shaped. In fact, the value of Q_N is the same at the

two polar extremes of $\theta = 0$ and $\theta = 1$, and the income accruing to natives is minimized when $\theta = 0.5$, regardless of the value of k . Put differently, natives gain the most when immigrants cluster in one region, regardless of where they cluster, and natives gain the least when immigrants allocate themselves randomly across regions.

Figure 2 also illustrates the intuition behind this result for the special case where $k = 1$. Suppose that all immigrants cluster in the high-wage region ($\theta = 1$). The net gain to natives is then given by the triangle PQR . In contrast, suppose that all of the immigrants end up in the low-wage region ($\theta = 0$). The net gain to natives then equals the triangle $P'Q'R'$, which is obviously equal in area to triangle PQR . The assumption of identical and linear demand curves in the two regions effectively builds in the result that the net gain to natives is the same whenever there is complete clustering, regardless of where immigrants cluster.¹²

This conclusion also depends crucially on the assumption that the native work force is immobile. It is easy to show that natives benefit more when immigrants cluster in high-wage regions as long as natives can move across regions and it is costly to make that internal move. After all, the initial regional wage gap would have eventually motivated some native workers to move across regions. The clustering of income-maximizing immigrants in the high-wage region reduces the number of natives who need to engage in internal migration and hence reduces the migration costs that natives have to incur.

To illustrate this point in a simple framework, suppose that immigrants enter the country first, and that natives then base their internal migration decisions on the postimmigration regional wage gap. Suppose further that, although costly, the internal migration of natives is instantaneous and complete, in the sense that all natives who need to move to equalize wages across regions do so immediately. The number of natives who need to move across regions is then given by

$$(11) \quad R = \frac{L_2 - L_1}{2} = \frac{(1 - 2\lambda)[1 + k(1 - 2\theta)]N}{2}.$$

Define the “net” income accruing to natives as

12. Note that this “neutrality” result hinges crucially on the assumed linearity of the labor demand curves. With constant-elasticity demand curves (as implied by a Cobb-Douglas production function), the convexity of the demand curve would imply that the gains accruing to natives are larger when $\theta = 1$ than when $\theta = 0$.

$$(12) \quad Q_N = Q - wM_1 - wM_2 - C(R),$$

where $w = w_1 = w_2$, and $C(R)$ gives the migration costs associated with R native workers moving across regions, with $C'(R) > 0$. Because natives “fill in” to arbitrage the regional wage gap regardless of where immigrants choose to cluster, it should be evident that the quantity $Q - wM_1 - wM_2$ in equation 12 is independent of θ . In the end, half of the labor force end up in region 1 and half in region 2, and wages are equalized. The relationship between Q_N and θ , therefore, depends entirely on how the geographic sorting of immigrants affects migration costs. Inspection of equations 11 and 12 shows that the larger the fraction of immigrants who cluster in the high-wage region (that is, the greater is θ), the fewer natives need to move across regions, the lower is the level of migration costs, and the larger is the net income that accrues to the native population.¹³

In fact, the increase in migration costs that natives must incur if immigrants are to cluster in the low-wage region can be substantial and may well swamp any benefits resulting from the clustering effect. Let R_0 be the number of natives who would have to move to equate wages if all immigrants clustered in the low-wage region ($\theta = 0$), and let R_1 be the number of natives who would have to move if all immigrants clustered in the high-wage region ($\theta = 1$). Equation 11 then implies that

$$(13) \quad \frac{R_0}{R_1} = \frac{1+k}{1-k}.$$

The implications of equation 13 are easily grasped with a numerical example. Suppose $k = 0.5$, so that half as many immigrants enter the country as complete immigration would require. The ratio in equation 13 then equals 3. In other words, native migration is three times as large when immigrants cluster in the low-wage region as when they cluster in the high-wage region. The additional migration costs, therefore, could easily outweigh the benefits that immigrants impart to natives when they cluster in the low-wage region.¹⁴

13. This discussion implicitly assumes that expenditure on migration vanishes from the economy rather than being transferred to other persons. This is likely to be the case if the main component of migration costs is the disutility associated with leaving a familiar environment and starting over again in a different and unfamiliar area.

14. A numerical example also helps to illustrate this point. Suppose there are 100 million native workers in the economy and that immigration increases the supply of workers by 10 percent, so that there are 10 million immigrants. Suppose further that $\lambda = 0.4$ and

In sum, the endogenous clustering of immigrants in the high-wage region is optimal in two different ways: it increases total national income, *and* it maximizes the income that accrues to natives net of migration costs. By moving the economy from an initial equilibrium with a regional wage gap to a new equilibrium with either a single national wage (in the case of complete immigration) or a smaller regional wage gap, immigrants generate two distinct types of benefits for natives.¹⁵ First, they raise national income through the traditional immigration surplus: because the capital stock is fixed, immigrants increase the profits of native capitalists by more than they lower the earnings of native workers. Second, they help narrow the gap between marginal products in the two regions, maximize the increase in GDP that accrues to natives, *and* reduce the volume of migration costs that natives would have had to incur. It is this second type of gain that results from the fact that immigration greases the wheels of the labor market.

In a multiregion framework *and* for a given volume of immigration, it seems sensible to define the gains that accrue to natives from the geographic sorting of immigrants in another way. How much do natives benefit from the income-maximizing behavior of immigrants relative to how much they would have benefited if the immigrants had chosen locations in some other way? Obviously, this operational definition of the gain is inherently ambiguous, because one must first define the nonoptimal behavior that might determine the geographic distribution of immigrants. Throughout the analysis, I define the baseline as the income that would have accrued to natives if immigrants had simply replicated the geographic sorting of the native population. In other words, suppose that a fraction λ of the immigrants choose to live in region 1. The gain that accrues to natives is then given by

$$(14) \quad \Delta_N = Q_N \Big|_{\theta=1} - Q_N \Big|_{\theta=\lambda}.$$

$k = 0.5$. In this example, 5 million natives would have to move to equalize wages when immigrants cluster in the high-wage region, and 15 million would have to move when immigrants cluster in the low-wage region. Even if the annualized migration cost were as low as \$1,000, the additional migration cost incurred by natives because of the “inefficient” immigrant clustering would be \$10 billion, easily swamping most available estimates of the gains from immigration.

15. This analysis is closely related to the study of the benefits from trade when there is an intersectoral difference in marginal products; see Hagen (1958) and Magee (1972, 1973).

The variable Δ_N includes two distinct types of benefits. First, an immobile native population gains as immigrants cluster rather than replicate the regional distribution of the native population. In the simple framework presented in this section, the benefits that arise from immigrant clustering are the same regardless of whether immigrants cluster in the high-wage or the low-wage regions.¹⁶ However, the fact that natives eventually move in response to interregional wage differences—together with the fact that these moves are costly—implies that Δ_N captures an additional benefit: the reduced costs of internal migration. In the remainder of the paper I will refer to the sum of the two types of benefits captured by the variable Δ_N as the *efficiency gain* from immigration. The efficiency gain thus measures how much natives gain from a fixed volume of immigration simply because immigrants choose to settle in high-wage regions.

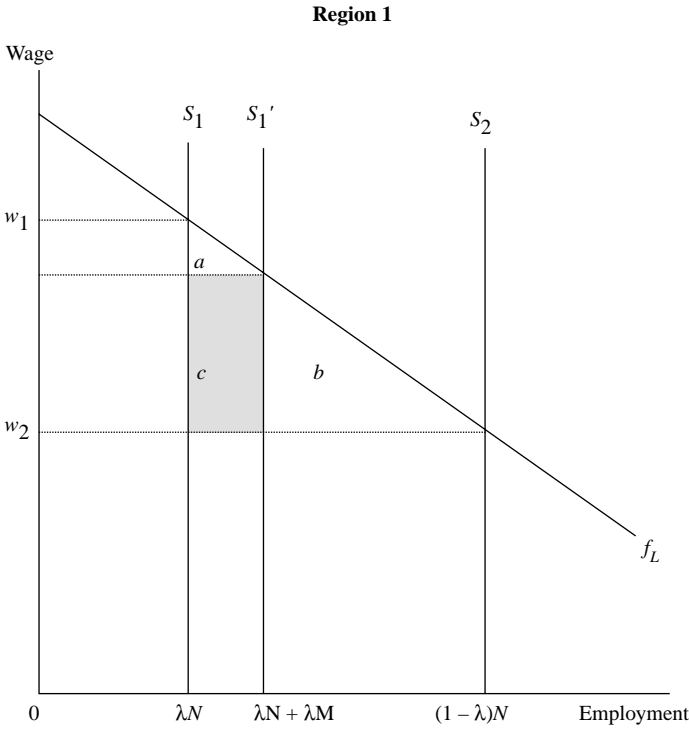
It is worth stressing that the choice of a baseline in equation 14 plays a crucial role in any calculation of the efficiency gain. The arbitrary nature of this choice, however, does not alter an important implication of the analysis: without *any* intervention on the part of the native population, immigration by income-maximizing persons not only maximizes national income but also maximizes the efficiency gain, the additional net income that accrues to the native population.

Figure 3 illustrates the nature of the efficiency gain defined in equation 14. For simplicity, I show the labor market conditions in region 1 only, and I ignore the savings in migration costs. Initially, the labor market is in equilibrium with wages w_1 and w_2 . As drawn, $\lambda = 0.25$, so that one-quarter of the natives live in region 1 and the rest in region 2 (with supply curves S_1 and S_2 , respectively). Suppose there is complete immigration and that all immigrants locate in the high-wage region ($k = 1$ and $\theta = 1$). The economy is now in a single-wage equilibrium, and natives gain by the size of the triangle formed by the union of areas A, B, and D.

Suppose instead that immigrants did not behave optimally in making their location decisions, and suppose further that $\theta = 0.25$, so that immigrants replicate the geographic sorting of the native population. This sorting of immigrants shifts the supply curve in region 1 to S'_1 . The immigrants

16. The notion of an equilibrium where immigrants are clustered in the low-wage area is not as far-fetched as it sounds. During the late 1980s, for example, Sweden routinely placed refugees in regions outside Stockholm, which are coincidentally the country's low-wage regions. See Åslund, Edin, and Fredriksson (2000).

Figure 3. Efficiency Gain Accruing to Natives^a



Source: Author's model as described in text.
 a. Assumes complete migration and an immobile native work force.

who locate in the high-wage region generate a gain of triangular area a for natives, whereas those who locate in the low-wage region generate a gain of area b for natives. It is evident that the optimal sorting increases native income over that with the nonoptimal sorting by the rectangular area c . This rectangle is the gain accruing to natives, for a given volume of immigration, over and above that from a sorting that simply replicates the geographic sorting of the native population.¹⁷

17. Although the discussion has focused on determining how the gains from improved efficiency are distributed between immigrants and natives, there are equally interesting distributional consequences within the native population. As with the immigration surplus in the one-sector model, the gain from immigration defined in equation 14—abstracting from the savings in migration costs—accrue to native capitalists. In contrast, the gains attributable to the savings in migration costs accrue to native workers.

Note that the theoretical framework presented in this section makes extensive use of the assumption that the two regions of the economy have the same labor demand curve. The theoretical implications are less straightforward when the two regions have different demand curves, as could result from underlying differences in the (fixed) endowment of physical capital. It is still the case, of course, that total GDP increases most when immigrants cluster in the high-wage region. Abstracting from the savings in migration costs, however, natives now have the most to gain when immigrants cluster in the region with the more inelastic demand curve.¹⁸ After all, for a fixed number of immigrants, the size of the triangle that accrues to natives is larger when the demand curve is steeper. The region with the more inelastic demand is not necessarily the region with higher wages. However, it should be clear that the bunching of immigrants in the “wrong” region—from the perspective of total economic efficiency—would lead, in the long run, to more native migration and increase migration costs for the native population. It might also greatly reduce any gains arising from immigrants pursuing a location strategy that does not maximize their income.

Finally, the one-period framework summarized in this section shows that the interaction among immigrants clustering in high-wage regions, the regional wage structure, and native internal migration can increase the income accruing to the native population. A more complete description of this interaction requires embedding the income-maximizing behavior of immigrants in a multiperiod model of native internal migration, one that allows for natives to adjust slowly to the presence of regional wage differences. This dynamic model is presented below and used to provide a back-of-the-envelope calculation of the efficiency gain.

Data

I examine the link between interstate wage differentials and the location decisions of immigrants and natives using data from the 1960–90 Public

18. This extension of the model raises a number of very interesting and policy-relevant questions. For example, an inelastic labor demand curve in a particular region may reflect a relatively low volume of physical capital in that region. Immigration policy could then be used to build up the capital stock in the low-capital region by granting entry to persons willing to invest in those regions (and, perhaps not coincidentally, also increase the short-run

Use Microdata Samples of the decennial census. The sample extracts used in the analysis include all civilian workers aged eighteen to fifty-nine who do not live in group quarters. The immigrant extracts form a 5 percent random sample of the population in 1980 and 1990, and a 1 percent random sample in 1960 and 1970. The native extracts form a 1 percent random sample in all years. I define a worker to be an immigrant if he or she was born abroad and is either a noncitizen or a naturalized citizen; all other persons are classified as natives.

I begin the empirical analysis by constructing a log wage index to measure the relative wage of a skill group in a particular state at a particular time. Five skill groups are defined in terms of educational attainment: less than nine years of schooling, nine to eleven years of schooling, twelve years of schooling (high school graduates), thirteen to fifteen years of schooling, and at least sixteen years of schooling (college graduates).¹⁹ The wage index is calculated as follows. Let $w_{ijk}(t)$ be the wage of worker i , residing in state j , belonging to skill group k , in census year t . I then used the sample of *native* workers to estimate the following regression model separately in each census for the years 1960, 1970, and 1980:²⁰

$$(15) \quad \ln w_{ijk}(t) = \mathbf{X}_{ijk}(t) \beta(t) + \mathbf{v}_{jk}(t) + u_{ijk}(t),$$

where $\mathbf{X}_{ijk}(t)$ gives a vector of socioeconomic characteristics indicating the worker's sex and age (defined as a vector of dummy variables indicating

gains that accrue to the native population). This argument can be used to justify the employment creation program in current U.S. immigration policy, where visas are "sold" at lower prices to those persons willing to invest in regions that have relatively poor economic prospects.

19. I also conducted the analysis with the skill groups defined in terms of occupation. The results were qualitatively similar to those reported.

20. Ideally, one would want to use the sample of immigrant workers to predict the wage that a foreign-born worker could expect to earn in each of the states. Because immigrants are highly clustered in a very small number of states, however, there are many empty (j, k) cells, preventing the calculation of a complete series of the log wage index. Moreover, Jaeger's (1996) analysis of the 1980 and 1990 censuses indicates that changes in relative supplies of immigrants and natives within a sex-education group have little effect on the wage gap between immigrants and natives for that group. This evidence suggests that immigrants and natives may be nearly perfect substitutes in production within broad education categories, further justifying the use of the log wage index estimated in the native population to approximate the regional wage dispersion faced by potential migrants. The regression for the 1960 census has 434,195 observations, that for the 1970 census has 567,620, and that for the 1980 census has 860,365.

whether the worker is aged eighteen—to twenty-four, twenty-five to thirty-four, thirty-five to forty-four, forty-five to fifty-four, or fifty-five to sixty-four); $\mathbf{v}_{jk}(t)$ gives a vector of fixed effects for state-education groups (j, k) at time t ; and $u_{ijk}(t)$ is the error term, assumed uncorrelated with all the independent variables in the model. The dependent variable and all the variables in vector \mathbf{X} are normalized to have a mean of zero in each census. The log wage index $v_{jk}(t)$ can then be interpreted as the (adjusted) wage differential, in percent, between the wage in state-education group (j, k) and the mean wage in the United States at time t . Note that the log wage index does not adjust for cost-of-living differences across states. The empirical analysis reported below will control for these differences by including a vector of state fixed effects in second-stage regression models.

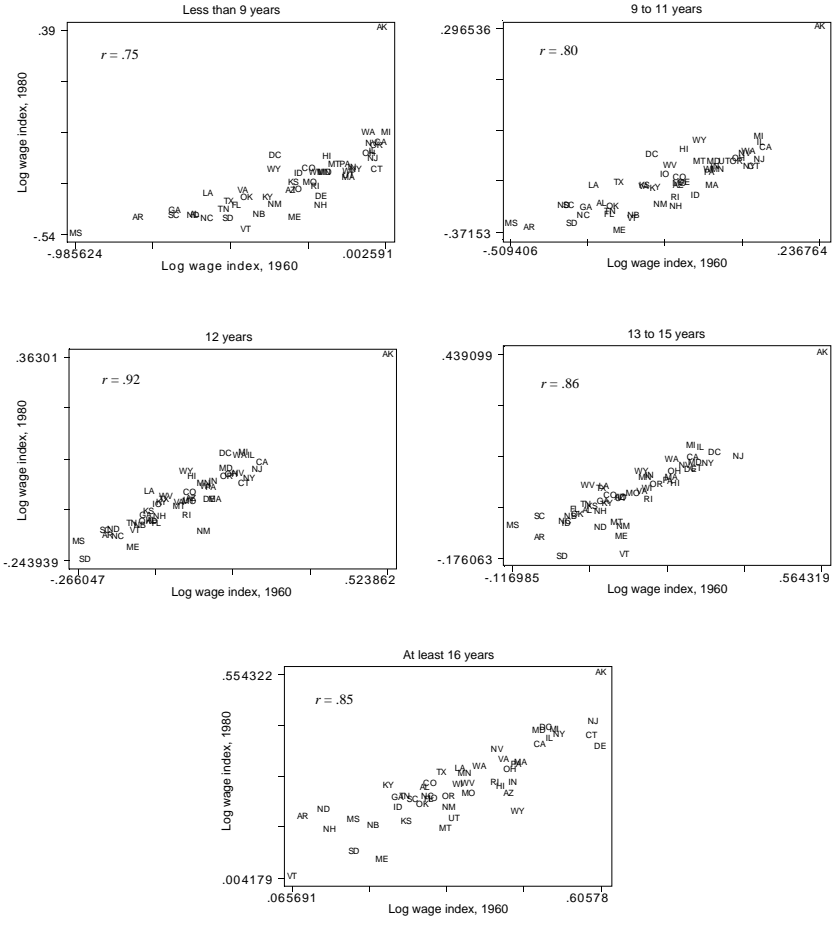
Figure 4 illustrates the interstate variation in the log wage index revealed by the 1960 and 1980 census data. Not surprisingly, there is a great deal of dispersion in adjusted wages across states, so that different states offer different opportunities to similarly skilled workers.²¹ Consider, for instance, the wage opportunities available to a college graduate in 1980. If he or she chose to live in Wyoming, the state at the 20th percentile (so that ten states offered lower wages), the log wage index took on a value of 0.18. If that college graduate chose instead to live in Nevada, the state at the 80th percentile, the log wage index took on a value of 0.35. In other words, interstate wage differentials are sizable even when we ignore the ten states at each end of the wage distribution.

Although high-wage states tend to offer high wages to all workers regardless of educational attainment, this correlation is far from perfect. In 1980 the correlation between the log wage index of workers with less than nine years of schooling and that of college graduates was only 0.72.²² To illustrate how the same state may offer relatively different opportunities to different types of workers, consider the log wage indices for New York and California. Both are high-wage states, but New York's wage advantage is particularly pronounced for highly educated workers. In 1980 sixteen states paid higher wages than New York to workers with less than nine years of schooling, but only five states paid higher wages to college graduates. In contrast, California's wage offer to workers with less than nine

21. Topel (1986).

22. The remaining correlations between college graduates and other workers are as follows: 0.73 for workers with nine to eleven years of schooling, 0.82 for workers with twelve years, and 0.90 for workers with thirteen to fifteen years.

Figure 4. Interstate Dispersion in Adjusted Wages, by educational Attainment, 1960 and 1980^a



Source: Author's calculations based on data for persons aged eighteen to fifty-nine from the 1960 and 1980 Public Use Micro-data Samples of the U.S. Census.

a. Measured by the log wage index, which is the sex- and age-adjusted differential between the log wage of workers in a particular state-education group and the mean log wage in the United States. It is calculated from a log wage function estimated separately for each census year.

years of schooling was the fourth highest in the nation, but its wage offer to college graduates was only the ninth highest.

Finally, figure 4 shows that the relative wages that states offer to workers with particular skills changes over time. The correlations between the 1960 and the 1980 log wage indices range from 0.75 to 0.92, depending on the education group. As a result, workers in a given education group might wish to live in different states at different times. For instance, Washington, D.C., offered the eighth-highest wage to college graduates in 1960. By 1980, however, the District's offer to college graduates had risen to become the third highest in the nation.

There are, therefore, substantial wage differences across states for particular skill groups. This paper argues that immigrants should be particularly responsive to these differences. This hypothesis, however, would seem to contradict a well-known stylized fact: immigrants have clustered and continue to cluster in a relatively few states. In 1990, 74 percent of newly arrived immigrants (those who had been in the country for less than five years) lived in one of the six main immigrant-receiving states: in descending order these are California, New York, Florida, Texas, New Jersey, Illinois. In contrast, only 36 percent of natives lived in those states.

Although this clustering might raise serious doubts about the validity of my argument, it is simply not true that all immigrants cluster in the same states. It turns out that different types of immigrants tend to live in different states, and that the nature of the clustering has changed over time. Table 1 describes the geographic distribution of newly arrived immigrants. In 1990 half of all new immigrants with less than nine years of schooling lived in California, compared with only a quarter of those with a college education. In contrast, 9.2 percent of immigrants with less than nine years of schooling, and 14.9 percent of immigrants who were college graduates, lived in New York. Overall, the data reveal that although fewer than 20 percent of immigrants who were high school dropouts lived outside the six main immigrant-receiving states, almost 40 percent of the immigrants with a college degree did so.

Moreover, the differences in the geographic sorting of immigrants cannot be fully accounted for by the job structures offered by the various states. As the bottom panel of table 1 shows, the ratio of the percentage of immigrants with less than nine years of schooling who live in California to the percentage of similarly skilled natives who live there rose from 4.1 percent to 9.4 percent between 1970 and 1990. In contrast, the same

Table 1. Newly Arrived Immigrant Population in the Six Largest Immigrant-Receiving States, by Educational Attainment, 1970 and 1990

State	<i>Less than nine years</i>		<i>Nine to eleven years</i>		<i>Twelve years</i>		<i>Thirteen to fifteen years</i>		<i>At least sixteen years</i>	
	1970	1990	1970	1990	1970	1990	1970	1990	1970	1990
	<i>Percent of all immigrants</i>									
California	22.0	50.1	22.3	41.9	21.4	32.7	25.7	33.2	17.8	26.5
New York	23.4	9.2	24.6	14.8	24.4	18.6	19.8	14.1	19.1	14.9
Florida	9.1	5.6	6.7	8.1	6.7	8.1	5.8	7.8	3.2	5.0
Texas	5.2	10.1	2.6	7.2	3.2	5.1	2.6	5.4	3.9	5.2
New Jersey	8.9	2.9	7.3	4.1	6.8	5.8	3.7	4.7	4.7	6.3
Illinois	7.0	4.4	5.3	3.9	5.6	5.0	6.3	4.3	7.7	4.7
	<i>Percent of new immigrants relative to percent of natives^a (ratio)</i>									
California	4.1	9.4	2.7	4.9	2.2	4.4	1.7	2.6	1.5	2.2
New York	3.3	1.7	2.8	2.3	2.9	2.9	2.4	2.3	1.9	1.8
Florida	2.9	1.3	2.1	1.6	2.3	1.8	1.9	1.6	1.1	1.2
Texas	0.8	1.2	0.4	1.0	0.7	0.9	0.4	0.8	0.7	0.8
New Jersey	2.9	1.3	2.1	1.5	1.9	1.9	1.2	1.7	1.2	1.6
Illinois	1.4	1.2	1.0	1.0	1.0	1.1	1.1	0.9	1.4	1.0

Source: Author's calculations based on data for persons aged eighteen to fifty-nine from the 1970 and 1990 Public Use Microdata Samples of the U.S. Census.

a. New immigrants are those entering the United States within five years after the log wage index is measured.

ratio for college graduates rose only from 1.5 percent to 2.2 percent during that period. In sum, the stylized fact that most immigrants move to the same states misses an important part of the story: there is a great deal of dispersion in the residential choices made by different types of immigrants.

The theory advanced by this paper suggests that the relative supplies of immigrants and natives to various states will depend on interstate wage differentials. As noted above, I calculated the log wage indices in census year t (where $t = 1960, 1970, 1980$). I now calculate the measures of relative supplies by analyzing the location decisions of immigrants who arrived *soon after* the year in which the log wage index is calculated. Let $M_{jk}(t^*)$ be the number of immigrants who arrived "soon after" time t , reside in state j , and belong to skill group k , and let $M_k(t^*)$ be the total number of new immigrants who belong to that skill group. The group of newly arrived immigrants is composed of persons who entered the United States in the five-year period after the log wage index is measured, so that $t^* = t + 5$. The variables $N_{jk}(t^*)$ and $N_k(t^*)$ give the corresponding numbers of native workers in the state-education groups at that particular time. I

then define the index of relative supply for the state-education group (j, k) at time t^* as²³

$$(16) \quad Z_{jk}(t^*) = \frac{M_{jk}(t^*) / M_k(t^*)}{N_{jk}(t^*) / N_k(t^*)}.$$

The variable $Z_{jk}(t^*)$ measures the relative supply of newly arrived immigrants in education group k to state j . The denominator effectively “deflates” the supply of immigrant workers in a particular skill group to a particular state by the relative importance of that state in the employment of similarly skilled native workers. The relative supply index equals 1 when immigrant and native workers belonging to the same education group have the same geographic distribution. The index would be greater than 1 if immigrants in education group k were overrepresented in state j .

I estimated the index of relative supply for each state-education group using the 1970, 1980, and 1990 censuses. Each census provides information on the person’s state of residence five years before the census. To minimize the amount of time that elapses between the measurement of the log wage index (time t) and the observation of a state of residence for a worker (time t^*), I used the mid-decade measure of location to calculate the relative supply indices. In other words, the 1960 log wage index (calculated from the 1960 census) will be related to the 1965 relative supply measure (calculated from the 1970 census), and similarly for the 1970s and the 1980s.

The cross-sectional relationship between the log wage index and the index of relative supply is summarized in table 2. For each of the five education groups, the table differentiates between the states paying the highest and those paying the lowest wages. The highest-paying states for an education group are those where the log wage index for that group ranks in the top five, and the lowest-paying states are those where the log wage index ranks in the bottom five. For each set of states I then calculated the average index of relative supply. Table 2 strongly suggests a behavioral clustering effect for new immigrants, at least in the cross section. Consider, for example, the geographic distribution of workers who are high school

23. The supplies of immigrants and natives are calculated using the sample of all persons who are not living in group quarters and who are eighteen to fifty-nine years old at the middle of the decade.

Table 2. Relative Labor Supply of New Immigrants in High- and Low-Wage States, by Educational Attainment^aRelative supply index^b

<i>Census year and educational attainment</i>	<i>New immigrants relative to natives</i>		<i>New immigrants relative to earlier immigrants</i>	
	<i>Five highest-wage states</i>	<i>Five lowest-wage states</i>	<i>Five highest-wage states</i>	<i>Five lowest-wage states</i>
<i>1960</i>				
Less than nine years	2.681	0.017	1.237	0.333
Nine to eleven years	1.937	0.262	1.095	2.258
Twelve years	1.688	0.368	1.045	1.792
Thirteen to fifteen years	1.480	0.397	1.046	1.306
At least sixteen years	1.238	0.289	0.890	0.956
<i>1970</i>				
Less than nine years	1.250	0.022	0.770	0.328
Nine to eleven years	1.003	0.065	0.739	0.307
Twelve years	1.029	0.207	0.737	0.670
Thirteen to fifteen years	0.921	0.264	0.790	0.958
At least sixteen years	1.083	0.384	0.885	1.214
<i>1980</i>				
Less than nine years	4.964	0.045	1.210	0.691
Nine to eleven years	2.709	0.082	1.196	0.513
Twelve years	2.360	0.143	1.217	0.523
Thirteen to fifteen years	1.880	0.192	1.134	0.721
At least sixteen years	1.920	0.312	1.143	0.850

Source: Author's calculations based on data for persons aged eighteen to fifty-nine from the 1960–90 Public Use Microdata Samples of the U.S. Census.

a. High- and low-wage states are those with high and low values, respectively, of a log wage index calculated for the indicated census year and skill group. The index measures the sex- and age-adjusted differential between the log wage of workers in the indicated education group in a given state and the mean log wage in the United States. Relative supply indexes are calculated from pooled data for the five highest-wage and the five lowest-wage states. New immigrants are those entering the United States within five years after the log wage index is measured; earlier immigrants have been in the country for at least five years.

b. Calculated as newly arrived immigrants from a particular state-education group as a percent of all newly arrived immigrants in that education group, divided by the same ratio for natives (or earlier immigrants).

graduates. In 1980 the relative supply index in the five “best” states for high school graduates is 2.4. In contrast, new immigrants are relatively absent from the five states that offer the lowest wages for high school graduates: the relative supply index in the “worst” states is 0.14. Generally, new immigrants tend to be overrepresented in the states that offer the highest wages, and underrepresented in the states that offer the lowest wages.²⁴

24. I also calculated the statistics reported in table 2 using alternative breakdowns of the “best” and “worst” states, such as the ten (rather than five) states offering the highest and the ten offering the lowest wages. The qualitative conclusions were very similar.

It turns out that the newest immigrant arrivals are overrepresented in high-wage states not only relative to natives, but also relative to immigrants who arrived in earlier waves. To show this, the last two columns of the table use a slightly different definition of the relative supply index. Let $E_{jk}(t^*)$ be the number of immigrants in earlier waves who reside in state j and belong to skill group k at time t^* , and let $E_k(t^*)$ be the *total* number of these earlier immigrants belonging to that skill group. The earlier immigrants have been in the United States for at least five years before the measurement of the log wage index.²⁵ One can then define an alternative relative supply index:

$$(17) \quad Z'_{jk}(t^*) = \frac{M_{jk}(t^*) / M_k(t^*)}{E_{jk}(t^*) / E_k(t^*)}.$$

Equation 17 provides a simple way of netting out the impact of ethnic networks on residential choice. Suppose that these networks lower the costs of migrating to the areas where the ethnic enclaves are located by transmitting information about economic opportunities in those areas and by providing a welcoming environment (in terms of language and culture) to new immigrants. If these networks are very effective and new immigrants simply move to the areas where their compatriots already reside, the new immigrants should be living in exactly the same places as the older immigrants. In contrast, if new immigrants are more responsive to the changing economic environment, they will be overrepresented relative to the older immigrants in those states that offer the highest wages for their particular skills.

The calculations reported in table suggest that the two cohorts of immigrants locate themselves in somewhat different states, with the new immigrants tending to be overrepresented in those states that offer the best economic opportunities for their skills (this was particularly true in 1980).²⁶ Consider, for example, workers who have between nine and

25. For example, the group of earlier immigrants defined in the 1970 census would include immigrants who arrived in the United States before 1955 (since the relevant log wage index is measured as of 1960).

26. Note, however, that there is little evidence that the new immigrants in the 1960s (that is, the 1960–64 arrivals) cluster in very different places than the earlier immigrants (the pre-1955 arrivals). All of these immigrants, however, arrived before the resurgence of immigration to the United States, and a relatively large fraction were refugees. It is unclear that income maximization plays an important role in determining where these types of immigrants settle when they first enter the country.

eleven years of schooling. In 1980 the relative supply index defined in equation 17 was 1.2 in the five states offering the highest wages, but only 0.5 in the five states offering the lowest.

The descriptive analysis in this section thus finds sizable interstate wage differentials among workers who have similar skills and a tendency for different types of immigrants to live in different states. Most important, the interstate wage differentials seem to be correlated with the residential choices made by new immigrants—relative to the choices made by natives and by earlier immigrant waves.

Empirical Results

Admittedly, this cross-sectional correlation between regional wage differentials and the subsequent location decisions of new immigrants could be spurious, because immigrants, for reasons unrelated to economic opportunities, may simply be moving to those states (such as California and New York) that happen to pay high wages. It turns out, however, that the relative supply of immigrants to a particular state varies over time in response to relative wage changes in that state. Consider the following first-difference regression model:

$$(18) \quad Z_{jk}(t_1^*) - Z_{jk}(t_0^*) = \theta[v_{jk}(t_1) - v_{jk}(t_0)] + \eta_j + \gamma_k + \phi_t + \varepsilon,$$

where η_j is a state fixed effect, γ_k is an education fixed effect, and ϕ_t is a period fixed effect. The unit of observation in this regression is a state-education group during a particular decade. (For example, it may be the change in the relative supply index observed for high school graduates between the early 1960s and the early 1970s, or that for college graduates between the early 1970s and the early 1980s.) The data are pooled across the two decades, so that the regression has 510 observations (two periods \times five education groups (fifty-one states, the fifty-first being the District of Columbia).

To illustrate the nature of the regression in equation 18, consider the observation that measures the change in the relative supply of new immigrants between the 1970s and the 1980s. The variable $Z_{jk}(t_1^*)$ gives the relative supply index (as of 1985) for the new immigrants with skills k who arrived between 1980 and 1984, whereas $v_{jk}(t_1)$ gives the log wage index as

of 1980. Similarly, the variable $Z_{jk}(t_0^*)$ gives the relative supply index (as of 1975) for the new immigrants with skills k who arrived between 1970 and 1974, whereas $v_{jk}(t_0)$ gives the log wage index as of 1970. The parameter q then estimates the sensitivity of the relative supply index to the wage changes that occurred *within* a particular state-education group over the decade.²⁷

Basic Regressions

Table 3 reports estimates of θ from alternative specifications of the model in equation 18. Each coefficient in the table comes from a different specification. All regressions are weighted by $(n_0^{-1} + n_1^{-1})^{-1}$, where n_t gives the number of observations in the state-education group (j, k) at time t .

The coefficients reported in the first row of table 3 are from regressions that do not include the state and education fixed effects in equation 18; those in the second row are estimated from models that do. The state fixed effects play a particularly important role because they control for state-specific changes in the cost of living between 1970 and 1980 (as well as for any other state-specific factors that might change the relative supply of immigrants).²⁸ The regression results are not affected qualitatively by the inclusion of the state and education fixed effects. Therefore I will limit the discussion to the results obtained from the more general (and preferred) specification.

27. The generic first-difference regression model in equation can also be estimated in terms of a level specification. With the data pooled over all three decades, the following regression model can be estimated:

$$Z_{jk}(t) = \theta v_{jk}(t) + (\eta_j \times \varphi_t) + (\gamma_k \times \varphi_t) + (\gamma_k \times \eta_j) + \varepsilon_{jk}(t),$$

where $(\eta_j \times \varphi_t)$ represents a vector of fixed effects that fully interact the cell's state of residence and the period fixed effect; $(\gamma_k \times \varphi_t)$ gives a vector of fixed effects interacting the cell's education and the period fixed effect; and $(\gamma_k \times \eta_j)$ gives a vector of fixed effects interacting the cell's education and the state of residence. If the level regression is weighted by the sample size of the (j, k) cell, the numerical value of the coefficient θ is the same in this level specification as in equation .

28. To see how the state fixed effects in a first-difference regression control for differences in the cost of living, consider a generic regression model $\Delta y_j = a + b \Delta \log(w_j/p_j) + e$, where $\Delta \log w_j$ is the change in the logarithm of the wage in state j over some time period, and $\Delta \log p_j$ is the change in the logarithm of the price index. Note that this regression is equivalent to $\Delta y_j = a + b \Delta \ln w_j - b \Delta \ln p_j + e$. This regression can be estimated either by including a measure of the cost-of-living index (which is not available before the 1970s) or, more generally, by including a vector of state fixed effects.

Table 3. Estimating the Sensitivity of Relative Labor Supply of Immigrants to Interstate Wage Differences^a

<i>Sample and specification</i>	<i>Fixed effects in model?^b</i>	<i>Estimated coefficient on log wage index</i>			
		<i>Immigrants relative to natives</i>	<i>Earlier immigrants relative to natives</i>	<i>New immigrants relative to natives</i>	<i>New immigrants relative earlier immigrants</i>
1960–80	No	0.580 (0.295)	-0.245 (0.277)	1.150 (0.355)	1.149 (0.360)
1960–80	Yes	0.643 (0.253)	-0.560 (0.250)	1.754 (0.373)	2.145 (0.438)
1960–70	Yes	-0.525 (0.585)	-0.551 (0.496)	0.258 (0.936)	1.554 (1.092)
1970–80	Yes	-0.476 (0.831)	-0.613 (0.886)	0.507 (0.556)	3.019 (0.751)
1960–80, men only	Yes	0.817 (0.307)	-0.409 (0.305)	1.993 (0.438)	3.557 (0.971)
1960–80, women only	Yes	0.499 (0.219)	-0.674 (0.230)	1.526 (0.360)	1.830 (0.495)
1960–80, IV ^c	Yes	-2.752 (6.588)	0.327 (5.521)	2.254 (7.936)	44.356 (44.479)

Source: Author's calculations using data for persons aged eighteen to fifty-nine from the 1950–90 Public Use Microdata Samples of the U.S. Census.

a. The dependent variable for each specification is the change in the relative supply index over a given period. The independent variable is the change in the log wage index. Standard errors are reported in parentheses. The multiple-decade specifications pool the data for 1960–70 and 1970–80 and have 510 observations (five education groups in fifty-one states over two periods). These regressions include a dummy variable for the decade from which the observation was drawn. Regressions estimated for single decades have 255 observations. All regressions are weighted by $(n_0^{-1} + n_1^{-1})^{-1}$, where n_t gives the number of observations for the state-education group in year t .

b. Specifications with fixed effects include vectors indicating the state of residence and educational attainment.

c. Instrumented by the previous decade's wage growth in that state-education group.

In the first column of the table, the relative supply index compares, for each state, the *total* number of immigrants—regardless of the year of migration—with the number of natives. In the second column the index compares the number of earlier immigrants with that of natives, and in the third column the index compares the number of new immigrants with that of natives (the relative supply index defined in equation 16). The fourth column uses the relative supply index defined in equation 17, which contrasts the residential locations of new immigrants and earlier immigrants, bypassing the location choices made by native workers altogether.

The positive regression coefficient reported in the first cell of each of the first two rows indicates that the relative supply of immigrants in a particular skill group rose in those states where the wage offered to that skill group was also rising. The sensitivity of immigrant supply to inter-

state wage differentials, however, differs across immigrant groups. In particular, the location choices of earlier immigrants relative to those of natives are not positively related to within-state changes in the log wage index. In contrast, new immigrants are very responsive to wage changes. To get a sense of how responsive they are, it is useful to convert the coefficient θ into a relative supply elasticity, $d \ln Z/d \ln v$. Abstracting from scale effects, this elasticity gives the percentage change in the relative number of immigrants who choose to reside in a particular state for a given percentage change in the wage. This elasticity is given by the ratio θ/Z . The mean value of the relative supply index for new immigrants (in the 1980s) is 1.4. The estimated supply elasticity for new immigrants relative to natives is then 1.3 ($1.754 \div 1.4$).

The regression results reported in the first three columns of table 3 ignore the possibility that nonwage factors, such as the pull of ethnic enclaves on potential migrants, may partly determine the residential choices of immigrants relative to those of natives. As I argued earlier, one can partly net out the impact of ethnic enclaves by comparing the residential locations of new immigrants with those of earlier immigrants. The fourth column of the table reports the regression coefficients from the model that uses this alternative index of relative supply.

The evidence indicates that the location decisions of new immigrants exhibit greater sensitivity to interstate wage differentials than those of earlier immigrants. The relative supply elasticity is also 1.3.²⁹ The analysis thus suggests that, even when one controls for idiosyncratic factors that help determine the location decisions of immigrants in the United States, it is still the case that the supply of new immigrants to particular labor markets is very responsive to the regional wage structure.

In their studies of the impact of immigration on native wages, Borjas, Richard Freeman, and Lawrence Katz, as well as Robert Schoeni, report that the sign of the geographic correlation between past immigration and current wages seems to depend on the period under analysis.³⁰ In some decades the correlation between wage growth in a particular labor market and immigration into that labor market is positive, whereas in other decades it is negative. It is important, therefore, to investigate whether

29. The mean of the relative supply index for new immigrants relative to earlier immigrants is 0.9.

30. Borjas, Freeman, and Katz (1997); Schoeni (1997).

the evidence supporting the excess sensitivity hypothesis holds up in other periods. The third and fourth rows of table 3 reestimate the regressions using the 1960–70 and the 1970–80 periods, respectively. The excess sensitivity hypothesis is roughly consistent with the evidence in each of the decades under analysis, although the results are stronger for the 1970–80 period.

The fifth and sixth rows of table 3 replicate the (pooled) analysis in the samples of men and women, respectively. The relative supply elasticities tend to be more positive for men than for women.³¹ The weaker supply elasticity found among women is consistent with the hypothesis that the location decision of female immigrants is strongly influenced by family considerations. For instance, the sample of female immigrants may contain many tied movers and tied stayers, contaminating the correlation between a worker's wage and the migration decision.³² The sample of female immigrants likely contains many women whose location decisions are tied to those of their foreign-born husbands, or who are tied stayers because they are married to native men and hence face relatively high fixed costs of moving across states.

Finally, there may be some contemporaneous correlation between wage growth in a particular state-education group in any given decade and the residual e in equation 18.³³ One common solution to this problem is to use as an instrument for the decade's wage change the wage change experienced by the same state-education group in the *previous* decade. This procedure requires estimating the log wage index in the 1950 census, to get the change in the log wage index for each group during 1950–60. I reestimated the regression model in equation 14 on data from the 1950 census, using the same sample restrictions and variable definitions as in the analysis of the subsequent censuses.³⁴

31. The mean of the relative supply index for new male immigrants relative to male natives is 1.5, so that the relative supply elasticity in the male sample is 1.3. The mean of the relative supply elasticity for new female immigrants relative to female natives is 1.4, and the relative supply elasticity in the female sample is 1.1.

32. Mincer (1978).

33. One possible source of endogeneity bias could be reverse causation between the relative supply index and the log wage index. For example, the sample of earlier immigrants (who arrived at least five years before the observation of the log wage index) could have themselves affected the wage in the local labor market.

34. The empirical analysis uses the "sample line" observations from the 1950 Public Use Microdata Sample, since these are the only observations that contain information on educational attainment (the sample-line observations form a 1/330th sample). The regressions

The last row of table 3 reports the instrumental variables regression estimates. Although the evidence is qualitatively consistent with the results implied by the ordinary least-squares coefficients, the instrumental variables procedure leads to unstable results, with coefficients that fluctuate dramatically across specifications and have very large standard errors. Part of the problem is that the wage growth experienced by a particular state-education group in decade $\tau - 1$ may not be a valid instrument for the wage growth in decade τ . In fact, there is only a very weak correlation in wage growth across decades. The correlation between wage growth in a given (j, k) cell across subsequent decades is -0.04 .³⁵ Because the log wage index predicted from the first-stage regression is effectively constant within state-education groups, the instrument has relatively little variance relative to the error in the equation.³⁶

Sensitivity of Regression Results

Because the analysis emphasizes the residential location decisions made by immigrants, it is useful to account for some of the characteristics that help differentiate the immigrant population in the United States. As we have seen, immigrants tend to cluster in relatively few states, and this clustering has been particularly pronounced in California. In 1990, 33.8 percent of immigrants resided in California, up from 14.6 percent in 1960. As shown in the last section, California tends to be a high-paying state for all workers, regardless of their educational attainment. Suppose that immigrants cluster in California simply because it is geographically close to some important source countries (such as Mexico), or because it offers amenities (such as a pleasant climate) that provide a more welcoming

estimated in the 1950 census use the sample weights provided in the data file. The log wage regression in the 1950 census has 90,363 observations. Note further that the 1950 public use microdata sample does not contain any information on wages for workers in Alaska and Hawaii, which became states after 1950. As a result, the second-stage relative supply regressions that use the instrumental variables procedure have 500 observations.

35. The correlation is $+0.14$ between 1950–60 and 1960–70, and -0.22 between 1960–70 and 1970–80. Borjas, Freeman, and Katz (1997) show that the correlation between wage growth in 1970–80 and that in 1980–90 became much more negative than over previous decades. It would be difficult to justify the use of the previous decade's wage growth as an instrument for the current decade's growth unless one can first understand why the regional wage structure changed in such a dramatic fashion during that period.

36. In addition, the predicted last period's wage growth is highly collinear with the vector of education and state fixed effects included in the first-difference regression model.

environment for new immigrants. The positive correlation between the relative supplies of immigrants and interstate wage differences within a skill group could then arise spuriously. The first row of table 4, however, shows that the empirical results are not affected when the relative supply regressions exclude all state-education groups that refer to workers in California. Even outside California, new immigrants are more sensitive to interstate wage differentials than either earlier immigrants or natives.

One could also argue that the key results may be driven by the predominance of the Mexican population in the immigrant flow. In 1990 about 22 percent of immigrants living in the United States had been born in Mexico, and 58 percent of these Mexican immigrants were living in California, a state that offered relatively high wages. As the second and third rows of table 4 show, the results are similar when the regressions are reestimated in samples that either exclude Mexican immigrants or exclude both Mexican immigrants and workers living in California.

It is also worth investigating whether the relative supply of immigrants to particular states exhibits excess sensitivity even when we estimate the models within particular national origin groups. To calculate the relative supply indices in sufficiently large samples, the analysis must use broadly defined groups, such as European immigrants, Asian immigrants, and Latin American immigrants. The evidence obtained from these regressions, however, is consistent with the basic thrust of the evidence. The relative supply of new immigrants within these aggregated groups—even when compared with the relative supply of their previously arrived compatriots—shows excess sensitivity to interstate wage differentials.

Finally, I have assumed throughout the analysis that newly arrived immigrants can freely choose their residential location when they enter in the United States. One group of immigrants, however, has relatively little choice over where to reside when they first enter the country, namely, refugees. The U.S. Department of State assigns individual refugees to sponsoring private voluntary agencies that provide them with a variety of social services, including initial resettlement in the United States.³⁷ These sponsoring agencies determine the location of resettlement, which depends partly on the match between a refugee's socioeconomic background and the availability of jobs and services in particular localities—as perceived by the sponsoring agency. As a result, the initial location of refugees could

37. U.S. Department of State (2000).

Table 4. Robustness of Estimated Relative Supply Effect to Changes in Sample Specification^a

<i>Sample</i>	<i>Estimated coefficient on log wage index</i>			
	<i>Immigrants relative to natives</i>	<i>Earlier immigrants relative to natives</i>	<i>New immigrants relative to natives</i>	<i>New immigrants relative earlier immigrants</i>
Workers outside California	0.484 (0.166)	-0.689 (0.172)	1.642 (.401)	2.272 (0.488)
Natives and non-Mexican immigrants	0.811 (0.210)	-0.813 (0.181)	3.042 (0.503)	3.182 (0.542)
Natives and non-Mexican immigrants outside California	0.274 (0.158)	-0.889 (0.197)	1.949 (0.430)	3.400 (0.605)
Natives and European immigrants	-0.260 (0.184)	-0.882 (0.204)	1.847 (0.429)	2.253 (0.675)
Natives and Asian immigrants	0.920 (0.449)	-0.711 (0.532)	1.372 (0.750)	1.568 (0.980)
Natives and Hispanic immigrants	1.276 (0.241)	-0.275 (0.430)	1.616 (0.592)	2.446 (0.965)
Natives and nonrefugee immigrants	0.522 (0.246)	-0.684 (0.249)	1.753 (0.344)	2.003 (0.369)

Source: Author's calculations using data for persons aged eighteen to fifty-nine from the 1950–90 Public Use Microdata Samples of the U.S. Census.

a. Using the 1960–80 specification with fixed effects (see table 3). Standard errors are in parentheses.

well be consistent with the key implications of the model proposed in this paper, but it may not be. Unfortunately, the census does not contain any information on the type of visa used by an individual to enter the United States. To approximate the refugee population, therefore, I classify all immigrants who originate in the main refugee-sending countries as refugees and all other immigrants as nonrefugees.³⁸

The last row of table 4 reports the coefficients estimated from the sample of nonrefugees. The evidence clearly indicates that the location decision of nonrefugees is consistent with the hypothesis proposed in this

38. Thirteen countries account for 90 percent of refugees awarded permanent residence status during the 1970s and 1980s: Afghanistan, Bulgaria, Cambodia, Cuba, Czechoslovakia, Ethiopia, Hungary, Laos, Poland, Romania, Thailand, the former Soviet Union, and Vietnam.

paper: they are disproportionately more likely to choose to live in the high-wage states. It would be of interest to estimate the regression model in the sample of refugees. The refugee population, however, is relatively small, hovering around 10 percent of the immigrant population during the period under study. The data lack sufficient variation, therefore, to allow a full-scale study of the relationship between the location of refugees (as chosen by the sponsoring agency) and interstate wage differentials.

Native Internal Migration

Up to this point, the empirical analysis has emphasized that the self-selected sample of new immigrants is very responsive to interstate wage differentials. In fact, the basic idea developed in this paper applies equally well to the self-selected sample of native workers who choose to move across states. Just as new immigrants should be clustered in the states that offer the highest wages for their types of skills, so the self-selected sample of native movers should also be clustered in those states that offer the highest wages for *their* skills.³⁹

The census reports a worker's state of residence five years before the census date.⁴⁰ One can then use these data to create two alternative samples of natives: native movers, who moved across state lines during the relevant five-year period, and native stayers, who remained in the same state over the five-year period. It would seem ideal to replicate the regression analysis in equation 18 by defining a new relative supply index that compares the location decisions of the native movers with those of native stayers. Unfortunately, this comparison is not quite right. Consider the data available for the 1980s. The 1980 log wage index was calculated using the 1980 census. The 1990 census provides the eventual geographic location of native movers—as of 1990—for those natives who moved between 1985 and 1990. The sample of native stayers that can be defined in these data,

39. This alternative test of the theory could help strengthen the argument that the results reported in tables 3 and 4 *cannot* be attributed to a spurious correlation created by the possibility that immigrants tend to move to a few states, which just happen to offer relatively high wages. Native movers, after all, do not typically move to the same six immigrant-receiving states. Only 34.1 percent of natives who moved between 1985 and 1990 moved to one of those states, compared with 74 percent of the new immigrant arrivals.

40. The 1980 census data contain information on place of residence in 1975 only for half of the sample, randomly selected. The analysis summarized below, therefore, uses a 0.5 percent random sample of native workers in 1980.

therefore, contains both natives who did not move after 1985, and natives who moved between 1980 and 1985 but did not move thereafter. In short, the so-called sample of native stayers is contaminated by the presence of some native movers.

In fact, the regression coefficients reported in the first column of table 5 suggest that native movers are *less* responsive to interstate wage differences than native stayers. One can try to circumvent the measurement problem by using alternative control groups. The second column of the table refines the definition of native stayers by estimating the relative supply index only in the subsample of stayers who are forty to fifty years old—that is, among workers who are past their prime years for internal migration but before the retirement age. This subsample of stayers, therefore, is likely to contain relatively few movers. In fact, there is now a very weak positive correlation between the relative supply index and relative wages in the fixed-effects specification, suggesting that native movers are somewhat more likely to reside in those states that pay the highest wages for the skills they have to offer.

One way of avoiding the measurement problem altogether is simply to ignore the group of native stayers and compare the native movers with immigrants. The third column of table 5 defines the relative supply index by contrasting the location decisions of native movers with those of earlier immigrants. As long as there are substantial migration costs, the sample of earlier immigrants is “stuck” in the state where they first entered the United States, so that this sample approximates an immobile native sample. Table 5 reports that the relative supply of native movers is more sensitive to interstate wage differentials than that of the earlier immigrants.

Finally, one can also define the relative supply index by contrasting the location of native movers with that of newly arrived immigrants. If the two groups were equally responsive to interstate wage differentials, the estimated impact of the log wage index would be zero. The regression coefficient, however, is negative and significant when the fixed effects are included, indicating that newly arrived immigrants are more responsive to interstate wage differentials than the self-selected sample of native movers.

Overall, the evidence reported in table 5 does not support the hypothesis that native movers behave “just like” new immigrants. Part of the problem arises from the data: the construction of the census data prevents the correct calculation of the relevant group of native stayers. It may also be

Table 5. Estimating the Sensitivity of Relative Labor Supply of Native Internal Migrants to Interstate Wage Differences^a

<i>Sample</i>	<i>Estimated coefficient on log wage index</i>			
	<i>Native movers relative to native "stayers"</i>	<i>Native movers relative to older native "stayers"</i>	<i>Native movers relative to earlier immigrants</i>	<i>Native movers relative to new immigrants</i>
Without fixed effects	-1.064 (0.184)	-1.280 (0.236)	3.421 (1.895)	-9.919 (7.406)
With fixed effects	-1.372 (0.221)	0.517 (0.441)	3.716 (2.406)	-22.005 (9.478)

Source: Author's calculations using data for persons aged eighteen to fifty-nine from the 1960–90 Public Use Microdata Samples of the U.S. Census.

a. Using the 1960–80 specification with fixed effects (see table 3). Standard errors are in parentheses. Older native stayers are those aged forty to fifty. New immigrants are those entering the United States within five years after the log wage index is measured; earlier immigrants have been in the country for at least five years.

the case, however, that the data simply are not consistent with the theory that native movers are income maximizers. The available evidence in the internal migration literature suggests that natives respond to interstate wage differentials and move to high-wage states. However, much of this evidence is somewhat dated and does not directly address the possibility that, in recent years, net native migration to Sunbelt states may have been motivated by reasons other than income maximization.⁴¹ This possibility increases the importance of immigrant flows in arbitraging regional wage differentials.

Between 1985 and 1990, 10.9 million native workers (11.0 percent of the native work force) moved across state lines. During the same period, 880,000 "preexisting" immigrants (11.3 percent of the immigrants who arrived before 1985) also moved across state lines.⁴² Finally, 2.1 million new immigrant workers entered the U.S. labor market during the same period. The new immigrants, therefore, accounted for 15.1 percent of all workers who moved across state lines between 1985 and 1990, even though these workers only accounted for 1.9 percent of the work force in 1990. The flow of new immigrants into the United States, therefore, has a

41. Greenwood (1997) surveys the extensive internal migration literature.

42. Bartel (1989), in a study of the internal migration decisions of immigrants, concludes that immigrants enter the United States through a limited number of gateway cities and that their subsequent internal migration within the United States is not very sensitive to regional wage differentials.

disproportionate impact on the evolution of the regional wage structure—even if all native movers are income maximizing when making their migration decisions. The new immigrants make up many of the marginal workers whose location decisions help equilibrate the national labor market and improve labor market efficiency.

Immigration and Wage Convergence

The labor supply responses documented in the previous section suggest that the geographic distribution of new immigrants in the United States should help reduce regional wage differentials. This finding has significant macroeconomic implications, for it suggests that regional wage convergence will be faster among those skill groups and in those periods that experience high levels of immigration.

A large literature examines the rate of wage convergence across states in the United States.⁴³ These studies typically estimate the generic regression model

$$(19) \quad \ln w_{j,t+1} - \ln w_{jt} = \alpha + \beta \ln w_{jt} + \varepsilon_{jt},$$

where w_{jt} is the wage in state j at time t , and β is the convergence coefficient, which will be negative if there is wage convergence across states. The values of β (measured at an annual rate) estimated in the literature range around -0.02 , suggesting that the half-life of interstate wage differentials is roughly thirty-five years.⁴⁴ The hypothesis examined in this paper

43. See, for example, Barro and Sala-i-Martin (1991, 1992); Blanchard and Katz (1992).

44. To derive the steady-state level of income in the national economy, we can write the difference equation in terms of the initial wage in the state. Ignoring the subscript j , the wage at time t is

$$\ln w(t) = \alpha[1 + (1 + \beta) + (1 + \beta)^2 + \dots + (1 + \beta)^{t-1}] + (1 + \beta)^t \ln w(0),$$

where $\ln w(0)$ is the logarithm of the initial wage. The steady-state wage is obtained by letting $t \rightarrow \infty$. It is given by $\ln w^* = -\alpha/\beta$. One can then write the current period wage as

$$\ln w(t) = [1 - (1 + \beta)^t] \ln w^* + (1 + \beta)^t \ln w(0).$$

The initial log wage gap has been cut by half when $\ln w(t)$ is the simple average of $\ln w^*$ and $\ln w(0)$. This implies that the half-life can be calculated from the equation $(1 + \beta)^t = 0.5$. The time required to cut the initial wage gap in half is then given by $t^* \approx -\ln 2/\beta$.

suggests that there exists a structural relationship linking the coefficient β in any particular labor market and the size of the immigrant flow into that market between periods t and $t + 1$.

I use data drawn from the 1950–90 censuses to estimate a particular specification of the regression model in equation 19. Recall that my empirical framework has differentiated across skill groups throughout the analysis. Consider the following convergence regression model:

$$(20) \quad v_{jk,t+1} - v_{jkt} = \alpha_{kt} + \beta_{kt}v_{jkt} + \varepsilon_{jkt},$$

where $v_{jk}(t)$ is the log wage index estimated for workers in state j and skill group k in census year t , and the parameter β_{kt} is the convergence coefficient describing the evolution of the regional wage structure for that skill group in the $(t, t + 1)$ time period. I annualized the dependent variable by dividing by ten, so that b gives the annual rate of regional wage convergence. I estimated the regression model in equation separately for each skill group in each time period 1950–60, 1960–70, 1970–80, and 1980–90. The empirical analysis, therefore, yields a total of twenty estimated convergence coefficients (five education groups over four periods).

Table 6 summarizes the convergence coefficients estimated for each of the cells. Sixteen of the coefficients are negative, and only one of the four positive coefficients is significantly different from zero. The mean convergence coefficient (weighted by the inverse of the square of the standard error) is -0.02 , which is quite similar to the “consensus” convergence coefficient reported in the literature.

The hypothesis developed in this paper suggests that there should be greater wage convergence in those labor markets (defined for a particular skill group over a particular time period) that experience larger immigrant flows. Let the index of immigrant penetration in a particular labor market be given by

$$(21) \quad g_{kt} = \ln \frac{M_k(t, t+1)}{N_{kt}},$$

where $M_k(t, t + 1)$ is the total number of immigrants in skill group k who entered the United States between periods t and $t + 1$, and N_{kt} is the native-born population in skill group k at the beginning of the period. I then test the theoretical implication that high immigration leads to faster wage convergence by estimating the second-stage regression model

Table 6. Estimated Wage Convergence and Immigrant Penetration, by Educational Attainment

<i>Educational attainment</i>	<i>Convergence coefficient^a</i>	<i>Standard error</i>	<i>Immigrant penetration^b</i>
<i>1950–60</i>			
Less than nine years	0.004	0.004	-4.231
Nine to eleven years	0.009	0.006	-4.320
Twelve years	-0.004	0.004	-4.316
Thirteen to fifteen years	-0.014	0.007	-3.836
At least sixteen years	-0.030	0.006	-3.633
<i>1960–70</i>			
Less than nine years	-0.025	0.003	-3.574
Nine to eleven years	-0.022	0.003	-4.233
Twelve years	-0.016	0.004	-3.987
Thirteen to fifteen years	-0.005	0.005	-3.547
At least sixteen years	-0.004	0.006	-3.123
<i>1970–80</i>			
Less than nine years	-0.040	0.003	-2.625
Nine to eleven years	-0.046	0.004	-3.723
Twelve years	-0.037	0.005	-3.764
Thirteen to fifteen years	-0.045	0.005	-3.069
At least sixteen years	-0.037	0.004	-2.720
<i>1980–90</i>			
Less than nine years	-0.018	0.008	-1.821
Nine to eleven years	-0.008	0.014	-2.784
Twelve years	-0.002	0.015	-3.658
Thirteen to fifteen years	0.014	0.015	-2.898
At least sixteen years	0.023	0.011	-2.693
Sample mean, 1950–90	-0.022	n.a.	-3.571

Source: Author's calculations using data for persons aged eighteen to fifty-nine from the 1950–90 Public Use Microdata Samples of the U.S. Census.

a. Estimated by equation 20 and weighted by $(n_0^i + n_1^i)^{-1}$, where n_t gives the number of observations for the state-education group in year t .

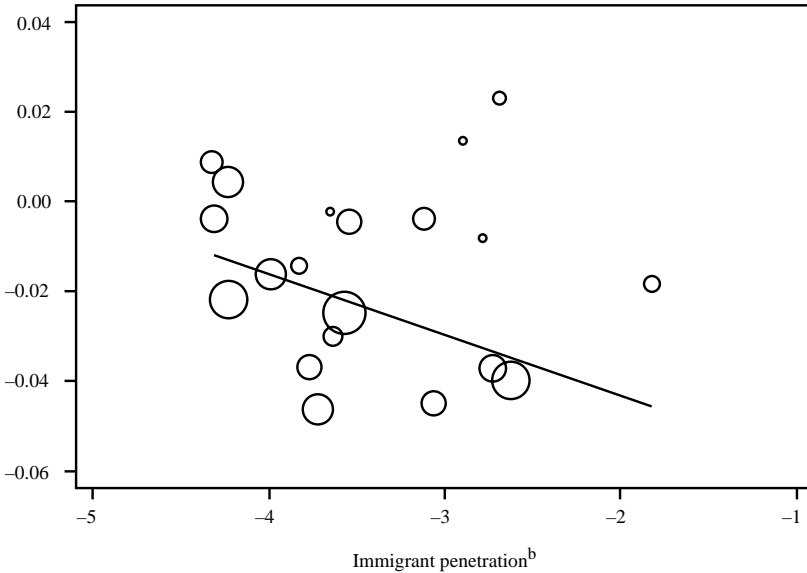
b. Measured as the log ratio of the total number of immigrants in a particular skill group who entered the United States over the sample period to the size of the native-born population in that skill group at the beginning of the period.

$$(22) \quad \beta_{kt} = a + bg_{kt} + \xi.$$

Because the dependent variable in equation 22 is subject to sampling error, I weight the regression by the inverse of the square of the standard error. This procedure helps adjust for any potential heteroskedasticity in the disturbance.

Figure 5 presents a scatter diagram illustrating the basic data available for the second-stage regression. (The area of the circles corresponds to the weights of the convergence coefficients as described above.) There is a

Figure 5. Relationship between Wage Convergence and Immigrant Penetration, 1950–90



Source: Author's calculations based on data for persons aged eighteen to fifty-nine from the 1950–90 Public Use Microdata Samples of the U.S. Census.

a. Each circle represents a convergence coefficient estimated in a particular decade for a particular skill group. Estimates are weighted by the inverse of the square of the standard error of the convergence coefficient; the larger the area of the circle, the larger the weight of that estimate in the regression.

b. Immigrant penetration is measured as the log ratio of the number of immigrants in a particular skill group who arrived during a particular decade to the number of natives in that skill group at the beginning of the decade.

clear downward-sloping relationship between the convergence coefficient in any particular market (defined by time period and skill group) and the relative number of immigrants entering that labor market. Table 7 reports the results of estimating the basic regression models on these data. The simplest specification, reported in the first row, takes each of the twenty labor markets as independent observations; a numerically strong and statistically significant correlation is found between convergence and immigration.

To illustrate the magnitude of the correlation, consider the following thought experiment. Suppose the convergence coefficient is on the order of -0.02 , implying a half-life of regional wage differences of thirty-five years. In the 1950s the average value of m_k (across skill groups) was equal to -4.1 . Put differently, immigration during this decade increased labor sup-

Table 7. Explaining Wage Convergence with Immigrant Penetration^a

<i>Specification</i>	<i>Coefficient on immigrant penetration^b</i>	<i>R²</i>	<i>No. of observations</i>
Without fixed effects	−0.013 (0.006)	0.229	20
With education fixed effects	−0.021 (0.008)	0.366	20
With education and period fixed effects	−0.001 (0.005)	0.753	20

Source: Author's calculations using data for persons aged eighteen to fifty-nine from the 1950–90 Public Use Microdata Samples of the U.S. Census.

a. The dependent variable for each specification is the convergence coefficient estimated in a particular decade for a particular skill group (see table 6). The independent variable is the measure of immigrant penetration for that decade and skill group. Standard errors are reported in parentheses. Regressions are weighted by the inverse of the square of the standard error of the estimated convergence coefficient.

b. See table 6.

ply by around 1.6 percent. By the 1980s the average value of m_k was -2.5 , and total immigration increased labor supply by around 9.8 percent. The magnitude of the coefficient b in equation ($b = -0.013$) implies that this increase in immigration would lower the convergence coefficient by 0.021 unit, or from -0.02 to -0.04 . The half-life of interstate wage differentials would then be cut in half, to 17.3 years. Immigration, therefore, seems to significantly speed up the process of wage convergence. And this speeding up occurs precisely because immigration introduces into the U.S. labor market a sizable group of persons who are particularly responsive to interstate wage differentials.

The remaining rows of table 7 show estimates of the basic second-stage model using increasingly general specifications. The second row, for example, adds education fixed effects to the basic model. The addition of these education fixed effects implies that the relationship between wage convergence and immigration is now being estimated *within* education groups—identifying the impact of immigration from the time variation in the size of the immigrant flow for a particular skill group. The coefficient b remains negative and is actually larger, again suggesting that immigration plays an important role in determining the speed of wage convergence in the U.S. labor market.

The last row of table 7 adds variables indicating period fixed effects (that is, the decade in which the convergence coefficient was estimated). The coefficient b remains negative, but it is no longer statistically significant.

This is not too surprising since there is little variation left in the sample of twenty convergence rates to identify the relationship between immigration and the speed of wage convergence.

Overall, the empirical analysis of interstate wage convergence yields results consistent with the key hypothesis of this paper. Immigration greases the wheels of the labor market by introducing workers who are very responsive to regional wage differentials into the host country's labor market. As a result, immigrants speed up the process of wage convergence and improve labor market efficiency.

Estimating the Efficiency Gain

Equation 14 defined the efficiency gain from immigration as the net increase in GDP accruing to natives that results from the clustering of immigrants in high-wage regions relative to what would have been observed if immigrants had simply replicated the geographic sorting of the native population. The calculation of the efficiency gain requires that I first specify how natives respond to regional wage differentials. In particular, how would the regional wage structure have evolved in the absence of immigration? And how do natives respond to the entry of immigrants in different markets?

The speed of adjustment by natives plays an important role in calculations of the efficiency gain. Suppose, for example, that natives are very responsive to regional wage differentials and that migration costs are relatively low. Then the entry of M immigrants into the high-wage region generates only a short-run efficiency gain, because natives would have eliminated the wage gap between the two regions quickly and at very low costs. In this scenario the fact that immigration greases the wheels of the labor market is unlikely to play an important role in a cost-benefit analysis of immigration. If, however, migration costs are relatively high, and relatively few natives move across regions, the efficiency gain generated by immigration could be sizable and long lasting.

A simple model helps to formalize these ideas.⁴⁵ Suppose the United States has two regions ($j = 1, 2$) and that the production function is the same in each:

45. The framework is a special case of the multisector migration model presented in Borjas (1999b).

$$(23) \quad Q_{jt} = A L_{jt}^{\beta},$$

where Q_{jt} is the value of output in region j at time t , L_{jt} is employment, and $\beta < 1$. The regional labor demand function implied by this technology is

$$(24) \quad \ln w_{jt} = \alpha - \delta \ln L_{jt},$$

where w_{jt} is the wage in region j at time t , and $\delta = 1 - \beta > 0$. A fraction $\lambda < 0.5$ of the native population lives in region 1 at time 0. The total number of native workers in the labor market, $N = L_{10} + L_{20}$, is assumed fixed. The wage gap between the two regions before any immigrants enter the economy is

$$(25) \quad \ln w_{10} - \ln w_{20} = -\delta \ln \frac{\lambda}{1 - \lambda}.$$

Suppose that the United States admits M immigrants at the end of time 0 and that a fraction θ of these immigrants settle in region 1. The numbers of immigrants who settle in regions 1 and 2, respectively, are given by

$$(26) \quad M_1 = \theta k(1 - 2\lambda) N,$$

$$(27) \quad M_2 = (1 - \theta) k(1 - 2\lambda) N.$$

As before, the parameters k and θ help to measure the extent to which immigrants arbitrage wage opportunities across the two labor markets. In the special case of $k = 1$, the immigrant flow is sufficiently large to potentially bring the two labor markets into a single-wage equilibrium without any native internal migration. If all of these immigrants settled in the high-wage region, so that θ is also equal to 1, immigration would be “complete” and would bring about an immediate equalization of wages across regions.

Regional wage differentials induce a supply response in the native-born population, but this response occurs with a lag. For simplicity, assume that only natives move across regions and that their supply response is described by the following lagged adjustment function:

$$(28) \quad \ln \frac{L_{1,t+1}}{L_{2,t+1}} - \ln \frac{L_{1t}}{L_{2t}} = \sigma(\ln w_{1t} - \ln w_{2t}),$$

where $\sigma \geq 0$ is the labor supply elasticity. In other words, the rate of change in the regional allocation of labor between times t and $t + 1$ depends on the regional wage gap at time t .

Some tedious algebra reveals that the adjustment mechanism in equation implies that the regional employment and wage differentials at time t are given by

$$(29) \quad \ln L_{1t} - \ln L_{2t} = (1 - \delta\sigma)^t \ln \frac{\lambda + \theta k(1 - 2\lambda)}{(1 - \lambda) + (1 - \theta)k(1 - 2\lambda)},$$

$$(30) \quad \ln w_{1t} - \ln w_{2t} = -\delta(1 - \delta\sigma)^t \ln \frac{\lambda + \theta k(1 - 2\lambda)}{(1 - \lambda) + (1 - \theta)k(1 - 2\lambda)}.$$

I assume that $0 < (1 - \delta\sigma) < 1$ so that the national labor market converges to a single-wage equilibrium over time.

Finally, it is costly for natives to engage in internal migration. Suppose that these migration costs at time t are given by the following quadratic cost function:⁴⁶

$$(31) \quad C_t = \phi R_t^2,$$

where R_t gives the number of natives who move at time t . It is easy to show that the number of internal migrants is approximated by⁴⁷

$$(32) \quad R_t = \frac{1}{4} \sigma (N + M) \ln \frac{w_{1t-1}}{w_{2t-1}}.$$

The present value of GDP accruing to natives, net of migration costs, is given by

$$(33) \quad Q_N = \sum_{t=0}^{\infty} \left(\frac{1}{(1+r)^t} (AL_{1t}^{\beta} + AL_{2t}^{\beta} - w_{1t} k\theta(1-\theta)(1-2\lambda)N - \phi R_t^2) \right),$$

where r is the rate of discount. Let $Q_N|_{\theta=1}$ be the net income accruing to natives if immigrants are income maximizers and cluster in the high-wage region, and let $Q_N|_{\theta=\lambda}$ be their net income if immigrants replicate the geo-

46. The quadratic cost function implies that it is very expensive for all natives to move at the same time. The increasing marginal cost of migration can be justified if there are limited resources to facilitate the geographic move, and if rising demand for internal migration leads to more congestion and hence higher costs for the marginal internal migrant.

47. Note that $\ln(L_2/L_1) \approx (L_2 - L_1)/0.5(L_1 + L_2)$. Equation 32 follows from applying this approximation to the lagged adjustment function in equation 28.

graphic sorting of the native population. The measure of the efficiency gain from immigration defined earlier is simply the difference between these two quantities.

The magnitude of the efficiency gain will obviously depend on the value of the supply elasticity σ . Rather than assume a particular numerical value, I build into the simulation a behavioral assumption about native migration. Suppose that native migration decisions are efficient, in the sense that native migration maximizes the value of national income accruing to natives, net of migration costs. It is well known that if natives have perfect foresight and the migration response by native workers is efficient, the quadratic cost function leads to the lagged adjustment function in equation 28.⁴⁸ The supply elasticity s is then a sufficient statistic to summarize how native migration responds optimally to changes in the regional wage gap. The assumption of efficient migration greatly simplifies the estimation of the efficiency gain that accrues to natives. The annualized efficiency gain is given by⁴⁹

$$(34) \quad \Delta_N = \left(\max_{\sigma_1} rQ_N \Big|_{\theta=1} - \max_{\sigma_\lambda} rQ_N \Big|_{\theta=\lambda} \right),$$

where σ_1 is the supply elasticity measuring optimal native migration behavior when all of the immigrants choose to live in the high-wage region, and σ_λ is that elasticity when a fraction λ of the immigrants choose to live in the high-wage region. The separate maximization of each stream of national income ensures that the “right” number of natives move in each period under either regime.

Table 8 reports estimates of Δ_N for alternative values of the parameters.⁵⁰ The table also reports the supply elasticities that maximize the efficiency

48. Gould (1968); Hamermesh (1993). The rate of change of employment in the dynamic model of labor demand with quadratic adjustment costs is $dE_t/dt = \kappa(E^* - E_t)$, where E^* gives the steady-state employment and E_t gives employment at time t . The alternative definition of R_t given in equation 32 implies that $R_t - R_{t-1} = -\delta\sigma R_{t-1}$. This relationship is identical to that implied by the dynamic model of labor demand because the steady-state value of R^* is zero.

49. This comparison is related to the discussion in the trade literature that analyzes the optimal timing of trade liberalizations; see Lapan (1976) and Mussa (1986).

50. The algebraic structure of the model does not lead to a closed-form solution for the maximization problem in equation 34. The simulation conducts the maximization by calculating the efficiency gain under each regime for many alternative values of σ , and then choosing the value of the elasticity that maximizes the net gain.

Table 8. Estimates of Efficiency Gains from Immigration Accruing to Natives under Selected Parameter Values

Item	Initial regional wage gap ^a		
	Sizable (0.25)	Moderate (0.12)	Low (0.06)
<i>Parameter</i>			
λ^b	0.30	0.40	0.45
k^c	0.25	0.50	1.00
<i>Low migration cost^d</i>			
Net efficiency gain ^e	6.9	4.7	2.7
σ_1^f	0.4	0.6	0.0
σ_λ^g	0.3	0.3	0.3
<i>Medium migration cost</i>			
Net efficiency gain	9.0	6.1	3.5
σ_1	0.3	0.4	0.0
σ_λ	0.2	0.2	0.2
<i>High migration cost</i>			
Net efficiency gain	12.4	7.9	4.6
σ_1	0.2	0.3	0.0
σ_λ	0.1	0.1	0.1
<i>Prohibitive migration cost</i>			
Net efficiency gain	22.0	13.2	10.3
σ_1	0.0	0.0	0.0
σ_λ	0.0	0.0	0.0

Source: Author's calculations based on the theoretical model.

a. Difference in the log wage index before immigration.

b. The fraction of natives that live in the high-wage region.

c. The size of the immigrant flow relative to the supply imbalance between the two regions.

d. Average cost of migration equals half of annual income. For medium cost, average cost of migration equals annual income, and for high cost it equals twice annual income.

e. Efficiency gains (in billions of dollars) accruing to natives are the difference between the income that accrues to natives if all immigrants settle in the high-wage region and that if immigrants replicate the geographical distribution of the native population.

f. The supply elasticity that maximizes native income (net of migration costs) when immigrants cluster in the high-wage region.

g. The optimal supply elasticity when immigrants replicate the initial native geographical distribution.

gain in equation 34. All of the simulations assume that the real rate of interest is 3 percent, that the parameter β is 0.7 (so that the elasticity δ in the labor demand function is 0.3), that there are 100 million native workers in the economy, and that GDP in the preimmigration regime is \$10 trillion.⁵¹ Finally, the relation between the immigrant share of the work force ($m = M/N$) and the fraction of natives who reside in region 1 is

$$(35) \quad m = k(1 - 2\lambda).$$

51. The assumption that $\delta = 0.3$ is consistent with the evidence summarized in Hamermesh's (1993) survey of the labor demand literature.

All of the simulations impose the restriction that $m = 0.1$, so that immigration increases labor supply by 10 percent. This assumption greatly restricts the range of valid values for k and λ .

Suppose that $\lambda = 0.4$ (or, equivalently, that $k = 0.5$), so that 40 percent of natives live in the high-wage region, and that the observed volume of immigration is half as large as required to bring about an immediate equalization of wages across regions. In this scenario the initial wage gap between the two regions is 12.2 percent. The simulations in table then indicate that the efficiency gain accruing to natives is modest. Consider first the case where natives are immobile, so that the supply elasticity of natives is set to zero, regardless of how many immigrants enter the country and where they settle. The last panel of the table shows that the annual efficiency gain from immigration is \$13.2 billion.

Suppose now that immigration costs are “low,” in the sense that the average cost of a native moving from region 2 to region 1 is half the income per worker in the country (\$35,000).⁵² The efficiency gain accruing to natives falls to \$4.7 billion. The efficiency gain is now relatively small because the low migration costs imply that natives would have quickly moved across regions anyway, so that there is little to be gained from immigration. The efficiency gain accruing to natives increases to \$7.9 billion when the migration costs are “high,” in the sense that the migration cost per worker, on average, is twice as large as income per worker.

The efficiency gains are substantially larger when there is a larger regional wage gap in the preimmigration period, that is when income-maximizing immigrants have the most to contribute to the narrowing of regional wage differences. Consider, for instance, the situation when $\lambda = 0.3$, so that the initial wage gap between the two regions is 25.4 percent. The efficiency gain accruing to natives is now approximately twice that calculated when the preimmigration regional wage gap was only 12.2 percent, ranging from \$6.9 billion to \$12.4 billion annually.

Although these efficiency gains are trivial in the context of a \$10 trillion economy, they are not trivial when compared with the traditional immigration surplus stressed in earlier calculations. After all, the efficiency gain

52. The assumption that $\beta = 0.7$ implies that 70 percent of the \$10 trillion in national income accrues to workers. Since there are 100 million workers, income per worker equals \$70,000. Equation 31 implies that the average cost of migration equals ϕR . I obtain the value of β by evaluating the expression for average costs at the point that would be observed if 1 percent of native workers (or 1 million workers) migrated in any particular year.

measured here accrues to natives *in addition* to the immigration surplus that arises from having more workers in an economy with a fixed (and native-owned) capital stock. The efficiency gain from immigration measures how much a *fixed* volume of immigration increases native income simply because the immigrants tend to cluster in the high-wage region. As I showed earlier, the immigration surplus is approximately \$10 billion when the factor price elasticity is 0.3. It turns out that the estimates of the efficiency gain roughly double the measured benefits from immigration.

The simulations reported in table may be unsatisfactory in one important way. As noted above, many studies in the wage convergence literature report that the half-life of differences in income per capita across U.S. states is around thirty years. It turns out, however, that the assumption of efficient migration built into the simulations implies a faster rate of convergence. The simulations yield an estimate of σ_1 (the supply elasticity when all immigrants cluster in the high-wage region) of 0.3 to 0.4 in the medium-cost scenario. This supply elasticity, in turn, implies that the half-life of the initial wage gap is around five to seven years.⁵³

Suppose that the empirical finding of a thirty-year half-life for regional wage differences is correct. The simplest explanation of the inconsistency between the simulations and this empirical finding is that migration costs are far higher than assumed in the calculations.⁵⁴ In fact, the exercise would generate a half-life of thirty years if I assumed that the average migration cost was around three times income per worker, or \$210,000. Although this estimate of migration costs may seem implausible, it is less so when viewed in the context of the *actual* migration flows across regions or countries. Consider, for example, the migration flow between Puerto Rico and the United States. Although there are no legal restrictions on the movement of Puerto Ricans to the U.S. mainland, a large income gap has persisted between these two regions for many decades.⁵⁵ The Penn World Tables report that in 1955 (chain-indexed) GDP per capita in the United States was four times as large as in Puerto Rico. Even by 1999, GDP per

53. Equations 25 and 30 indicate that the regional wage gap has been cut in half when $(1 - \delta\sigma)^t = 0.5$. The implied half-life is then given by $t^* = \ln 0.5 / \ln (1 - \delta\sigma)$.

54. It is also possible that native migration is inefficient because it fails to maximize native income net of migration costs.

55. Fitzpatrick (1971) gives a comprehensive history of the population flow between Puerto Rico and the United States.

capita, adjusted for purchasing power parity, was still 3.5 times as large as in the United States.

Not surprisingly, many Puerto Ricans have taken advantage of the huge income gap over the decades. Nearly 25 percent of Puerto Ricans migrated to the United States between the 1940s and the 1980s.⁵⁶ However, the interesting question is not why 25 percent of Puerto Ricans chose to migrate, but why 75 percent of Puerto Ricans chose *not* to. The Puerto Rican case, I conjecture, strongly suggests that migration costs could easily swamp income differentials that are perhaps on the order of three- to fourfold.

Of course, smaller income gaps across countries would induce even less migration. The recent experience of the European Union, as it extended membership to such low-wage countries as Spain and Portugal, is instructive. In 1990, for instance, GDP per capita in France was 80 percent greater than that in Greece, 70 percent greater than in Portugal, and 44 percent greater than in Spain.⁵⁷ Yet despite much concern that the accession of Greece, Portugal, and Spain into the European Union might generate substantial migration flows, these flows never materialized.

Consider, for example, what happened to migration to Germany—one of the EU countries with the highest GDP per capita—after the initial expansion. The number of Greek nationals living in Germany grew from 356,000 to 363,000 between 1994 and 1996, the number of Spanish nationals remained at about 132,000; and the number of Portuguese nationals grew from 118,000 to 131,000.⁵⁸ And the number of foreign-born nationals from the acceding countries was roughly stable in the other EU countries. It seems, therefore, that the income differences between the original EU countries and the initial wave of acceding countries, large as they were, were not sufficiently large to generate substantial migration flows. In short, there is strong circumstantial evidence suggesting that migration costs play a crucial role in inhibiting labor flows across regions or countries. This fact, of course, increases the importance of immigration as grease on the labor market of the host country's economy.

In the end, the hypothesis of efficient migration, and its link to migration costs and the actual size of population flows across regions, raises a

56. Ramos (1992).

57. World Bank, *World Development Indicators*, 1998.

58. Fröhlich (1997, p. 6).

number of important questions about the process of native internal migration, questions that remain unexplored in a literature that seems ripe for rediscovery.

Conclusion

Migration costs prevent many native-born workers from moving to those states that offer the best economic opportunities. Immigrant workers, in contrast, form a self-selected sample of persons who have chosen to incur those migration costs. As long as migration costs are mainly fixed costs, newly arrived immigrants in the United States will choose to live in those states that offer them the best economic opportunities. As a result, new immigrants should be clustered in those states that offer them the highest wages, and the location decisions of immigrant workers should be much more responsive to interstate wage differentials than those of natives.

This paper has tested this hypothesis using data drawn from the decennial censuses. The evidence indicates that new immigrants do indeed make different location decisions than native workers and than earlier immigrants. Moreover, these differences are strongly related to interstate wage differences within a particular skill group. In other words, new immigrants who have particular skills to offer are more likely to reside in those states that happen to offer the highest wages for those skills. The endogenous sorting of newly arrived immigrants to high-wage areas, in fact, speeds up the process of regional wage convergence.

The empirical evidence, therefore, suggests that immigrants may play an important—and neglected—role in the U.S. economy: They make up a disproportionately large fraction of the marginal workers whose location decisions arbitrage differences across labor markets. Put differently, immigration improves labor market efficiency. Moreover, it turns out that part of this efficiency gain accrues to natives, suggesting that existing estimates of the benefits from immigration may be ignoring a potentially important source of these benefits. However, a back-of-the-envelope calculation suggests that the efficiency gain is probably below \$10 billion a year.

The analysis also has provocative implications for the interpretation of studies that analyze how immigration affects the labor market opportunities of native workers. Most studies in this literature attempt to estimate the

impact of immigration by comparing the wages of native workers in different geographic areas and relating these geographic differences to measures of immigrant penetration in the local labor markets.

Because immigrants deliberately choose to enter those labor markets that offer the highest wages, it will be very difficult to document that increased immigration lowers the native wage in the penetrated geographic areas. The literature has typically attempted to control for this endogeneity problem by using instrumental variables, where the instrument for the measure of immigrant penetration in the local labor market is typically a variable indicating the number of immigrants who resided in that labor market at some point in the past. The evidence presented in this paper suggests that this is not a valid instrument. Immigrants cluster in those labor markets that offer them the best opportunities for the skills that they bring to the country, and hence the size of the preexisting stock of immigrants will not, in general, be uncorrelated with the wages offered by a particular locality.

The evidence presented in this paper suggests that there is much left to learn about the macroeconomic consequences of immigration. For the most part, the immigration literature in labor economics evaluates various aspects of the microeconomic impact of immigration or measures how immigrants perform in the host country's labor market. The fact that immigration speeds up the process of wage convergence raises a number of new questions, such as the link between immigration and economic growth, that deserve careful empirical investigation.

Finally, most of the studies in the immigration literature emphasize some measure of the cost that immigrants impose on various sectors of the host country's economy—whether they be the wage losses suffered by native workers or the increase in taxes borne by native taxpayers. Remarkably little attention has been paid to the possibility that immigrants impart a variety of benefits to the host country's economy. A great deal of research remains to be done to better understand the source and magnitude of these benefits.

