Volume Title: Immigration, Trade and the Labor Market

Volume Author/Editor: John M. Abowd and Richard B. Freeman, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-00095-8

Volume URL: http://www.nber.org/books/abow91-1

Conference Date: September 11-12, 1987

Publication Date: January 1991

Chapter Title: Immigration and Self-Selection

Chapter Authors: George J. Borjas

Chapter URL: http://www.nber.org/chapters/c6663

Chapter pages in book: (p. 29 - 76)

1 Immigration and Self-Selection

George J. Borjas

The insight that migrants may be systematically different from persons who do not choose to migrate has long played an important role in sociological and historical studies of the immigration phenomenon (see, e.g., the studies contained in Jackson 1969). The selectivity hypothesis has also played a major role in the modern economic literature that analyzes how immigrants do in the U.S. labor market. For example, the early studies of Chiswick (1978) and Carliner (1980) invoke the assumption that immigrants are positively selected from the population of the countries of origin to explain the cross-sectional empirical finding that immigrant earnings (after a short time period) "overtake" the earnings of natives with the same observed socioeconomic characteristics, such as age and education.[1]

My recent work in this area (Borjas 1985, 1987) has addressed two related questions raised by the early studies. Since most of the literature analyzing immigrant earnings focuses on the study of single cross-sectional data sets, my 1985 paper raised the possibility that the overtaking findings could be due to the fact that cross-sectional regressions confound aging and cohort effects.[2] The positive correlation between immigrant earnings and years of residence in the United States observed in the cross section could arise because immigrants "adapt" rapidly to the U.S. labor market or because earlier waves of immigrants differ in substantial ways (labor market productivities, unobserved

George J. Borjas is professor of economics at the University of California, San Diego, and a research associate of the National Bureau of Economic Research.

The author is grateful to Richard Freeman for many helpful discussions of the ideas presented in this paper, to Charles Brown for insightful suggestions and comments, and to Bernt Bratsberg for excellent research assistance. He is also grateful to the National Science Foundation (grants SES-8604973 and SES-8809281) and to the National Institute of Child Health and Human Development (grant R01-HD22344) for financial support.

abilities or skills) from more recent waves. Borjas (1985) adapted well-known techniques (see, e.g., Heckman and Robb 1983) to separately identify aging and cohort effects using the 1970 and 1980 U.S. Censuses. This methodology, which "tracks" synthetic cohorts of immigrants over time, showed that (*a*) immigrant assimilation was not as fast as the cross-sectional studies indicate, (*b*) the more recent immigrant waves performed substantially worse in the labor market than the early postwar waves, and (*c*) there was little likelihood that the most recent immigrant waves would ever earn substantially more than natives of comparable age and education.

An important insight provided by the study of synthetic cohorts is that invoking the assumption of positive selection, though it may be correct for some cohorts of immigrants, may be completely wrong for other cohorts of immigrants. This raises the important question of exactly which factors determine whether immigrants are positively or negatively selected from the population in the countries of origin. Borjas (1987) presents an initial attempt to address this problem and derives a simple economic model of selection on the basis of unobserved characteristics (which, after all, form the focus of much of the literature on immigrant earnings). This model, which will be discussed in detail below, shows that there is no general law stating that immigrants must be positively selected. In fact, under a reasonable set of conditions, it is likely that immigrants are negatively selected (i.e., persons who have below-average earnings and productivities are the most likely to migrate to the United States). My empirical analysis revealed that positive selection was more likely to characterize immigrants from the advanced industrial countries and negative selection was more likely to characterize immigrants from the Third World countries, who form the bulk of migration to the United States in the post-1965 period.

This paper expands my earlier work in a number of significant ways. The theoretical analysis below will argue that, although most of the literature has focused on the role that selection in unobserved characteristics plays in determining immigrant earnings, there is also selection in observed characteristics such as education. The theoretical framework clearly shows that it is completely possible for the most educated persons to migrate to the United States (i.e., positive selection in education) but for these persons to be the least productive in the population of highly educated persons (i.e., negative selection in unobserved characteristics). The analysis below presents a number of propositions that yield insights into the process that determines the selection of immigrants in these separate dimensions of "quality."

The empirical analysis in this paper expands my previous work in two ways. First, it presents a detailed analysis of the U.S. earnings of immigrants by focusing on the roles played by selection in both observed and unobserved characteristics. It will be seen that a number of the theoretical predictions are confirmed by the data. Second, it is clear that potential migrants can choose from a number of potential host countries. The empirical analysis below will

present a systematic study of the selection biases generated by the sorting of migrants among three potential countries of destination: Australia, Canada, and the United States. The evidence indicates that both country-of-origin and country-of-destination characteristics play an important role in determining the performance of immigrants in any labor market.

1.1 Theory of Immigration

1.1.1 The Roy (1951) Model

Migration is assumed to flow from country 0, the country of origin or the "home" country, to country 1, the country of destination or, for concreteness, the United States. This simple framework ignores three potential complications. First, it is likely that persons born in the United States also consider the possibility of migrating to other countries, and perhaps many of them do so. Second, even persons choosing the United States as a country of destination may find that things do not work out (or perhaps work out much better than expected), and some return migration is thereby generated. Third, individuals contemplating migration in a particular country of origin enter the "immigration market" in which a number of other host countries (such as Australia and Canada) compete for the immigrant's human and physical capital. Little is known about the size and composition of the migrant flows from the United States to other countries; hence, these possibilities are ignored in what follows. Much more, however, is known about the size and composition of the flows from any given home country to each of three potential host countries (Australia, Canada, and the United States), and the implications of the simpler two-country model will be applied below to the more general framework where potential migrants not only decide whether to migrate but also choose a country of destination.

Residents of the home country face an earnings (w) distribution given by

$$\ln w_0 = X\delta_0 + \varepsilon_0, \tag{1}$$

where X is a vector of socioeconomic characteristics with value δ_0 in country 0, and the disturbance ε_0 is independent of X and is normally distributed with mean zero and variance σ_0^2.

The earnings distribution facing individuals in the United States is given by

$$\ln w_1 = (1 - M)X\delta_n + MX\delta_1 + \varepsilon_1, \tag{2}$$

where M is a dummy variable indicating whether the individual is foreign born or native. The vector δ_n gives the value that the U.S. labor market attaches to the socioeconomic characteristics X for natives. This valuation may differ because of discrimination or other unobserved factors from the value δ_1 that the labor market attaches to the characteristics brought in by potential migrants. The disturbance ε_1 is again independent of X (and M) and is normally distrib-

uted with mean zero and variance σ_1^2. Finally, the random variables ε_0 and ε_1 have correlation coefficient ρ.

Equations (1) and (2) completely describe the earnings opportunities facing a potential migrant (as well as U.S. natives). Three questions are raised by this simple framework. First, what factors determine the size of the migration flow generated by the income-maximization hypothesis? Second, what types of selection in the unobserved characteristics ε are created by the endogenous migration decision? Third, what types of selection in the observed characteristics X are created by the endogenous migration decision?

The migration decision is determined by the sign of the index function:

$$(3) \qquad I = \ln\left(\frac{w_1}{w_0 + C}\right) \approx [X(\delta_1 - \delta_0) - \pi] + (\varepsilon_1 - \varepsilon_0),$$

where C gives the level of mobility costs, and π gives a "time-equivalent" measure ($\pi = C/w_0$) of the costs of migrating to the United States.

The level of migration costs C is likely to vary among individuals for two reasons. First, there are time costs associated with migration, and these time costs are likely to be higher for persons with higher opportunity costs. Second, there are transportation costs associated with migration. These direct costs include not only the air fare (which is likely to be constant across individuals) but also moving expenses of family and household goods, and it is reasonable to suppose that these expenses may also be a positive function of w_0. These assumptions give little hint as to how the time-equivalent measure of mobility costs, π, varies across individuals. It is instructive to assume first that π is constant across individuals since the main implications of the Roy model are clearest in this special case. The analysis below will show that the treatment of π as a random variable in the population does not substantially alter the analysis and will, in some instances, reinforce the conclusions of the simpler model.

Since migration to the United States occurs when $I > 0$, the emigration rate from the country of origin for persons of given characteristics X is given by

$$(4) \qquad P(X) = \text{pr}\{v > -[X(\delta_1 - \delta_0) - \pi]\} = 1 - \Phi(z),$$

where $v = \varepsilon_1 - \varepsilon_0$, $z = -[X(\delta_1 - \delta_0) - \pi]/\sigma_v$, and Φ is the standard normal distribution function. If the characteristics X have a joint density function given by $f(x)$, then the emigration rate from country 0 is given by

$$(5) \qquad P = \int_{x \in \Omega} P(x) f(x) dx.$$

Equations (4) and (5) summarize the (rather obvious) economic content of the theory of migration proposed by Hicks (1932) and further developed in Sjaastad (1962). In particular, the emigration rate is a negative function of mean income in the home country ($\mu_0 = X\delta_0$), a positive function of mean

income in the United States ($\mu_1 = X\delta_1$), and a negative function of migration costs. Much of the literature on the internal migration of persons in the United States is devoted to testing these theoretical predictions (see the survey in Greenwood 1975).

The immigration literature, on the other hand, has historically focused on explaining not the size of migration flows but their composition or labor market quality. As far back as 1919, for example, Douglas was asking whether the skill composition of immigrant cohorts was constant across successive immigrant waves. The theory of migration contained in equations (1)–(5) has important implications about the selection biases that characterize the pool of migrants in terms of *both* unobserved and observed characteristics. Consider initially the selection mechanism in the unobserved characteristics ε. In particular, consider the conditional expectations $E(\ln w_0 \mid X, I > 0)$ and $E(\ln w_1 \mid X, I > 0)$. Note that these means condition on two dimensions: the observed characteristics X and the decision to migrate. Under the normality assumptions, these conditional means are given by

$$E(\ln w_0 \mid X, I > 0) = X\delta_0 + \frac{\sigma_0\sigma_1}{\sigma_v}\left(\rho - \frac{\sigma_0}{\sigma_1}\right)\lambda, \tag{6}$$

$$E(\ln w_1 \mid X, I > 0) = X\delta_1 + \frac{\sigma_0\sigma_1}{\sigma_v}\left(\frac{\sigma_1}{\sigma_0} - \rho\right)\lambda, \tag{7}$$

where $\lambda = \phi(z)/P(X)$, and ϕ is the density of the standard normal. The variable λ is inversely related to the emigration rate and will be positive as long as some persons find it profitable to remain in the country of origin (i.e., $P[X] < 1$).

Let $Q_0 = E(\varepsilon_0 \mid X, I > 0)$, $Q_1 = E(\varepsilon_1 \mid X, I > 0)$, and $k = \sigma_1/\sigma_0$. The variables Q_0 and Q_1 measure the "quality" (in terms of unobserved characteristics) of the migrant pool. The Roy model identifies three cases of substantive interest.[3]

Positive Selection, $Q_0 > 0$ and $Q_1 > 0$.

This type of selection exists when migrants have above-average earnings in the country of origin (for given characteristics X) and also have U.S. earnings that exceed the earnings of comparable U.S. natives (ignoring the possibility that immigrant earnings may be reduced because of their ethnic or racial background). Inspection of equations (6) and (7) shows that the necessary and sufficient conditions for this type of selection to occur are

$$\rho > \frac{1}{k}, \quad k > 1. \tag{8}$$

If ρ is sufficiently high, and if income is more dispersed in the United States than in the country of origin, immigrants arriving in the United States will be selected from the upper tail of the home country's income distribution and

will outperform comparable natives on arrival in the United States. Intuitively, this occurs because the home country, in a sense, is "taxing" high-ability workers and "insuring" low-ability workers against poor labor market outcomes. Since high-income workers benefit relatively more than low-income workers from migration to the United States (regardless of how much higher mean incomes in the United States may be relative to the country of origin), a brain drain is generated, and the United States, with its greater opportunities, becomes a magnet for persons who are likely to do well in the labor market.

Negative Selection, $Q_0 < 0$ *and* $Q_1 < 0$

This type of selection is defined to exist when the United States draws persons who have below-average incomes in the country of origin and who, holding characteristics constant, do poorly in the U.S. labor market. The necessary and sufficient conditions for negative selection to occur are

$$\rho > k, \quad k < 1. \tag{9}$$

Negative selection also requires that ρ be "sufficiently" positive but that the income distribution in the country of origin be more unequal than that in the United States. Intuitively, negative selection is generated when the United States "taxes" high-income workers relatively more than the country of origin and provides better insurance for low-income workers against poor labor market outcomes. This opportunity set leads to large incentives for low-ability persons to migrate, since they can improve their situation in the United States, and to decreased incentives for high-ability persons to migrate, since income opportunities in the home country are more profitable.

Refugee Sorting, $Q_0 < 0$ *and* $Q_1 > 0$

This kind of selection occurs when the United States draws below-average immigrants (in terms of the country of origin) but migrants have above-average earnings in the U.S. labor market. The necessary and sufficient condition is

$$\rho < \min\left(\frac{1}{k}, k\right). \tag{10}$$

In other words, if ρ is negative or "small," the composition of the migrant pool is likely to resemble a refugee population. For instance, it is likely that ρ is negative for countries that have recently experienced a Communist takeover. After all, the change from a market economy to a Communist system is often accompanied by structural changes in the income distribution and by confiscation of entrepreneurial assets and redistribution to other persons. The Roy model suggests that immigrants from such systems will be in the lower tail of the "revolutionary" income distribution but will outperform the average U.S. native worker.

The basic Roy model thus provides a useful categorization of the factors

that determine the quality or composition (in terms of unobserved characteristics) of the migrant pool. Even at this level, several important implications are generated that give some insight into a number of empirical findings in the literature. For example, many studies have documented the fact that refugee populations perform quite well in the U.S. labor market when compared to native workers of similar socioeconomic characteristics. These empirical results are explained by the income-maximization hypothesis and by the fact that these refugee populations, prior to the political changes that led to a worsening of their economic status, were relatively well off in the country of origin. It is, therefore, unnecessary to resort to the arbitrary distinctions between "economic" and "noneconomic" migrants to explain the refugee experience.

The Roy model also provides an interesting explanation for the empirical finding that the quality of migrants to the United States has declined in the postwar period (where quality is defined by the wage differential between migrants and natives of the same measured skills). Prior to the 1965 amendments to the Immigration and Nationality Act, immigration to the United States was regulated by numerical quotas. The distribution of the fixed number of quotas across countries was based on the ethnic population of the United States in 1920 and thus encouraged migration from (some) Western European countries and strongly discouraged immigration from other continents, particularly Asia. The favored countries have one important characteristic: their income distributions are probably much less dispersed than those of countries in Latin America or Asia. The 1965 amendments abolished the discriminatory restrictions against immigration from non-European countries, established a twenty thousand numerical limit for legal migration from any single country (subject to both hemispheric and worldwide numerical limitations), and led to a substantial increase in the number of migrants from Asia and Latin America. The new flow of migrants thus originates in countries that are much more likely to have greater income inequality than the United States.[4] It would not be surprising, therefore, if the quality of immigrants declined as a result of the 1965 amendments.

The theoretical analysis yields two equations that can guide empirical analysis. These equations are given by

$$Q_1 = g(\mu_0, \mu_1, \pi, \sigma_0, \sigma_1, \rho), \tag{11}$$

$$Q_1 = h(\sigma_0, \sigma_1, \rho)\lambda. \tag{12}$$

Equation (11) gives a "reduced-form" equation, where immigrant quality in the United States (i.e., the wage differential between migrants and natives of equal measured skills) is a function of all the primitive parameters of the model (i.e., the parameters of the two income distributions and migration costs). My earlier paper (Borjas 1987) provides a detailed analysis of the theoretical restrictions implied by the income-maximization hypothesis on the direction of the effects of the various variables in the model. These effects are

usually ambiguous and can be categorized in terms of "composition effects" and "scale effects." In particular, a change in any variable α will create incentives for a different type of person to migrate (the composition effect) and for a different number of persons to migrate (the scale effect).

Equation (12) is a "structural" equation and states that, if knowledge of λ is available, a subset of the parameters of the model *enters multiplicatively* through the h function (see eq. [7]). By holding λ constant, the structural equation essentially nets out the scale effect and leads to more unambiguous predictions of the effect of the exogenous variables on the quality of immigrants. It is important to note that the h function in (12) does *not* depend on mean income levels in the countries of origin and the country of destination or on the level of migration costs since these factors play a role only through the selectivity variable λ.

Three comparative statics results are implied by analysis of the λ-constant structural quality equation.

1. An increase in the variance of the income distribution in the home country leads to a decrease in the quality of migrants in the United States.

2. An increase in the variance of the income distribution in the United States leads to an increase in the quality of migrants in the United States.[5]

3. An increase in the correlation coefficient between earnings in the home country and earnings in the United States increases immigrant quality if there is positive selection and decreases immigrant quality if there is negative selection. The ambiguity arises because, the larger the correlation coefficient, the better the "match" between the two countries. The improvement in the match increases the quality of the immigrant flow if there is positive selection and decreases it if there is negative selection.

1.1.2 Random Mobility Costs

These insights have been derived from the simplest version of the Roy model, which treats mobility costs (defined as a fraction of potential income in the country of origin) as a constant in the population. This assumption may be restrictive, and it is important to ascertain how its relaxation affects the results of the model. Suppose that mobility costs are normally distributed in the population and can be written as

$$\pi = \mu_\pi + \varepsilon_\pi, \tag{13}$$

where μ_π is the mean level of mobility costs in the population, and ε_π is a normally distributed random variable with mean zero and variance σ_π^2. The random variable ε_π may be correlated with ε_0 and ε_1 and the correlation coefficients are given by $\rho_{\pi 0}$ and $\rho_{\pi 1}$, respectively. The conditional expectations of migrant incomes in the home and destination countries are now given by

$$E(\ln w_0 \mid X, I > 0) = X\delta_0 + \frac{\sigma_0 \sigma_1}{\sigma_\nu} \left[\left(\rho - \frac{\sigma_0}{\sigma_1} \right) - \rho_{\pi 0} \frac{\sigma_\pi}{\sigma_1} \right] \lambda, \tag{14}$$

$$(15) \qquad E(\ln w_1 \mid X, I > 0) = X\delta_1 + \frac{\sigma_0\sigma_1}{\sigma_{v'}}\left[\left(\frac{\sigma_1}{\sigma_0} - \rho\right) - \rho_{\pi 1}\frac{\sigma_\pi}{\sigma_0}\right]\lambda,$$

where $v' = \varepsilon_1 - \varepsilon_0 - \varepsilon_\pi$.

Equations (14) and (15) show that the addition of migration costs does not affect any of the substantive results of the simplest version of the Roy model if migration costs are uncorrelated with earnings opportunities. However, if migration costs are correlated with earnings opportunities, the type of selection that is generated may change in either direction. Suppose, for example, that migration costs are positively correlated with earnings opportunities. For instance, high-ability persons may take longer to find appropriate jobs. This positive correlation makes both Q_0 and Q_1 more negative and hence increases the likelihood of negative selection. Conversely, if migration costs (measured in time units) and earnings opportunities are negatively correlated, the likelihood of positive selection is increased.

Two additional points about this more general model are worth stressing. First, the importance of variable migration costs in the analysis will diminish greatly if the variance in migration costs is relatively small compared to the variance in the income distributions. Second, regardless of how important migration costs are, the key result that negative selection is more likely from countries with high levels of income inequality and positive selection is more likely from countries with more equal income distributions is unaffected.

1.1.3 Selection in Observed Characteristics

Equation (4), the probit equation determining the migration rate, contains an additional insight: the migration rate is a function of X through the parameter $(\delta_1 - \delta_0)$. Hence, the migration of persons with larger levels of X is more likely if X has a higher return in the United States than in the country of origin, and the migration of persons with lower levels of X is more likely if the country of origin values the characteristic X more than the United States. A complementary analysis to the Roy model can be derived if it is assumed that the vector X consists of only one variable, say education (s), that this variable is uncorrelated with the disturbances in the earnings functions, and that this variable, too, is normally distributed in the population. The assumption of only one variable in the vector X is irrelevant since the results can be easily generalized to any number of variables. The assumption of normality, though unrealistic for some socioeconomic characteristics, does simplify the mathematics substantially and allows a useful extension of the Roy approach to the study of selection in observed skills and the analysis of the actual wage differential between immigrants and natives (as opposed to the standardized wage differential).

Suppose the earnings functions in the two countries are given by

$$(16) \qquad \ln w_0 = \mu_0 + \delta_0 s + \varepsilon_{0'}$$

$$\ln w_1 = \mu_1 + \delta_1 s + \varepsilon_1, \tag{17}$$

and that the education distribution in the population of the country of origin can be written as

$$s = \mu_s + \varepsilon_s, \tag{18}$$

where ε_s is normally distributed with mean zero and variance σ_s^2.

Assuming that mobility costs are constant, the emigration rate for the population in the country of origin is given by

$$P = \text{pr}\{(\varepsilon_1 - \varepsilon_0) + (\delta_1 - \delta_0)\varepsilon_s > -[(\mu_1 - \mu_0) + (\delta_1 - \delta_0)\mu_s - \pi]\} = 1 - \Phi(z^*), \tag{19}$$

where $t = (\varepsilon_1 - \varepsilon_0) + (\delta_1 - \delta_0)\varepsilon_s$, and $z^* = -[(\mu_1 - \mu_0) + (\delta_1 - \delta_0)\mu_s - \pi]/\sigma_t$.

Two interesting questions can be addressed within this framework. First, consider the conditional expectation of schooling of persons who do migrate. It is easy to show that

$$E(s \mid I > 0) = \mu_s + \frac{\sigma_s^2}{\sigma_t}(\delta_1 - \delta_0)\lambda. \tag{20}$$

Hence, the mean schooling of migrants will be less than or greater than the mean schooling of the population depending on which of the two countries values schooling more. Positive selection in schooling will be observed when $(\delta_1 - \delta_0) > 0$ so that the U.S. labor market attaches a higher value to schooling, while negative selection in schooling will be observed when $(\delta_1 - \delta_0) < 0$ so that highly educated individuals have little incentive to leave the country of origin.

It is important to stress that these selection conditions seem to have little to do with the conditions determining selection in unobserved characteristics. Any permutation of selection mechanisms in unobserved and observed characteristics is theoretically possible. Hence, negative selection in unobserved characteristics (or ability) may be jointly occurring with positive selection in education, or vice versa. Simply because the United States attracts highly educated persons from some countries does not imply that these highly educated persons are the most productive highly educated persons in that particular country of origin.

At a more fundamental level, however, the determinants of the two types of selection are not all that different. The sorting in observed characteristics is guided by international differences in the prices δ_0 and δ_1. In the case of unmeasured skills, the sorting is guided by the variances σ_0^2 and σ_1^2. In a sense, these variances measure the "prices" of unmeasured skills in the respective countries since these abilities are more highly rewarded in countries with higher levels of income inequality. The sorting in all the dimensions of skills,

therefore, is guided by the same basic process: skills flow to whichever country offers the highest price for them.

The actual mean earnings of the migrant pool in each of the two countries are given by

$$(21)\quad E(\ln w_0 \mid I > 0) = \mu_0 + \delta_0\mu_s + \left[\frac{\sigma_s^2}{\sigma_t}(\delta_1 - \delta_0)\delta_0 + \frac{\sigma_0\sigma_1}{\sigma_t}\left(\rho - \frac{\sigma_0}{\sigma_1}\right)\right]\lambda,$$

$$(22)\quad E(\ln w_1 \mid I > 0) = \mu_1 + \delta_1\mu_s + \left[\frac{\sigma_s^2}{\sigma_t}(\delta_1 - \delta_0)\delta_1 + \frac{\sigma_0\sigma_1}{\sigma_t}\left(\frac{\sigma_1}{\sigma_0} - \rho\right)\right]\lambda.$$

Mean earnings of migrants depend on the mean education of migrants, as given by (20), and on the mean level of their unobserved characteristics. Since the two kinds of selections are independent, *nothing* can be said about how the average migrant performs in the host country unless the kinds of selections that occurred in each of the two dimensions of quality are known. Nevertheless, it is of interest to document the net effect of the selection in all the various dimensions of skills on immigrant earnings, and the empirical analysis below presents a detailed study of the unstandardized earnings differential between immigrants and natives in the host country.

Equations (21) and (22) show that generalizations about the quality of immigrants based solely on observed education levels (or other measures of X) are extremely misleading. In addition, it is well known that observed characteristics such as education, age, marital status, health, etc. explain a relatively small fraction of earnings variation across individuals. It is not uncommon, for example, to find that the observed characteristics explain much less than a third of the variance in wage rates or weekly earnings. The selection in unobserved characteristics, therefore, is likely to be much more important empirically than the selection in observed characteristics.

A number of comparative statics results can be generated by analysis of equation (20). Perhaps the most interesting of these results is

$$(23)\quad 0 < \frac{\partial E(s \mid I > 0)}{\partial \mu_s} < 1.$$

That is, a one-year increase in the mean education level of the country of origin will increase the mean education level of persons who actually migrate to the United States, but this increase will be by less than one year. The intuition for this result follows from the fact that an increase in μ_s will change the size of the immigrant flow. Suppose, for concreteness, that $(\delta_1 - \delta_0) > 0$ so that there is positive selection in education. The increase in μ_s makes it worthwhile for more persons to migrate and thus dilutes the mean education level of the population of migrants. Hence, the increase in the conditional expectation is less than the increase in the population mean. An important implication

of this theoretical prediction is that the variance in education levels across immigrants (from different countries) in the United States will be smaller than the variance in education levels of the actual populations across countries in the world. In other words, the population of migrants in the United States is more homogeneous (in terms of education) than the populations of the different sending countries.

In general, equation (20) implies the existence of *observable* quality equations analogous to (11) and (12):

$$Q_1^* = g^*[\mu_0, \mu_1, \pi, \sigma_0, \sigma_1, \rho, \mu_s, \sigma_s, (\delta_1 - \delta_0)], \tag{24}$$

$$Q_1^* = h^*[\sigma_0, \sigma_1, \rho, \mu_s, \sigma_s, (\delta_1 - \delta_0)]\lambda, \tag{25}$$

where Q_1^* gives the mean level of the observed characteristics of immigrants in the United States. The estimation of (24) and (25), of course, is likely to be extremely difficult in practice since they introduce a number of primitive parameters (e.g., $\delta_1 - \delta_0$) that are unobservable and likely to remain so.

1.2 Empirical Framework

Recent empirical research on the earnings of immigrants stresses the importance of disentangling the cohort and aging effects that are confounded by a single cross section of data. In the analysis presented in this paper, two Censuses in the country of destination will be pooled (e.g., the 1970 and 1980 U.S. Censuses), and the following regression model will be estimated:

$$\ln w_{ij} = X_j\delta_i + \alpha_1 y_j + \alpha_2 y_j^2 + \sum_t \beta_t C_t + \gamma_i \pi_j + \varepsilon_{ij}, \tag{26}$$

$$\ln w_{nl} = X_l\delta_n + \gamma_n \pi_l + \varepsilon_{nl}, \tag{27}$$

where w_{ij} is the wage rate of immigrant j, w_{nl} is the wage rate of native person l, X is a vector of socioeconomic characteristics (e.g., education, age, etc.), y is a variable measuring the number of years that the immigrant has resided in the country of destination, C is a vector of dummy variables indicating the year in which migration occurred, and π is a dummy variable set to unity if the observation is drawn from the 1980 Census and zero otherwise.[6] The vector of parameters (α_1, α_2), along with the age coefficients in the vector X, provides a measure of the assimilation effect (i.e., the rate at which the age-earnings profile of migrants is converging to the age-earnings profile of natives), while the vector of parameters β estimates the cohort effects. The period effects are given by γ_i for immigrants and by γ_n for natives.

The model in equations (26) and (27) is underidentified. In particular, some of the right-hand-side variables in the immigrant earnings function are perfectly collinear. Suppose, for example, that the immigrant arrived in calendar year θ so that $C_\theta = 1$. Then

$$y = (T - k - \theta) + \pi k, \tag{28}$$

where T is the calendar year in which the latest cross section is observed, and k is the number of years separating the two cross sections. The variable capturing the period effect, therefore, is a linear combination of the cohort variable and of the years-since-migration variable. Obviously, two cross sections cannot be used to identify three separate effects: period, cohort, and aging effects.

In order to estimate the structural parameters describing the extent of immigrant assimilation and cohort quality change, a restriction must be imposed on the size of the period effect in the migrant population. A reasonable, though unverifiable, assumption is that the period effect experienced by immigrants (γ_i) is identical to the period effect experience by natives (γ_n). In other words, changes in the wage rate due to shifts in aggregate economic conditions affect the immigrant and native wage levels by the same relative magnitude. It is easy to show that this restriction is sufficient to identify all the structural parameters in equations (26) and (27) exactly. This theoretical restriction leaves some amplitude for its empirical implementation since the choice of the native base is essentially arbitrary. The choice of a native base for the various immigrant groups under study will be discussed in detail below.

There are two dimensions of migrant quality that can be calculated from the estimated regressions in (26) and (27): (*a*) the entry wage of immigrants when they arrive into the United States and (*b*) the rate at which this wage changes over time. To simplify the empirical analysis, the two measures will be combined into a single measure of immigrant quality. In particular, let $\bar{w}_i(\theta)$ be the entry wage of an immigrant cohort that arrives in the United States at age twenty in calendar year θ, and let $\bar{w}_n$ be the entry wage of a comparable (in terms of all observable economic variables) native person who enters the labor market at age twenty. Similarly, let g_i be the rate at which the earnings of immigrants grow over their lifetime, and let g_n be the growth rate for natives. Finally, let r be the rate of discount (assumed to be the same for migrants and natives). If persons are infinitely lived, the present values associated with the earnings streams of migrants and natives are given by

$$V_i(\theta) = \int_0^\infty \bar{w}_i(\theta) e^{-(r-g_i)t}\, dt = \bar{w}_i(\theta) / (r - g_i), \tag{29}$$

$$V_n = \int_0^\infty \bar{w}_n e^{-(r-g_n)t}\, dt = \bar{w}_n(r - g_n). \tag{30}$$

The percentage difference in present values between immigrants of cohort θ and natives is defined by

$$(31) \quad \ln[V_i(\theta)/V_n] = [\ln \bar{w}_i(\theta) - \ln \bar{w}_n] - \ln(r - g_i) + \ln(r - g_n),$$

and a first-order approximation (using the assumption that earnings growth rates are small relative to the discount rate) yields

$$(32) \quad \ln [V_i(\theta) / V_n] \approx [\ln \bar{w}_i(\theta) - \ln \bar{w}_n] + \frac{g_i - g_n}{r}.$$

Hence, the percentage difference in the present value of the earnings streams faced by immigrants and natives is an additive function of the wage differential at the time of entry and of the difference in earnings growth rates over the life cycle.[7]

The present value differential in (32) can be easily evaluated from the estimates of equation (26) and (27) if two assumptions are made. First, the rate of discount is assumed to be 5 percent. Clearly, the assumption of any higher discount rate would lead to a worsening of relative immigrant earnings since the latter part of the working life cycle (where immigrants tend to do better) would be more heavily discounted. Second, the growth rates g_i and g_n must be evaluated from the age and years-since-migration coefficients in the earnings functions in (26) and (27). The quadratic specification for age and years since migration in the earnings functions implies that the growth rate is not constant over time. The empirical analysis below will define the growth rate g_i and g_n by

$$(33) \quad \hat{g}_i = (Y_i[\bar{X}, 50, 30, \theta] - Y_i[\bar{X}, 20, 0, \theta])/30,$$

$$(34) \quad \hat{g}_n = (Y_n[\bar{X}, 50] - Y_n[\bar{X}, 20])/30,$$

where $Y_i[X, A, y, \theta]$ is the predicted (ln) earnings for an immigrant with characteristics X, at age A, with y years of residence in the United States, and who migrated in cohort θ. Similarly, $Y_n[X, A]$ gives the predicted earnings for a native with characteristics X at age A. In other words, the average growth rate experienced by immigrants and natives between ages twenty and fifty (evaluated at the mean characteristics of the migrant population, $\bar{X}$) is used for estimation of the growth rate in the present value expressions.

This approach has the useful property that the growth rates (for both immigrants and natives) are a linear function of regression coefficients, and, since the entry wages are given by $Y_i[\bar{X}, 20, 0, \theta]$ for immigrants and $Y_n[\bar{X}, 20]$ for natives, the present value expressions in (33) and (34) are also linear functions of regressions coefficients; hence, a standard error can be easily evaluated.

This approach makes a departure from the tradition in the empirical literature that analyzes immigrant earnings. The entire literature essentially focuses on the estimation of entry wage levels and on the calculation of "overtaking" points (if they exist). This type of analysis is not useful if overtaking points occur rather late in the life cycle (or if they do not occur at all), as some recent evidence suggests. The empirical use of the present value of earnings is much more consistent with the theoretical content of the theory of migration and

deemphasizes the somewhat misleading concept of overtaking points. The analysis of the success of migrant groups in the United States, to borrow from the human capital theory that guided much early research on immigrant earnings, should be based not on the calculation of wage differentials at given ages but on the life-cycle wealth accumulated by migrants and natives. Hence, the present value approach used in the empirical sections of this paper is much more in the tradition of the human capital literature and of the Roy model of immigration developed in the previous section.

1.3 Earnings of Immigrants in the United States

1.3.1 Data and Descriptive Statistics

This section analyzes the relative earnings of immigrants in the U.S. labor market. The data are drawn from the 1970 2/100 U.S. Census (obtained by pooling the 5% SMSA and County Group Sample and the 5% State Sample) and the 1980 5/100 A sample. The complete samples are used in the creation of the immigrant extracts, but random samples are drawn for the native "baseline" populations.[8] The analysis is restricted to men aged 25–64 who satisfied five sample selection rules: (1) the individual was employed in the calendar year prior to the Census; (2) the individual was not self-employed or working without pay; (3) the individual was not in the armed forces (as of the survey week); (4) the individual did not reside in group quarters; and (5) the individual reported annual earnings exceeding $1,000. Throughout this section, the dependent variable is the logarithm of the individual's wage rate in the calendar year prior to the Census. The individual's wage rate is defined as the ratio of annual earnings to annual hours worked. In the 1970 Census, annual hours worked is given by the product of weeks worked in 1969 and hours worked in the Census week, while, in the 1980 Census, annual hours is the product of weeks worked in 1979 and usual hours worked per week in that calendar year.

Forty-one countries were chosen for analysis. These countries were selected on the basis that both the 1970 and the 1980 Censuses contained a substantial number of migrants from that country. In particular, it is necessary to have at least eighty observations of persons born in a particular foreign country in the pooled 2/100 1970 Census to enter the sample of forty-one countries. The countries thus chosen account for over 90 percent of all immigration to the United States between 1951 and 1980. It must be noted, however, that this restriction omits some countries that during the late 1970s became important source countries (e.g., Vietnam). Since two Censuses are required for the complete identification of the parameters of the model presented in section 1.2, however, a systematic analysis of the relative earnings of these migrants will have to await the 1990 Census.

Table 1.1 begins the empirical analysis by presenting the unstandardized

differential between the log wage rate of the various migrant groups and "natives." In these statistics, the native population is defined as the group of U.S.-born white, non-Hispanic, non-Asian men aged 25–64. Perhaps the most striking finding in the table is the fact that migrants from European countries tend to have wage rates that often exceed the wages of white natives, while migrants from Asian or Latin American countries tend to have wage rates that are substantially below those of white natives.

Table 1.1 also presents the relative earnings of the 1965–69 cohort of migrants as of 1970, the relative earnings of the same cohort in 1980, and the relative earnings of the 1975–79 cohort as of 1980. These statistics yield important insights into the process of assimilation (the rate at which the earnings of migrants and natives are converging) and into the extent of productivity differences across successive cohorts. The "tracking" of the 1965–69 cohort across Censuses shows that the relative earnings of this cohort of migrants improved over time for most national groups. At the same time, the comparison of successive immigrant cohorts (i.e., the comparison of the 1965–69 cohort as of 1969 and the 1975–79 cohort as of 1979) shows that for some countries the relative earnings of migrants increased while for other countries the relative earnings of migrants decreased substantially. For example, the most recent migrant from France in 1970 was earning about 8 percent less than natives at the time of entry, while the most recent migrant from France in 1980 was earning about 22 percent more than natives at the time of entry. Conversely, the most recent migrant from India in 1970 earned about 4 percent more than white natives at the time of entry, but the most recent migrant from India in 1980 was earning 21 percent less than white natives at the time of entry.

Table 1.2 continues the descriptive analysis by presenting the mean (completed) education level of four different cohorts of immigrants that arrived in the 1960–80 period. Since the education data available in the Census does not differentiate between education obtained prior to immigration and education obtained in the United States after immigration, the mean education levels for the 1970–74 and 1975–79 cohorts are obtained from the 1980 Census, and the mean education levels for the 1960–64 and 1965–69 cohorts are obtained from the 1970 Census. This use of the available data is designed to minimize the contamination of the education variable by postmigration schooling.

The statistics in table 1.2 are consistent with the well-known secular increase in education levels over time for practically all migrant cohorts. It is worth noting, however, that for some countries the increase in education has been quite small (e.g., Portugal) while for others (e.g., Norway) it has been amazingly large. As the theoretical analysis in section 1.1 shows, these truncated education means can be understood only in terms of the population means of the education distribution in the countries of origin. To provide some insights into the extent of self-selection on the basis of education, table 1.2 also presents mean education levels calculated for the population in the countries of origin. The mean education level for the 1960s is calculated using

Table 1.1 **Unstandardized U.S. Earnings of Immigrants Relative to White Natives**

	1970			1980			
Country	All Immigrants	1965–69 Cohort	Sample Size	All Immigrants	1965–69 Cohort	1975–79 Cohort	Sample Size
Europe:							
Austria	.1969	.2182	380	.2108	.3598	−.1258	746
Czechoslovakia	.1229	.0466	398	.1483	.1141	.0273	872
Denmark	.1208	.1803	141	.2387	.3570	.4241	291
France	.1109	−.0766	317	.1071	.1158	.2237	952
Germany	.1600	.1095	2,399	.1577	.2350	.2646	6,499
Greece	−.1722	−.3704	634	−.1874	−.2556	−.3392	2,328
Hungary	.1304	−.0631	650	.1059	.1027	−.1805	1,356
Ireland	−.0369	−.0260	754	.0688	.0737	−.1421	1,580
Italy	−.0150	−.1707	3,068	−.0124	−.0790	−.1616	7,236
Netherlands	.0643	.1412	430	.1717	.2179	.2824	1,161
Norway	.1653	.2629	243	26.96	.4183	.2444	408
Poland	.0392	−.0952	1,629	.0165	.0207	−.3698	3,278
Portugal	−.1913	−.2406	349	−.2104	−.1949	−.3240	2,213
Romania	.1153	−.1915	259	.0551	.0928	−.2913	614
Soviet Union	.0813	−.1048	907	−.0533	−.0578	−.2856	2,104
Spain	−.1572	−.3480	210	−.0417	−.0184	−.2143	730
Sweden	.1485	.2573	221	.2392	.4570	.1617	335
Switzerland	.2424	.0095	177	.3307	.2121	.4735	397
United Kingdom	.1669	.1902	2,231	.2111	.3188	.1924	5,475
Yugoslavia	.0353	−.1382	646	.0546	−.0191	−.1706	1,967
Asia and Africa:							
China	−.1543	−.3459	880	−.2212	−.1324	−.5372	3,875
Egypt	−.0073	−.2127	136	.0737	.3222	−.2892	696
India	.1667	.0413	363	.1221	.4050	−.2085	3,629
Iran	−.0116	−.3556	121	−.0545	.1375	−.2237	1,027
Israel	.0707	−.1951	141	−.0274	−.0392	−.2483	789
Japan	.0535	.0519	228	.1362	.1492	.2020	1,634
Korea	−.0781	−.2183	142	−.0881	.2409	−.3007	2,013
Philippines	−.1920	−.2389	816	−.0707	.0694	−.3143	4,955
Americas:							
Argentina	.0319	−.1644	218	−.0096	.0086	−.1428	834
Brazil	.0212	−.0993	101	.0485	.1407	.0481	345
Canada	.1072	.1084	3,430	.1258	.1440	.1739	7,083
Colombia	−.1452	−.2337	254	−.2313	−.2027	−.4464	1,760
Cuba	−.2822	−.4461	1,960	−.1828	−.2698	−.5392	6,837
Dominican Republic	−.3576	−.5157	210	−.4768	−.4319	−.6785	1,605
Ecuador	−.2343	−.4511	174	−.2473	−.2858	−.5229	1,097
Guatemala	−.1940	−.5372	82	−.3425	−.2182	−.5977	723
Haiti	−.3041	−.3061	130	−.3726	−.2296	−.6536	1,133
Jamaica	−.1645	−.2462	263	−.2132	−.1245	−.3604	2,061
Mexico	−.4094	−.6021	3,122	−.3975	−.3431	−.6402	24,955
Panama	−.0187	−.1899	101	−.0761	−.1263	−.3663	584
Trinidad and Tobago	−.1561	−.2909	86	−.1488	−.0685	−.4150	782

Table 1.2 **Completed Years of Schooling in Immigrant Cohorts**

	Year of Arrival				Mean Education in Population	
Country	1975–79	1970–74	1965–69	1960–64	1970s	1960s
Europe:						
Austria	14.8	13.9	13.4	12.8	8.7	6.7
Czechoslovakia	15.4	14.5	14.1	12.5	10.2	9.1
Denmark	15.5	13.6	16.1	11.6	11.2	8.5
France	15.6	14.8	14.5	12.8	11.1	7.0
Germany	15.2	14.2	13.3	12.0	10.7	10.1
Greece	11.1	9.9	8.8	10.9	9.2	6.2
Hungary	13.6	13.5	12.3	12.6	10.6	7.2
Ireland	13.8	13.1	12.9	11.3	9.1	8.1
Italy	10.6	8.5	6.8	7.5	9.1	5.6
Netherlands	15.9	15.1	14.1	12.3	10.4	8.8
Norway	15.2	15.6	14.0	11.7	9.9	7.2
Poland	12.7	11.9	10.7	9.5	11.2	7.0
Portugal	6.6	6.7	5.2	5.8	8.2	3.5
Romania	13.7	14.5	11.6	11.9	9.5	5.3
Soviet Union	14.3	13.5	10.5	11.3	11.4	8.1
Spain	13.2	11.3	10.3	9.9	8.0	4.4
Sweden	15.4	15.8	15.5	14.4	10.3	8.7
Switzerland	15.4	15.4	14.5	13.6	8.7	6.7
United Kingdom	15.1	14.7	13.7	13.1	10.8	9.9
Yugoslavia	11.0	10.6	10.7	9.4	9.7	3.5
Asia and Africa:						
China	11.3	12.8	12.8	13.2	8.4	4.3
Egypt	15.9	16.2	15.5	15.1	5.7	4.0
India	16.1	17.6	16.7	17.0	4.9	2.2
Iran	15.2	16.3	15.3	15.5	3.6	1.3
Israel	14.2	13.8	13.5	14.0	9.8	7.0
Japan	15.7	14.7	15.4	15.0	11.2	9.2
Korea	14.0	14.9	15.8	16.5	8.0	5.0
Philippines	14.2	14.9	14.8	13.9	8.2	5.1
Americas:						
Argentina	13.6	12.1	12.0	12.6	8.7	6.3
Brazil	15.4	13.1	12.6	12.8	8.6	2.8
Canada	14.6	13.7	12.9	11.4	10.3	8.5
Colombia	11.9	11.3	10.6	11.5	5.0	2.2
Cuba	11.3	9.9	9.5	11.9	8.3	4.1
Dominican Republic	8.9	9.1	8.4	7.9	6.2	3.6
Ecuador	10.9	11.0	10.4	11.3	6.2	3.4
Guatemala	9.0	9.7	9.9	12.0	2.9	1.5
Haiti	10.2	12.1	12.0	11.2	3.2	1.7
Jamaica	11.3	10.9	10.7	10.6	9.5	4.5
Mexico	6.5	6.8	6.1	6.0	6.1	2.9
Panama	13.1	12.7	12.4	11.1	10.1	5.9
Trinidad and Tobago	11.7	12.0	11.0	14.4	7.9	7.1

enrollment data in the various countries of origin during the 1950s, while the mean education level for the 1970s is calculated using enrollment data in the various countries of origin during the 1960s. The "lagged" construction of the variable giving mean education levels in the country of origin is designed to account for the fact that, in the samples used here, the average person migrated at about age 20. The relevant education distribution, therefore, is given by that of persons enrolled in school a few years earlier.[9]

The means in table 1.2 present a remarkable picture. Even after allowing for the substantial errors involved in calculating the population means for each country of origin, the truncated means are almost always much greater than the population means. For example, the mean of education in Haiti is about three years, but the most recent Haitian immigrants report ten years of education in the 1980 Census. Surprisingly, the two statistics are most similar for Mexico, where both immigrants and the Mexican population have 6–7 years of education. Overall, table 1.2 suggests that immigrants are positively selected on the basis of education. The model presented earlier implies that this result is consistent with the hypothesis that the "rate of return" to education is greater in the United States than in most countries of origin. However, it is also consistent with the hypothesis that migration costs are lower for persons with higher education levels. This conjecture has received intensive study in the internal migration literature (Schwartz 1968).

1.3.2 Basic Regression Results

The regression model in equations (26) and (27) was estimated on each of the 41 countries under analysis using the pooled 1970 and 1980 Census data. As noted earlier, the choice of the native baseline is an important step in the estimation procedure. In this section, the reference group is chosen according to the racial/ethnic background of the population of each country of origin. The estimation uses the white, non-Hispanic, non-Asian sample of native men as the reference group for migrants from Europe, Canada, and the Middle East. The group of Asian natives is the reference group for migrants from all other Asian countries. The group of Mexican natives is the reference group for Mexican migrants, and the group of "other Hispanic" men is the reference group for persons from all other Spanish-speaking countries in the American continent. Finally, the group of black natives is the reference group for migrants from countries with predominantly black populations (i.e., Haiti, Jamaica, and Trinidad and Tobago).

The definition of the reference group in terms of the racial/ethnic background of the immigrant population is a simple way of specifying different period effects for the various immigrant groups. Presumably, the effect of changes in aggregate economic conditions on immigrant earnings is likely to be better approximated by the period effects experienced by populations that closely resemble the immigrant group. It is important to note, however, that, although the baseline populations differ across the 41 countries, the calcula-

tion of the present value differentials defined in equation (32) will *always* be relative to white, non-Hispanic, non-Asian natives (as in table 1.1). In other words, the use of alternative reference groups is simply used to "net out" the period effect in the 1980 Census, and, after controlling for period effects, all comparisons between migrants and natives are conducted with respect to the "white" population.

Initially, the regression model is estimated using a detailed list of demographic controls. These include education, marital status, health status, and metropolitan residence (as well as age and age squared). The calculated present value differentials estimated from the 41 runs of the model are presented in table 1.3 for each of the 6 cohorts identifiable in the Census data. It is worth stressing that these present value differentials measure the differences in earnings among migrants and white natives of equal measured skills and hence are empirical counterparts to the quality measure Q_1 defined in terms of unobserved characteristics.

Table 1.3 shows that there are substantial differences in the "abilities" of migrant groups across the 41 countries of origin. Immigrants from European countries (particularly Western European countries) tend to do quite well relative to white natives of comparable socioeconomic characteristics. Recent immigrants from the United Kingdom, for example, can expected about 10 percent larger earnings over their lifetime than comparable white natives, recent immigrants from France will earn about 8–19 percent more than comparable white natives, and recent immigrants from Sweden will earn about 10–20 percent more than white natives over their lifetime.

On the other hand, immigrants from most Asian and Latin American countries do not perform well in comparison to white natives of equal observable skills. Recent immigrants from Taiwan, for example, will earn about 16–34 percent less over their lifetime than comparable white natives, immigrants from Israel will earn about 20–30 percent less, immigrants from Argentina about 20 percent less, and immigrants from Colombia about 24–38 percent less. An immigrant's birthplace plays an important role in determining the type of selection that characterizes the migrant flow.

In addition, table 1.3 shows that, even within a given country of origin, there are sizable differences in the unobserved quality of immigrants across the various cohorts. The quality of immigrants from some countries has been increasing rapidly, while the quality of immigrants from other countries has been declining rapidly. For instance, the most recent French immigrants have a higher earnings potential than earlier cohorts (particularly those arriving before 1970), while the most recent Polish migrants have much lower earnings potential than migrants of earlier cohorts. Similarly, the most recent Canadian immigrants earn about 8–15 percent more than most of the earlier cohorts, while the most recent Mexican immigrants earn about 9–13 percent less than the earlier Mexican cohorts.[10]

Table 1.3 **Present Value Differentials between Immigrants and Natives**

	Year of Arrival					
Country	1975–79	1970–74	1965–69	1960–64	1950–59	< 1950
Europe:						
Austria	−.0841	.1344	.1945	.0707	−.0004	−.0312
	(−.74)	(1.25)	(2.84)	(1.33)	(−.01)	(−.44)
Czechoslovakia	−.0141	−.0546	−.0036	.0609	.0182	.0596
	(−.14)	(−.73)	(−.10)	(1.01)	(.42)	(.90)
Denmark	.4432	.0623	.1522	−.0010	−.0434	.1105
	(2.76)	(.35)	(1.45)	(−.01)	(−.65)	(.94)
France	.1879	.0829	−.0415	−.1179	−.0626	.0539
	(2.29)	(1.15)	(−.74)	(−2.57)	(−1.56)	(.81)
Germany	.0733	.0638	.0385	.0115	.0150	.1174
	(1.69)	(1.50)	(1.44)	(.60)	(.97)	(4.26)
Greece	−.1060	−.1818	−.1344	−.0402	−.0381	−.1230
	(−2.00)	(−5.08)	(−4.40)	(−1.10)	(−1.39)	(−2.28)
Hungary	.1542	−.1132	−.0128	−.0389	.0380	.1441
	(−1.94)	(−1.81)	(−.26)	(−.86)	(1.45)	(2.55)
Ireland	.1267	.0817	.1758	.0676	−.0252	−.2171
	(1.58)	(1.39)	(3.75)	(2.14)	(−.84)	(−4.82)
Italy	.0498	.0424	.0693	.0839	.0695	.0627
	(1.30)	(1.75)	(3.77)	(5.04)	(5.10)	(2.48)
Netherlands	.2815	−.0917	.0936	.0264	−.0422	−.1736
	(3.66)	(−1.11)	(1.69)	(.70)	(−1.40)	(−2.77)
Norway	.1880	.2468	.1757	.2017	.1437	−.0290
	(1.45)	(1.56)	(1.74)	(2.55)	(2.48)	(−.35)
Poland	−.1926	.0727	.0784	.0387	.0526	.0764
	(−4.11)	(1.95)	(2.65)	(1.66)	(2.44)	(2.31)
Portugal	.0293	.0348	.0785	.0954	.0871	.1746
	(.51)	(.82)	(2.31)	(2.44)	(2.18)	(2.11)
Romania	−.2030	.0911	−.0050	−.0253	.0534	−.0041
	(−2.12)	(1.21)	(−.10)	(−.39)	(1.04)	(−.01)
Soviet Union	−.2641	−.0309	−.0332	−.0456	.0203	.0322
	(−4.42)	(−.55)	(−.57)	(−1.06)	(.68)	(.67)
Spain	.1047	.1287	.0518	−.0022	−.1186	−.1001
	(1.17)	(2.01)	(.96)	(−.01)	(−2.22)	(−.94)
Sweden	.1141	.1621	.2205	.0721	.0001	.0153
	(1.01)	(1.15)	(1.97)	(.78)	(.01)	(.14)
Switzerland	.2395	.1071	.1407	.0967	.0594	.0264
	(2.15)	(.93)	(1.69)	(1.41)	(1.05)	(.26)
United Kingdom	.1052	.0910	.0948	.0449	.0098	−.0432
	(3.11)	(2.88)	(4.19)	(2.32)	(.60)	(−1.47)
Yugoslavia	.0602	.0746	.0625	.1389	.1089	.0237
	(.93)	(1.84)	(1.94)	(3.90)	(4.15)	(.39)

(continued)

Table 1.3 (continued)

Country	Year of Arrival					
	1975–79	1970–74	1965–69	1960–64	1950–59	<1950
Asia and Africa:						
China	−.3662	−.3362	−.2274	−.1842	−.1228	−.0944
	(−7.58)	(−10.26)	(−8.62)	(−6.88)	(−4.20)	(−1.88)
Egypt	−.1597	−.1186	−.0588	−.0980	−.0900	.1511
	(−1.70)	(−1.46)	(−.82)	(−1.34)	(−1.33)	(.84)
India	−.3365	−.1635	−.0497	.0391	.0371	.1138
	(−5.28)	(−2.94)	(−.98)	(.78)	(.73)	(.94)
Iran	.0751	−.0084	.0215	−.0470	−.0207	−.0143
	(.77)	(−.10)	(.32)	(−.71)	(−.32)	(−.10)
Israel	−.3304	−.2346	−.2766	−.1978	.0060	.2476
	(−3.50)	(−3.32)	(−4.18)	(−3.10)	(.10)	(1.72)
Japan	.0741	−.0145	−.0496	−.0401	−.0799	−.1887
	(.95)	(−.22)	(−.81)	(−.71)	(−1.67)	(−1.39)
Korea	−.1840	−.1162	−.0966	−.0738	−.2214	−.0144
	(−2.00)	(−1.51)	(−1.31)	(−1.04)	(−3.55)	(−.01)
Philippines	−.1884	−.0778	−.0689	−.1075	−.1856	−.1108
	(−4.01)	(−2.53)	(−2.75)	(−3.26)	(−5.94)	(−2.15)
Americas:						
Argentina	−.2537	−.2723	−.1908	−.0822	−.0497	.0221
	(−2.97)	(−3.69)	(−3.10)	(−1.47)	(−.83)	(1.44)
Brazil	.0679	−.0944	.0623	.0782	−.0006	−.3063
	(.54)	(−.91)	(.75)	(.88)	−.01)	(−1.59)
Canada	.1440	.0497	.1149	.0681	.0359	−.0065
	(4.07)	(1.51)	(5.27)	(4.24)	(2.45)	(−.26)
Colombia	−.3764	−.2372	−.1562	−.0614	.0920	−.2959
	(−4.99)	(−3.62)	(−2.65)	(−1.11)	(1.69)	(−1.77)
Cuba	−.2711	−.0850	−.1366	−.0687	−.1752	−.0870
	(−5.89)	(−2.92)	(−6.08)	(−3.52)	(−7.87)	(−1.61)
Dominican Republic	−.1566	−.0628	−.0399	−.0338	−.0311	−.0904
	(−5.73)	(−3.52)	(−2.75)	(−2.39)	(−1.94)	(−2.99)
Ecuador	−.2965	−.1742	−.2348	−.0657	−.0810	.0581
	(−3.16)	(−2.08)	(−3.07)	(−.89)	(−1.20)	(.28)
Guatemala	−.3163	−.2695	−.2551	−.2085	−.0959	.3290
	(−2.68)	(−3.17)	(−3.32)	(−2.33)	(−1.13)	(1.85)
Haiti	−.4721	−.2447	−.1227	−.0189	−.1056	−.4107
	(−4.81)	(−2.85)	(−1.56)	(−.22)	(−1.29)	(−2.05)
Jamaica	−.2958	−.1505	−.1078	−.2182	−.0780	−.0451
	(−4.48)	(−3.24)	(−2.72)	(−4.51)	(−1.67)	(−.51)
Mexico	−.1566	−.0628	−.0399	−.0338	−.0311	−.0904
	(−5.73)	(−3.52)	(−2.75)	(−2.39)	(−1.94)	(−2.99)
Panama	−.2717	−.0221	−.1267	−.0972	−.1131	.0544
	(−2.20)	(−.24)	(−1.80)	(−1.47)	(−1.74)	(.42)
Trinidad and Tobacco	−.2433	−.0774	−.0438	−.0002	.0981	−.1023
	(−2.15)	(−.95)	(−.64)	(−.01)	(1.04)	(−.57)

Note: The *t*-ratios are presented in parentheses.

1.3.3 Determinants of Selection in Unobserved Characteristics

The Roy model suggests that the quality differentials documented in table 1.3 can be "explained" by economic and political characteristics of both the various countries of origin and the United States at the time of migration.

Because it is easier to obtain such data for the post-1960 period, and also to maintain comparability with the analysis that will be conducted in the next section across host countries, the empirical study in this section focuses on explaining the variation in quality across the four cohorts that arrived in the post-1960 period. Hence, there are 164 observations (41 countries times 4 cohorts per country) in the data set analyzed here. The aggregate variables used in the analysis are, for the most part, obtained from my earlier study (Borjas 1987) and are described in table 1.4. They include measures of political conditions in the country of origin, mobility costs, and characteristics of the income distribution (the mean and the variance).

The empirical analysis of the differences in the present value differentials between immigrants and natives in the 164-observation data set is presented in table 1.5.[11] The first column of the table presents estimates of the reduced-form equation derived in (11). This regression reveals that a relatively small number of country-specific variables explains a large fraction of the inter- and

Table 1.4 **Definition of Aggregate Variables**

Variable	Definition	Mean (Standard Deviation)
FREE	= 1 if the country had a competitive party system at the time of migration, 0 otherwise	.48 (.50)
COMMUNIST	= 1 if the country had a Communist government at the time of migration, 0 otherwise	.17 (.38)
LOSTFREE	= 1 if the country lost a competitive party system within the last 10 years, 0 otherwise	.08 (.27)
INEQUALITY	= ratio of household income of the top 10 percent of the households to the income of the bottom 20 percent of the households ca. 1970	7.50 (6.08)
UNEMPLOYMENT	Unemployment rate in the United States at the time of migration	6.25 (1.64)
USLAW	= 1 if migration occurred after 1970, 0 otherwise	.5 (.50)
ENGLISH	Fraction of 1975–80 cohort of immigrants who speak English well or very well	.74 (.25)
DISTANCE	Number of air miles (in thousands) between the country's capital and the nearest U.S. gateway (Los Angeles, Miami, and New York)	3.73 (1.98)
ΔGNP	Difference in (ln) GNP per capita between the country of origin and the United States at the time of migration	−1.39 (1.05)

Note: For additional details on the creation of these variables, see Borjas (1987).

Table 1.5 **Determinants of Differences in Unobserved Characteristics**

	Reduced-form Equation		Structural Equation[a]	
Variable	Coefficient	t	Coefficient	t
CONSTANT	−.1574	(−1.61)	−.0537	(−1.70)
FREE	.0410	(1.45)	.0336	(4.31)
COMMUNIST	.0113	(.37)	.0072	(.69)
LOSTFREE	−.0333	(−.93)	−.0106	(−.86)
INEQUALITY	−.0040	(−1.79)	−.0029	(−4.56)
UNEMPLOYMENT	.0334	(1.81)	.0108	(1.70)
USLAW	−.1593	(−2.61)	−.0505	(−2.42)
ENGLISH	.0797	(1.70)	. . .	. . .
DISTANCE	.0003	(.05)	. . .	. . .
ΔGNP	.0495	(4.02)	. . .	. . .
R^2	.402		.382	
$\hat{h}$	. . .	. . .	−.0167	(−5.24)
$\hat{h}$(USLAW = 0)	. . .	. . .	.0085	(.79)
$\hat{h}$(USLAW = 1)	. . .	. . .	−.0419	(−3.79)

[a]All the variables in the structural equation are interacted with λ, the selection variable. For details, see eq. (12).

intracountry variance in the unobserved quality of immigrants. Many of the aggregate characteristics are statistically significant. Consider, for instance, the variable measuring the extent of income inequality in the country of origin. The coefficient of this variable is negative and marginally significant, as predicted by the theory. Similarly, the difference between mean GNP in the country of origin and mean GNP in the United States is positive and significant, indicating the fact that migrants from countries with advanced economies are characterized by larger levels of unobserved abilities or productivities.

It is worth stressing that the measure of income inequality not only is statistically significant but also has a sizable *numerical* effect on the quality of the immigrant flow. This point is best illustrated by considering two countries: the United Kingdom and Mexico. The inequality measure takes on a value of 4.0 for the United Kingdom and of 12.3 for Mexico. The regression coefficient in table 1.5 suggests that, holding all other factors constant, Mexican immigrants earn 3–4 percent less than British immigrants simply because of the selectivity effects of higher levels of income inequality.

Three other variables seem to be quite important in the regression. The first measures the English proficiency of the immigrant pool. Immigrants from countries where English is prevalent do much better in the United States than immigrants from non-English-speaking countries. Second, the unemployment rate in the United States is an important determinant of immigrant quality: the higher the unemployment rate at the time of migration, the better the quality of the migrant pool. This result is consistent with the Roy model if unemployment particularly effects the earnings opportunities of less-skilled workers.

For instance, an increase in the unemployment rate will worsen the opportunities for persons in the lower end of the ability (i.e., income) distribution and hence will lead to reduced incentives for these persons to migrate. The quality of the self-selected immigrant pool increases as a result of the withdrawal of these persons from the immigration market.

Finally, the reduced-form regression in table 1.5 introduces a dummy variable signaling whether the cohort arrived in the post-1970 period. Recall that U.S. immigration policy was changed drastically by the 1965 amendments (which became fully effective in 1968). Hence, post-1968 cohorts, holding constant characteristics of the country of origin, should differ from earlier cohorts. This is precisely what the results in table 1.5 indicate. In particular, post-1970 cohorts have nearly 16 percent lower (relative) earnings over the life cycle than immigrants who arrived prior to the change in U.S. policy. This result provides striking evidence of a significant structural shift that occurred in the unobserved quality of U.S. immigrants in the last two decades.

As noted earlier, since data exist on the emigration rate of immigrants from any given country of origin (i.e., the number of immigrants in a particular cohort as a fraction of the population of the country of origin at the time of migration), the selectivity variable λ can be calculated, and the structural equation in (12) can be estimated. The structural equation is written as $Q_1 = h\lambda$, and the h-function can be approximated by $h = \beta Z$, where Z is the vector of variables proxying for the relevant primitive parameters. Hence, the empirical counterpart to (12) is $Q_1 = \beta(Z\lambda)$. This structural equation is presented in the second column of table 1.5. The selectivity variable directly controls for changes in mobility costs and means of income distributions, and these variables are omitted from the structural regression. Remarkably, the structural equation leads to estimates that are highly significant and very supportive of the Roy model. In particular, the inequality variable becomes negative and very significant, the unemployment variable remains positive and significant, and the dummy variable indexing post-1970 cohorts remains negative and strong.

The estimated regression parameters can be used to calculate $\hat{h} = \hat{\beta}Z$. The estimates of $\hat{h}$ are presented at the bottom of column 2 in table 1.5. Three estimates are presented: one evaluated at the mean of all the variables, a second one evaluated at the same means but letting the dummy variable USLAW index pre-1970 cohorts, and a third evaluated at the same means but letting the dummy variable USLAW index post-1970 cohorts. These simulations show that there seemed to be weak positive selection prior to 1970 but very strong negative selection in the post-1970 period.

1.3.4 Determinants of Selection in Education

As noted earlier, self-selection occurs not only on the basis of unobserved ability but also on the basis of observed characteristics such as education. Table 1.2 documented the strong differences in educational attainment across immigrant groups from different countries. In addition, it was seen that the

observed educational attainment of immigrant groups differed from the mean educational attainment of the population in the country of origin. It is of interest, therefore, to analyze whether the same conceptual framework that explains the differences in unobserved characteristics also explains the differences in educational attainment.

The key implication of the Roy model is that highly educated persons are most likely to originate in countries that have a low rate of return to education (relative to that found in the United States). Put another way, holding the mean educational attainment in the population of the source country constant, there should be a negative relation between the mean educational attainment of immigrant national origin groups and the rate of return to schooling in the source country.

A detailed study by Psacharopoulos (1973) reports the private rate of return to higher education for a number of countries. Unfortunately, there are only 15 countries in common between his sample and the sample of countries that are important sources of immigration. Nevertheless, using the data presented in table 1.2 for each of the four post-1960 cohorts (so that there are 60 observations), the following regressions were estimated:

$$E(s \mid I > 0) = \underset{}{13.02} + \underset{(9.78)}{.23\,\mu_s} \underset{(1.90)}{} \underset{(-2.44)}{-9.76r}, \qquad R^2 = .236,$$

$$E(s \mid I > 0) = \underset{}{8.01} + \underset{(4.81)}{.66\mu_s} \underset{(4.51)}{} \underset{(-2.79)}{-9.82r} \underset{(-4.24)}{-\ 1.44\ \Delta\text{GNP}}, \qquad R^2 = .421$$

where the dependent variable is the mean educational attainment of the immigrant group in the United States, μ_s is the mean educational attainment in the source country, r is the rate of return to education in the source country, ΔGNP is the percentage difference in per capita GNP between the source country and the United States at the time of migration, and the t-statistics are presented in parentheses.

The most important finding in these regressions is the significant negative effect of the rate of return to schooling in the source country on the educational attainment of the immigrant pool. A 10 percentage point increase in the rate of return to schooling decreases the mean educational attainment of the immigrant group by almost one year. The key implication of the Roy model, therefore, is confirmed by the data. The educational composition of the sample of immigrants is determined by the relative payoff to schooling in the source country.

The regressions also indicate that the mean level of educational attainment in the country of origin has a positive effect on the mean educational attainment of immigrants and that the coefficient, as predicted by the theory, is between zero and one. I should add, however, that this confirmation of the theory—like the results regarding the rate of return to schooling—should be treated with some caution because the data on international differences in mean education levels and rates of return are measured with substantial error.

1.3.5 The Unstandardized Wage Differential between Immigrants and Natives

Up to this point, the empirical analysis has focused on ascertaining the types of selection that occur on the basis of unmeasured skills and on the basis of educational attainment. As noted in the theoretical section, selection occurs along every single dimension of skills that is valued by the labor market, with particular skills flowing to countries that value those skills the most. As a result of these selection processes, the actual earnings of the typical immigrant in the United States are likely to differ substantially from the earnings of the typical white native.

To determine the effect of self-selection in the various dimensions of skills on the earnings gap between immigrants and natives, the 41 regression models were reestimated without including any demographic characteristics in the vector X (except, of course, age and age squared so as to trace the age-earnings profile over the working life). The resulting present value differentials are presented in table 1.6. As can be seen by comparing tables 1.3 and 1.6, the relative earnings of some immigrant groups are lowered significantly when no standardization for demographic variables is conducted. For instance, the most recent immigrant cohort from Mexico earns about 47.8 percent less than white natives, but they earn only about 15.7 percent less than demographically comparable white natives. On the other hand, the most recent cohort of immigrants from the United Kingdom earns about 18.8 percent more than white natives, but only about 10.5 percent more than demographically comparable white natives. In other words, some immigrant groups have demographic characteristics that are much less valuable than those of natives, while other immigrant groups have demographic characteristics that are much more valuable than those of the typical native.

It turns out that the same country-specific variables that explain the variation in standardized earnings differentials among source countries also explain the variation in the unstandardized differentials. Table 1.7 analyzes the determinants of the intercountry variation in unstandardized immigrant earnings (analogous to the second-stage regressions on standardized earnings presented in table 1.5). It is worth noting that, as before, immigrants originating in countries with high levels of income inequality have lower earnings than immigrants originating in other countries. This is not surprising since the level of income inequality can be interpreted as a summary index of skill prices. Thus, the key prediction of the Roy model is confirmed by the analysis of the actual earnings (as opposed to the standardized earnings) of immigrants.

1.4 Immigrant Sorting among Host Countries

The last section showed that the labor market performance of immigrants currently living in the United States is strongly influenced by the economic and political characteristics of the country of origin at the time of migration.

Table 1.6 **Unstandardized Present Value Differentials between Immigrants and Natives**

	Year of Arrival					
Country	1975–79	1970–74	1965–69	1960–64	1950–59	<1950
Europe:						
Austria	.0142	.1987	.3104	.1157	.0535	.1135
	(.10)	(1.72)	(4.20)	(2.03)	(1.27)	(1.49)
Czechoslovakia	.0561	.0045	.0805	.1289	.1191	.1745
	(.50)	(.03)	(1.54)	(1.99)	(2.58)	(2.46)
Denmark	.6004	.0933	.3028	.0409	.0312	.1331
	(3.54)	(.49)	(2.73)	(.50)	(.44)	(1.07)
France	.2658	.1464	.0173	−.0847	.0116	.1896
	(3.05)	(1.89)	(.28)	(−1.73)	(.26)	(2.68)
Germany	.1673	.1054	.0937	.0359	.0526	.2464
	(3.64)	(2.32)	(3.27)	(1.78)	(3.29)	(8.44)
Greece	−.1990	−.2996	−.2326	−.0534	−.0370	−.0627
	(−3.52)	(−7.83)	(−7.13)	(−1.37)	(−1.26)	(−1.09)
Hungary	−.1137	−.0799	.0442	.0412	.0925	.3162
	(−1.34)	(−1.19)	(.82)	(.85)	(3.36)	(5.20)
Ireland	.0148	.0306	.1284	.0676	.0299	−.1139
	(.17)	(.49)	(2.57)	(2.00)	(.94)	(−2.40)
Italy	−.0645	−.0947	−.0526	−.0132	.0007	.0344
	(−1.61)	(−3.72)	(−2.73)	(−.76)	(.03)	(1.29)
Netherlands	.4170	.0036	.2121	.1008	.0295	−.0980
	(5.17)	(.03)	(3.57)	(2.50)	(.90)	(−1.48)
Norway	.3004	.4233	.3319	.2751	.2018	.0117
	(2.20)	(2.50)	(3.12)	(3.24)	(3.26)	(.14)
Poland	−.2091	.0479	.0736	.0186	.0563	.1322
	(−4.22)	(1.21)	(2.33)	(.75)	(2.45)	(3.75)
Portugal	−.2826	−.2518	−.1962	−.1538	−.1270	.0159
	(−4.65)	(−5.66)	(−5.46)	(−3.70)	(−3.03)	(.17)
Romania	−.1151	.2340	.1141	.0808	.0724	.0895
	(−1.13)	(2.94)	(1.61)	(1.15)	(1.35)	(.83)
Soviet Union	−.2086	−.0061	−.0293	−.0118	.0429	.1402
	(−3.29)	(−.10)	(−.47)	(−.26)	(1.36)	(2.74)
Spain	.1182	.0601	.0221	−.0084	−.1335	−.0793
	(1.24)	(.87)	(.39)	(−.14)	(−2.35)	(−.69)
Sweden	.1826	.2013	.3170	.1426	.0867	.1199
	(1.20)	(1.36)	(2.64)	(1.44)	(1.18)	(.98)
Switzerland	.2868	.1716	.2385	.1904	.1589	.2197
	(2.42)	(1.39)	(2.67)	(2.60)	(2.66)	(1.99)
United Kingdom	.1875	.1705	.1898	.1198	.0905	.0490
	(5.22)	(5.05)	(7.88)	(5.80)	(5.20)	(1.58)
Yugoslavia	−.0115	.0043	.0123	.1105	.0938	.0549
	(−.17)	(.10)	(.36)	(2.90)	(3.36)	(.83)

Table 1.6 (continued)

Country	Year of Arrival					
	1975–79	1970–74	1965–69	1960–64	1950–59	<1950
Asia and Africa:						
China	−.3703	−.2659	−.1115	−.0379	.0157	−.0562
	(−7.24)	(−7.59)	(−3.97)	(−1.33)	(.51)	(−1.05)
Egypt	.0101	.0583	.1559	.0914	.1147	.3073
	(.10)	(.66)	(2.03)	(1.17)	(1.59)	(1.59)
India	−.0705	.1339	.2617	.3240	.2784	.2143
	(−1.04)	(2.28)	(4.87)	(6.11)	(5.24)	(1.68)
Iran	.1011	.1077	.1501	.1339	.2055	.2627
	(.94)	(1.39)	(2.09)	(1.89)	(2.90)	(1.68)
Israel	−.2692	−.1891	−.1761	−.0764	.1744	.3943
	(−2.66)	(−2.50)	(−2.49)	(−1.12)	(2.99)	(2.56)
Japan	.1857	.0683	.0565	.0927	.0500	−.0139
	(2.24)	(1.02)	(.87)	(1.55)	(1.00)	(−.10)
Korea	−.0846	.0047	.0826	.1141	−.0169	.3015
	(−.87)	(.03)	(1.06)	(1.52)	(−.26)	(1.30)
Philippines	−.1415	.0236	.0245	−.0109	−.1089	−.1758
	(−2.84)	(.73)	(.92)	(−.32)	(−3.31)	(−3.31)
Americas:						
Argentina	−.2239	−.3123	−.1801	−.0122	.0736	.4027
	(−2.47)	(−3.98)	(−2.76)	(−.20)	(1.17)	(2.47)
Brazil	.1575	−.0816	.0717	.0807	.0615	−.2328
	(1.19)	(−.74)	(.82)	(.87)	(.74)	(−1.14)
Canada	.1862	.0593	.1069	.0654	.0685	.0537
	(4.95)	(1.69)	(4.58)	(3.80)	(4.37)	(2.06)
Colombia	−.4451	−.3141	−.1903	−.0522	.1791	−.1150
	(−5.59)	(−4.53)	(−3.04)	(−.89)	(3.10)	(−.65)
Cuba	−.3450	−.1693	−.1877	−.0234	−.1658	−.0256
	(−7.13)	(−5.52)	(−7.93)	(−1.15)	(−7.08)	(−.45)
Dominican Republic	−.6794	−.4560	−.3963	−.3034	−.1546	−.0256
	(−8.37)	(−6.89)	(−6.81)	(−5.44)	(−2.37)	(−.33)
Ecuador	−.4016	−.2428	−.2789	−.0339	.0186	.1277
	(−4.06)	(−2.75)	(−3.44)	(−.44)	(.26)	(.58)
Guatemala	−.4927	−.4041	−.3604	−.2532	−.1017	.4213
	(−3.94)	(−4.48)	(−4.41)	(−2.68)	(−1.14)	(2.23)
Haiti	−.5907	−.3037	−.1307	.0069	−.0527	−.2614
	(−5.64)	(−3.31)	(−1.55)	(.10)	(−.61)	(−1.22)
Jamaica	−.3938	−.2388	−.1406	−.2124	−.0471	.0514
	(−5.62)	(−4.85)	(−3.35)	(−4.12)	(−.94)	(.55)
Mexico	−.4780	−.3518	−.3074	−.2798	−.2350	−.2918
	(−17.14)	(−19.94)	(−21.97)	(−20.59)	(−15.06)	(−9.42)
Panama	−.3263	−.0653	−.1376	−.0905	−.0433	.2219
	(−2.49)	(−.71)	(−1.85)	(−1.29)	(−.63)	(1.64)
Trinidad and Tobago	−.3313	−.1167	−.0435	.0693	.1741	.0340
	(−2.73)	(−1.35)	(−.60)	(.75)	(1.73)	(.17)

Note: The *t*-ratios are presented in parentheses.

Table 1.7 **Determinants of Differences in Unstandardized Earnings Differentials**

	Reduced-form Equation		Structural Equation[a]	
Variable	Coefficient	t	Coefficient	t
CONSTANT	−.4069	(−4.39)	−.0505	(−1.24)
FREE	.0372	(1.39)	.0692	(7.50)
COMMUNIST	.1056	(3.71)	.0344	(2.73)
LOSTFREE	−.0383	(−1.34)	−.0047	(−.34)
INEQUALITY	−.0061	(−2.30)	−.0041	(−4.00)
UNEMPLOYMENT	.0188	(1.04)	.0054	(.66)
USLAW	−.1322	(−2.13)	−.0354	(−1.27)
ENGLISH	.3917	(9.79)	. . .	. . .
DISTANCE	.0145	(3.38)	. . .	. . .
ΔGNP	.0133	(1.06)	. . .	. . .
R^2	.754		.552	
h	. . .	. . .	−.0248	(−6.42)
h(USLAW = 0)	. . .	. . .	−.0071	(−.52)
h(USLAW = 1)	. . .	. . .	−.0425	(−2.77)

[a]All the variables in the structural equation are interacted with λ, the selection variable. For details, see eq. (12).

Potential emigrants in the source countries, however, *chose* to come to the United States instead of migrating to other potential countries of destination. In a sense, the observed pool of immigrants in the United States is the outcome of competition in the immigration market among various countries of destination. Different countries, by offering different immigration policies and different income distributions, will attract different kinds of immigrants.

Three countries, Australia, Canada, and the United States, have been the main countries of destination for permanent migrants in recent years. U.N. statistics, for example, report that, in the period 1975–80, nearly five million persons migrated to a different country, with nearly two-thirds of these individuals migrating to one of these three countries.[12]

Each of these countries, of course, is characterized by a long history of immigration. The size of the recent flows generated by the self-selection of immigrants into each of the three potential countries of destination is illustrated in table 1.8. Over the period 1959–81, about 14.7 million persons legally left the various countries of origin and migrated to either Australia, Canada, or the United States. Sixty percent of these migrants chose the United States as their destination, and the remainder were evenly split between Australia and Canada. Table 1.8 also shows that these statistics vary significantly between the early part of the period (1959–70) and the later (1971–81). Recent migrants are disproportionately more likely to select the United States as their destination (nearly two-thirds of migrants in the 1970s chose to do so) and disproportionately less likely to choose Australia as their destination (only 14 percent did so).

Migration Flows to the United States, Canada, and Australia

	Period of Migration										
	1959–70				1971–81				1959–81		
	Number (1,000s)	% to U.S.	% to Canada	% to Australia	Number (1,000s)	% to U.S.	% to Canada	% to Australia	Number (1,000s)	% to U.S.	% to Canad
	115.1	37.5	29.6	32.8	220.5	48.3	32.4	19.3	335.5	44.6	31.5
	2,111.6	84.9	13.4	1.7	2,687.7	81.0	15.9	3.1	4,799.3	82.7	14.8
	708.3	69.5	19.2	11.3	2,580.8	73.5	17.7	8.7	3,289.0	72.7	18.0
dom	1,322.9	20.3	28.8	50.9	751.1	18.4	31.7	49.9	2,074.0	19.6	29.8
uding U.K.)	2,583.4	47.5	28.9	23.6	1,309.2	55.7	26.0	18.3	3,892.6	50.3	27.9
	123.7	18.9	32.5	48.6	176.9	23.5	19.4	57.2	300.5	21.6	24.8
	6,965.0	55.2	23.3	21.5	7,726.2	65.9	20.3	13.8	14,690.9	60.8	21.7

S. Department of Commerce, *Statistical Abstract of the United States*; U.S. Immigration and Naturalization Service, *Statistical Ye and Naturalization Service*; Canada Statistics, *Historical Statistics of Canada* and *Canada Yearbook*; Australian Department of Im s, *Australian Immigration* (1982).

These aggregate statistics mask important country-of-origin differences. During the period 1971–81, the United States was less likely to attract immigrants from Africa, the United Kingdom, and Europe and significantly more likely to attract immigrants from Asia or North and South America. Canada, on the other hand, seemed a relatively attractive destination for immigrants from Africa, the United Kingdom, and Europe, while Australia was the destination of choice for persons emigrating from the United Kingdom: nearly half the two million persons who left the United Kingdom in the period 1959–81 migrated to Australia.[13]

1.4.1 Migration Policies in Host Countries[14]

One important constraint on the size and the composition of the flow of migrants to potential host countries is the set of statutes and policies used by the various countries to screen the applicant pool. U.S. immigration policy, prior to the 1965 amendments to the Immigration and Nationality Act, was guided by the objective of restricting the migration of persons whose national origin did not resemble the national origin and ethnic composition of the United States population in 1920. The 1965 amendments abolished the "discriminatory" national origin quota system and instituted the goal of family reunification as the main objective of U.S. immigrant policy. These changes, as we saw above, may have been responsible for a very large decline in the unobserved skills of immigrants admitted by the United States.

Canadian immigration policy, until 1961, also had a preferential treatment of immigrants originating in Western European countries. The 1962 Immigration Act (and further relatively minor changes in the statutes and regulations through the 1970s) removed the country-of-origin and racial restrictions and shifted emphasis toward skill requirements. Under the new regulations, potential migrants who were not relatives of Canadian citizens or residents could enter Canada if they passed a "test." Applicants were graded and given up to 100 points according to a "point system," and 50 points were necessary to obtain permission to migrate to Canada. These points were given according to the applicant's education (a point per year of schooling, up to 20 points), occupational demand (10 points if the applicant's occupation was in strong demand in Canada), age (up to 10 points for applicants under the age of thirty-five, minus 1 point for each year over age thirty-five), a "personal assessment" by the immigration officer that was valued up to 15 points, etc. In 1976, the Canadian Immigration Act was amended to incorporate the goal of family reunification as an important policy objective.

Australian immigration policy has a long history of restricting the migration of persons who are not of British origin. These restrictions, known as the "White Australia Policy," operated both in terms of denying entry to persons of non-British or non–Northern European origin and in terms of denying financial assistance (to cover transportation and resettlement expenses) to undesirable migrants.

World War II raised doubts among Australian officials about the feasibility of defending a large continent with a small population, and a series of governments pursued a national policy of substantially increasing the number of immigrants who chose Australia as their destination. This objective, however, could not be achieved by allowing only British citizens to immigrate, and thus Australia began looking elsewhere for migrants (e.g., Germany, the Netherlands, Malta, Italy, and Greece all signed formal arrangements with the Australian government to recruit and assist persons from these countries in their migration to Australia). Further political changes in Australia led to the abolishment of the White Australia Policy in 1972. An immigration policy devoid of discrimination by national origin and race was announced, and a point system based on the Canadian system was instituted. During the early 1980s, Australia began to stress the concept of family reunification in its migration policy (see Birrell 1984). It is unlikely, however, that this shift in policy will have much effect on the 1981 Australian Census data that will be analyzed below.

The effect of these changes in immigration policy on the national origin composition of the immigrant pool in each of the countries is documented in table 1.9. In all host countries, the national origin of the immigrant population has changed drastically over time. For example, in both Canada and the United States, the share of migrants originating in European countries declined drastically between the 1960s and the 1970s. During the 1960s, 23.5 percent of immigrants to Canada originated in the United Kingdom, and an additional 46.0 percent originated in other European countries. During the 1970s, the fractions had fallen to 15.2 and 21.7 percent, respectively. Conversely, the fraction of immigrants originating in Asia was only 8.4 percent during the 1960s, and this fraction had increased to 29.1 percent during the 1970s.

Table 1.9 shows that the United Kingdom accounted for nearly half the migrants to Australia during the 1960s but for only a third of the migrants during the 1970s. A similar decline is observed in the fraction of Australian immigrants originating in other European countries: from 40.8 to 22.4 percent. On the other hand, the fraction of immigrants from Asia increased from 5.3 to 21.1 percent, a fourfold increase in a ten-year period.

1.4.2 Data and Descriptive Statistics

The data are drawn from Public Use Samples of the Censuses available for each of the three destination countries. The U.S. data are identical to that used in the previous section and require no further description.

The Canadian Censuses were conducted in 1971 and 1981. Both these Censuses have the important characteristic that they report the year in which foreign-born persons arrived in Canada. Hence, the aging/cohort decomposition described in section 1.2 can be carried out. The 1971 data for both immigrants and natives residing in Canada are a 1/100 random sample of the

Table 1.9 **Migration Flows into the United States, Canada, and Australia, 1959–81**

	Period of Migration																	
	United States						Canada						Australia					
	1959–70		1971–81		1959–81		1959–70		1971–81		1959–81		1959–70		1971–81		1959–81	
Origin	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total	Number (1,000s)	% of Total
Africa	43.2	1.1	106.5	2.0	149.7	1.7	34.1	2.1	71.5	4.6	105.6	3.3	37.8	2.5	42.5	4.0	80.2	3.1
America	1,792.0	46.6	2,175.7	42.7	3,967.7	44.3	283.5	17.5	427.9	27.3	711.4	22.3	36.1	2.4	84.1	7.9	120.2	4.7
Asia	492.2	12.8	1,898.1	37.2	2,390.3	26.7	136.3	8.4	457.3	29.1	593.6	18.6	79.8	5.3	225.4	21.1	305.1	11.9
United Kingdom	268.8	7.0	138.5	2.7	407.3	4.6	381.2	23.5	237.8	15.2	619.0	19.4	672.9	44.9	374.8	35.1	1,047.7	40.9
Europe	1,228.2	31.9	729.5	14.3	1,957.7	21.9	745.4	46.0	340.1	21.7	1,085.5	34.0	609.8	40.8	239.6	22.4	849.4	33.1
Oceania and other	23.4	.6	41.5	.8	64.9	.7	40.2	2.5	34.3	2.2	74.5	2.3	60.1	4.0	101.1	9.5	161.1	6.3
Total	3,847.8		5,089.9		8,937.7		1,620.7		1,568.9		3,189.6		1,496.3		1,067.5		2,563.7	

Sources: U.S. Department of Commerce, *Statistical Abstract of the United States*; U.S. Immigration and Naturalization Service, *Statistical Yearbook of the Immigration and Naturalization Service*; Canada Statistics, *Historical Statistics of Canada* and *Canada Yearbook*; Australian Department of Immigration and Ethnic Affairs, *Australian Immigration* (1982).

Canadian population, while the 1981 micro file is a 2/100 random sample of the Canadian population. All observations that satisfy the sample restriction of being prime-age men (aged 25–64), not self-employed, not residing in group quarters, and whose records report positive annual earnings in the year prior to the Census are used in the analysis.

The Australian data used in this paper are drawn from the 1981 Census of Population and Housing, the only micro Australian Census file available at present. This Census file is a 1/100 random sample of the Australian population, and the entire sample (for both immigrants and natives) that satisfies the sample restrictions listed above is used.

Three important problems are raised by the Australian data. First, only one Census is available; therefore, the aging/cohort decomposition cannot be conducted. The Australian results, therefore, are not directly comparable to those for the other two countries. Nevertheless, a simple solution that allows some rough comparisons will be proposed below. Second, the Australian Census does not report annual earnings but instead reports annual incomes (which include nonsalary receipts). This problem may not be very serious since the analysis focuses on native/immigrant earnings differences, and self-employed persons are omitted from the study. Finally, the Australian Census (unlike the U.S. or Canadian data) does not contain good measures of labor supply. Hence, a wage rate for the year prior to the Census cannot be calculated. The empirical analysis in this section, therefore, will be conducted on the logarithm of annual earnings.

Table 1.10 presents summary statistics (mean earnings and education) as well as sample sizes for the various samples that will be used in the analysis.[15] In addition, table 1.10 decomposes the immigrant population in each of the host countries according to the continent of origin. This decomposition by continent (rather than country) is mandated by the fact that, in both the Australian and the Canadian Censuses, the decomposition by country leads to a very small number of observations for most countries. In addition, the Canadian Censuses identify the country of origin only for a select group of Western European immigrants.

The results for the United States, as expected, show a downward trend in the earnings of immigrants (relative to natives) over the decade. The average immigrant in 1970 earned, on average, about as much as the typical native worker. By 1980, however, immigrant earnings were about 15 percent below the native wage. The Canadian data show little change in the relative earnings of immigrants between 1971 and 1981. In both Censuses, the average immigrant had slightly higher earnings than the typical native worker. The exception seems to be immigrants originating in Latin America: their earnings are about 10 percent lower than those of Canadian natives in 1971 but 19 percent lower in 1981. The Australian Census shows that the typical immigrant in 1981 had roughly the same earnings as the typical native person and that the differential varied somewhat by country of origin.

Table 1.10 **Summary Statistics**

	Country of Destination: United States					
	1970			1980		
Country of Origin	ln (w)	EDUC	N	ln (w)	EDUC	N
Natives	8.99	11.5	28,978	9.61	12.7	15,071
Asia	8.88	13.3	3,495	9.47	14.6	25,288
Africa	8.88	13.9	172	9.40	15.3	2,622
Europe	9.06	10.8	16,922	9.69	12.1	42,734
Latin America	8.67	9.2	7,507	9.23	9.4	48,929
All immigrants	8.95	10.8	32,491	9.46	11.7	134,252
	Canada					
	1971			1981		
Natives	8.82	9.9	28,049	9.79	11.3	61,205
Asia	8.72	13.2	409	9.66	13.6	2,372
Africa	8.86	14.1	119	9.74	14.0	504
Europe	8.86	10.0	6,633	9.86	10.9	12,193
Latin America	8.72	12.0	223	9.60	12.1	1,229
All immigrants	8.86	10.5	8,018	9.81	11.7	17,417
				Australia, 1981		
Natives				9.39	11.6	23,086
Asia				9.34	12.9	1,074
Africa				9.45	13.1	267
Europe				9.34	11.4	7,799
Latin America				9.35	12.1	102
All immigrants				9.36	11.7	9,936

It is instructive to compare the Australian statistics with the relevant numbers for Canada and the United States. For instance, European immigrants in Australia actually have the lowest education levels of any of the migrant groups in Australia and have a wage disadvantage of only 5 percent. In Canada, European immigrants also tend to have slightly lower educational levels but higher earnings than natives, while in the United States European immigrants outperform all other immigrant groups despite the fact that they have lower educational levels than the native population. This comparison (as well as similar comparisons for other regions of origin) reveals the nonrandom sorting of migrants across the various host countries.

An important insight is provided by these statistics: generalizations about the productivity or earnings capacities of ethnic or national groups are misleading since they ignore the self-selectivity that generated the composition of the migrant pool in each of the host countries. In other words, there is no such thing as *the* effect of Asian ethnicity or race on immigrant earnings. The value

attached by the host country's labor market to ethnic/racial characteristics depends greatly on the kinds of selections that generated the particular flow of migrants.

1.4.3 1980–81 Cross-sectional Regressions

Since the aging/cohort decomposition cannot be conducted for the Australian data, it is instructive to begin the empirical analysis by focusing on the 1980–81 cross section. Table 1.11 presents the cross-sectional earnings function estimated separately in the samples of immigrants and natives in each of the three countries of destination. The regressions in the native samples are of interest mainly because they are so similar across the destination countries. The coefficients of age, marital status, and urbanization status all have the expected signs and are of similar magnitudes whether the labor market is in Australia, Canada, or the United States. The only coefficient that seems to be an outlier in the native samples is that of education in Australia, where the coefficient is almost twice as large as that in the United States or Canada.

The regressions in the immigrant samples are interesting because they illustrate the general result that practically all socioeconomic variables have a smaller effect among immigrants than among natives regardless of the country of destination. The earnings of immigrants are much less responsive to socioeconomic characteristics than the earnings of natives in these economies.

The immigrant regressions in table 1.11 also include a vector of variables indicating the time of migration.[16] An important use of these coefficients (and of the socioeconomic variables) is to predict the size of the wage differentials between immigrants and natives for each of the cohorts. These predictions are calculated using the mean socioeconomic characteristics of the immigrant sample in each of the host countries. In addition, these predictions are obtained by holding the age of immigration constant at 20 *for all cohorts*. Hence, the typical immigrant in the 1975–1980 cohort is 23 years old when the prediction is calculated, the typical immigrant in 1970–74 is 28 years old, etc. The predicted age-earnings profile, therefore, incorporates both aging and cohort effects. These profiles are presented in table 1.12.

The U.S. and Canadian profiles resemble the ones usually reported in the literature: the earlier cohorts, either because they are older and have been in the country longer or because there are vintage or cohort effects, do much better in the labor market than more recent cohorts. Table 1.12, however, shows that the Australian experience is very different. The Australian cross-sectional age-earnings profile for immigrants is essentially flat! In fact, it is impossible to find any statistical difference in the relative earnings of immigrants among the cohorts that arrived in Australia after 1950. Their relative earnings hover around 7–8 percent less than natives, and there is no discernible trend over time. This result implies that, if there is any assimilation effect in Australia, the quality of immigrants to Australia must have *increased* during the period 1960–80. Hence, a simple comparison of the cross-sectional

Table 1.11 **1980–81 Cross-sectional regressions**

	Country of Destination					
	United States		Canada		Australia	
Sample	Coeff.	*t*	Coeff.	*t*	Coeff.	*t*
Natives:						
CONSTANT	6.6488	(76.33)	7.0465	(193.01)	6.3522	(104.68)
EDUC	.0587	(33.92)	.0510	(76.26)	.0908	(58.77)
AGE	.0841	(20.17)	.0873	(49.42)	.0886	(32.01)
AGE^2	−.0009	(−18.00)	−.0009	(−45.21)	−.0011	(−34.61)
MAR	.3151	(23.53)	.2973	(51.10)	.2727	(31.31)
HLTH	−.3337	(−15.15)	. . .	. . .	. . .	. . .
URBAN	.1545	(12.07)	.1036	(22.78)	.1605	(16.61)
R^2	.193		1.71		.245	
All immigrants:						
CONSTANT	6.6378	(223.77)	7.3415	(95.72)	6.7307	(66.17)
EDUC	.0497	(133.61)	.0415	(40.97)	.0748	(35.59)
AGE	.0802	(55.39)	.0710	(19.31)	.0779	(16.86)
AGE^2	−.0009	(−51.35)	−.0008	(−18.44)	−.0010	(−18.70)
MAR	.2325	(50.52)	.2190	(18.42)	.2013	(14.16)
HLTH	−.3502	(−34.48)	. . .	. . .	. . .	. . .
URBAN	.0574	(9.43)	−.0016	(−.16)	.1079	(5.41)
1970–74 Wave	.2107	(36.81)	.1609	(9.73)	.0444	(2.11)
1965–69 Wave	.3141	(51.89)	.2816	(18.03)	.0491	(2.36)
1960–64 Wave	.3750	(56.74)	.2825	(15.39)	.0810	(3.68)
1950–59 Wave	.4436	(74.88)	.3679	(25.59)	.0811	(4.18)
Pre–1950 Wave	.4752	(64.63)	.4287	(17.50)	.1159	(4.63)
R^2	.226		.163		1.88	

regressions across the destination countries leads to an important finding about the trends that mark the self-selection of immigrant flows to the host countries over the last two decades.

1.4.4 Present Value Differentials

Since two Censuses are required to identify aging and cohort effects, the analysis of equations (26) and (27) is initially restricted to the U.S. and Canadian Censuses. Within each country of destination, five immigrant samples will be analyzed: the pooled sample and the subsamples of immigrants originating in Africa, Asia, Europe, and Latin America. The regression will contain a vector of demographic characteristics including education, marital status, health status (when available), and metropolitan residence. These regressions are used to calculate the present value differential between immigrants and demographically comparable natives for each of the cohorts. These present value differentials are presented in table 1.13. (The data presented in table 1.13 for Australia will be discussed in detail below.)

Table 1.12 **Earnings Differentials between Immigrants and Comparable Natives in 1980–81 Cross Sections**

Origin and Destination	Immigrant Cohort					
	1975–80	1970–74	1965–69	1960–64	1950–59	< 1950
All immigrants in:						
United States	−.3460	−.1534	−.0676	−.0239	.0177	.0045
	(−14.48)	(−.10.42)	(−6.91)	(−2.58)	(1.79)	(.39)
Canada	−.2271	−.1118	−.0286	−.0571	−.0020	.0558
	(−9.52)	(−6.61)	(−2.35)	(−3.99)	(−.22)	(2.78)
Australia	−.0810	−.0642	−.0814	−.0656	−.0796	−.0342
	(−2.51)	(−2.87)	(−4.98)	(−4.05)	(−6.06)	(−1.82)

Note: The *t*-ratios are presented in parentheses.

Consider initially the pooled sample of immigrants. Table 1.13 documents the systematic decline in the quality of immigrants arriving in the United States over the last two decades. For instance, the typical immigrant arriving in 1960–64 in the United States had only a slight wage disadvantage relative to a comparable native, while the typical immigrant arriving in the United States in 1975–79 has a wage disadvantage of nearly 27 percent over the life cycle as compared to the native baseline. Remarkably, the Canadian Censuses reveal very similar patterns: the 1960–64 migrant to Canada had a 6 percent wage disadvantage over the life cycle (relative to natives), while the disadvantage for the most recent migrants (1975–80) has increased to nearly 23 percent.

The American and Canadian trends are less similar when the analysis is restricted to men from a specific country of origin. For example, among European immigrants, the U.S. Census reveals a substantial decline in quality (from a 4 percent advantage to an 11 percent disadvantage) over the last twenty years, while the Canadian Census reveals a roughly stable wage differential between immigrants and natives over the post-1960 cohorts. Similarly, among Asian immigrants, the Canadian data reveal that the 1960–64 and the 1975–80 cohorts had essentially the same relative standing, while the U.S. data reveal a decline in quality from a 15 to a 27 percent disadvantage. These results, therefore, imply that at least part of the similarity between the United States and Canada at the aggregate level is due to the fact that the national origin composition of the cohorts shifted over time, away from European immigrants (who tend to do quite well in the labor market) to Asian and Latin American immigrants (who do much worse in the labor market).

As noted earlier, the Australian Census is available only for 1981. Since cohort and aging effects cannot be identified, the present value differentials cannot be calculated directly. Recall, however, that the 1981 cross-sectional regressions estimated in the Australian data showed that immigrants in

Table 1.13 **Present Value Differentials between Immigrants and Comparable Natives**

	Year of Arrival					
Group	1975–80	1970–74	1965–69	1960–64	1950–59	< 1950
All immigrants in:						
United States	−.2656 (−18.99)	−.1228 (−12.20)	−.0827 (−10.40)	−.0453 (−6.88)	−.0260 (−4.37)	−.0451 (−4.38)
Canada	−.2297 (−13.25)	−.1306 (−8.57)	−.0449 (−3.75)	−.0632 (−4.63)	−.0344 (−3.57)	.0212 (1.10)
Australia	.0149 (.46)	.0136 (.61)	−.0570 (−3.49)	−.0740 (−4.57)	−.1330 (−10.12)	−.0914 (−4.86)
African immigrants in:						
United States	−.3779 (−5.11)	−.3097 (−6.08)	−.1425 (−3.21)	−.1577 (−3.62)	−.1997 (−4.28)	−.1806 (−1.69)
Canada	−.4092 (−3.00)	−.4555 (−3.23)	−.2690 (−2.03)	−.3297 (−2.55)	−.2595 (−2.65)	.2108 (.61)
Australia	−.1688 (−1.01)	−.2197 (−1.90)	−.1191 (−1.42)	−.1317 (−1.26)	−.3413 (−.88)	−.4481 (−3.26)
Asian immigrants in:						
United States	−.2692 (−11.47)	−.4117 (−8.33)	−.1565 (−10.53)	−.1495 (−9.89)	−.2551 (−17.54)	−.2487 (−9.08)
Canada	−.3930 (−6.88)	−.3658 (−6.56)	−.2534 (−4.86)	−.3651 (−6.38)	−.3868 (−10.19)	.0637 (.54)
Australia	−.0634 (−.84)	.0022 (.04)	−.2348 (−4.75)	−.2367 (−3.63)	−.3817 (−7.42)	.0141 (.20)
European immigrants in:						
United States	−.1068 (−6.06)	−.0167 (−1.25)	.0218 (2.14)	.0436 (5.07)	.0307 (4.44)	.0219 (1.79)
Canada	−.0516 (−2.22)	.0113 (.55)	.0022 (.14)	−.0290 (−1.92)	.0116 (1.04)	.0423 (2.04)
Australia	.0745 (1.68)	.0350 (1.33)	−.0524 (−2.87)	−.0732 (−4.26)	−.1121 (−8.15)	−.0833 (−4.18)
Latin American immigrants in:						
United States	−.2716 (−14.62)	−.1273 (−9.53)	−.1243 (−11.42)	−.0841 (−8.91)	−.1282 (−13.56)	−.1629 (−8.18)
Canada	−.3312 (−3.77)	−.2820 (−3.25)	−.1693 (−2.10)	−.1230 (−1.46)	−.1757 (−3.07)	.1788 (.91)
Australia	.1671 (.61)	−.0677 (−.38)	−.3991 (−2.45)	−.2721 (−1.15)	.0827 (.15)	−.2868 (−.70)

Note: The *t*-ratios are presented in parentheses.

Australia face significantly different age-earnings profiles than their counterparts in the United States and Canada. In particular, in the cross section, there seems to be little relation between the earnings of immigrants in Australia and the length of residence in Australia. If there is *any* assimilation effect in Australia, therefore, this result must imply that the quality of immigrants to Australia has increased over the last two or three decades.

A rough estimate of this increase can be obtained if it is assumed that the unobserved assimilation effect experienced by immigrants in Australia resembles the assimilation effect of similar persons (i.e., persons from the same country of origin) in Canada or the United States. Given this approximation, the assimilation effects can then be subtracted from the Australian cross-sectional coefficients (thus netting out the role played by pure aging in the generation of the cross-sectional results), and the present value differentials can be computed for each of the cohorts. Since there are two sets of assimilation parameters (one for Canada and one for the United States), a number of different approximations can be calculated. In general, these experiments led to similar qualitative findings. In this paper, therefore, the assimilation rate used is the average of the two assimilation rates (i.e., the U.S. and Canada aging effects) experienced by immigrants from the same continent of origin.

Given these assimilation rates and the cross-sectional regressions estimated in the Australian Census for each region of origin, it is a simple matter to calculate the predicted present value differential between the various cohorts of immigrants and comparable natives in Australia. These predictions are also presented in table 1.13. Two substantive results are worth noting. As implied by the flat earnings profiles found in the (pooled) Australian cross section, the quality of immigrants to Australia increased slightly over the last twenty to thirty years. The typical immigrant in 1960–64 could expect a 7 percent wage disadvantage over his life cycle, while the typical immigrant in 1975–80 has no wage disadvantage relative to natives over his life cycle. Second, this increase in immigrant quality can essentially be found in every one of the national origin groups under analysis. For example, the typical European immigrant in the early 1960s had a 7 percent wage disadvantage, while the typical European immigrant in the late 1970s has a 7 percent wage advantage over natives. Similarly, the average Asian immigrant in the early 1960s had 24 percent lower earnings over his life cycle than natives, while the differential is only 6 percent (and insignificant) for the most recent migrants.

The data presented in table 1.13 provide a unique descriptive analysis of an important question. Which host countries are the "winners" and which are the "losers" in the immigration market? This comparison, of course, depends on the assumption that the native base across countries has a similar level of productivity and skills. This assumption makes the relative wage of immigrants across host countries directly comparable as an index of immigrant quality. The assumption that natives among the three host countries are roughly simi-

lar is not empirically verifiable. However, it does not seem unreasonable since all three countries share a common language and culture, have similar political and economic systems, and are at similar stages of economic development.

Given this assumption, the statistics presented in table 1.13 present an interesting story of the extent of self-selection in the generation of the foreign-born population in each of the countries. Consider the trends for the pooled sample. During the 1940s and 1950s, Australia was attracting immigrants who had lower productivities than the immigrants attracted by Canada and the United States. This type of selection, however, was drastically reversed during the 1960s, as both Canada and the United States began to attract persons who did not perform as well in the labor market and Australia began to attract immigrants with relatively high levels of unobserved skills.

1.4.5 Determinants of Immigrant Quality

Consider the following regression model:

$$Q_{ij}(t) = X_i(t)\alpha + X_j(t)\beta + \varepsilon_{ij}(t) \tag{35}$$

where $Q_{ij}(t)$ is the present value differential between immigrants and comparable natives of a cohort migrating from country i to country j at time t, $X_i(t)$ is a vector of variables describing conditions in the country of origin i at time t, and $X_j(t)$ is a vector of variables describing conditions in the country of destination j at time t.

The specification of (35) builds in a very important (and restrictive) assumption. In particular, the relative earnings of a person from country i in country j at time t are independent of events in other periods t' ($t \neq t'$), and, more important, they are also independent of conditions in other countries (particularly, they are independent of conditions in other potential countries of destination). This empirical framework, in a sense, introduces an "independence of irrelevant alternatives" assumption into the study. Although this assumption is not likely to be strictly satisfied, it does simplify the empirical analysis greatly. If the assumption was invalid, for instance, the right-hand side of (35) would have to be expanded to include the characteristics of all other potential countries of destination, and the increase in the number of variables would rapidly drive the number of degrees of freedom to zero.

Table 1.14 presents the estimates of the reduced-form equation in (35). The sample consists of 48 observations (4 continents of origin times 4 post-1960 cohorts times 3 countries of destination). The regression in table 1.14 reveals that a small number of characteristics of the countries of origin and the countries of destination do "explain" a very large fraction of the variance in the unobserved quality of immigrants. The variables in the reduced-form equation, for example, explain over 80 percent of the variance in the quality measures presented in table 1.13. Despite this success, however, it must be noted that, because the countries of origin are defined in terms of continents, the

Table 1.14 **Determinants of Immigrant Quality across Host Countries**

Variable	Coefficient	t
CONSTANT	.1252	(−2.77)
USLAW	−.0511	(−1.79)
UNEMPLOYMENT	.0011	(.18)
INEQUALITY(0)	−.0044	(−1.89)
INEQUALITY(1)	.0431	(4.35)
ΔGNP	.0903	(8.78)
R^2	.801	

Note: Key to additional variables: UNEMPLOYMENT = unemployment rate in the host country at the time of migration; INEQUALITY(0) = average income inequality (as defined in table 1.4) in selected countries from continent of origin in decade of migration; INEQUALITY(1) = inequality measure for destination countries in decade of migration; ΔGNP = difference in (ln)GNP per capita between sending and host countries at time of migration.

two variables measuring country-of-origin characteristics (the relative GNP level and the extent of income inequality) are, in effect, averaged over a large and diverse number of countries.[17] It is unclear what biases are caused by this aggregation, but it is important to remember that the coefficients in table 1.14 are, at best, suggestive of the underlying economic behavior.

Both the GNP of the continent of origin (relative to GNP per capita in the country of destination) and the inequality measure for the continent of origin affect the quality of migrants significantly. Migrants from wealthier regions do better no matter where they go, and migrants from regions with large levels of income inequality do worse than other migrants. Similarly, the inequality measure for the country of destination has a positive and significant effect on relative immigrant earnings, as predicted by the Roy model. Finally, the change in U.S. immigration policy (as measured by USLAW) has a negative and marginally significant effect and thus helps identify the effect of this major change in policy relative to other countries. The change in U.S. immigration policy lowered the earnings of migrants by 5 percent relative to the earnings of migrants who chose other countries of destination.

1.4.6 The Point System and Immigrant Quality

It is somewhat surprising that the cohort quality trends in Canada and the United States are so similar despite the major differences in immigration policies between the two countries. Immigration policies, however, can screen applicants only on the basis of *observed* demographic characteristics such as education, occupation, and age. The results summarized in table 1.13 show that even a stringent point system, such as that used by Canada, was unable to prevent a decline in immigrant skills in that country similar to that experienced in the United States.

On the other hand, however, the point system clearly affects the observable skills of the incoming immigrant flow. A clear way of ascertaining the impor-

tance of this direct effect is to reestimate the regression model on the pooled sample of immigrants without controlling for any demographic characteristics. The unstandardized differentials in the present value of lifetime earnings are presented in table 1.15. It is evident that, during the 1970s, immigrants admitted to the United States were substantially less skilled (relative to U.S. natives) than immigrants admitted to other host countries (relative to the natives of those countries).

It is also evident that the point system generated an immigrant flow into Canada that had relatively favorable socioeconomic characteristics. For instance, even though the most recent immigrants into Canada (the post-1975 arrivals) earned 23 percent less than comparable Canadian natives, they earned only 12.2 percent less than the typical Canadian native. These immigrants, therefore, have more favorable demographic characteristics than Canadian natives. Although the point system was unable to prevent the decline in immigrant skills, it greatly tempered the extent of the drop.

This fact has interesting implications. If a host country decides that it wishes to attract more skilled immigrants, a point system seems to be a very direct and simple way of doing so. At the same time, however, the point system has the major limitation that it cannot screen for unobservables, and these unobservables are major determinants of individual earnings. Hence, as long as economic and political conditions motivate relatively unskilled persons to emigrate, the point system may restrict entry to only those who pass the test, but the immigrant pool will be composed mostly of relatively unskilled persons with the acceptable demographic characteristics.

Table 1.15 **Standardized and Unstandardized Present Value Differentials in Host Countries**

	All Immigrants in:					
	United States		Canada		Australia	
Year of Arrival	Standardized	Unstandardized	Standardized	Unstandardized	Standardized	Unstandardized
1975–79	−.2656	−.3706	−.2297	−.1222	.0149	−.0020
	(−18.99)	(−23.91)	(−13.25)	(−6.65)	(.46)	(−.23)
1970–74	−.1228	−.2443	−.1306	−.0271	.0136	−.0377
	(−12.20)	(−21.94)	(−8.57)	(−1.67)	(.61)	(−.41)
1965–69	−.0827	−.1710	−.0449	.0568	−.0570	−.0976
	(−10.40)	(−19.52)	(−3.75)	(4.47)	(−3.49)	(−2.10)
1960–64	−.0453	−.1086	−.0632	−.0079	−.0740	−.0856
	(−6.88)	(−15.18)	(−4.63)	(.54)	(−4.57)	(−1.96)
1950–59	−.0260	−.0712	−.0344	.0094	−.1330	−.1803
	(−4.37)	(−11.12)	(−3.57)	(.91)	(−10.12)	(−5.56)
< 1950	−.0451	−.0562	.0212	.0286	−.0914	−.1997
	(−4.38)	(−5.05)	(1.10)	(1.40)	(−4.86)	(−7.21)

Note: The *t*-ratios are presented in parentheses.

1.5. Summary

Self-selection plays a dominant role in immigration (as it does in all other forms of turnover). There is selection in the determination of the composition of the persons who leave any given country, in terms of both observable characteristics (such as education) and unobservable characteristics (such as abilities and productivities). In addition, this nonrandom sample is then sorted across various possible host countries in a nonrandom way. Hence, the pool of immigrants in any host country is, in a sense, doubly self-selected: the pool of immigrants in the host country is composed of persons who found it profitable to leave the country of origin *and* who did not find it profitable to go anywhere else.

This paper attempts to use the economic theory of self-selection as a guide to understanding how immigrants perform in the labor market. The assumption of wealth-maximizing behavior provides important insights into the mechanics that guide the selection process.

The empirical analysis studied the role played by self-selection in the earnings of immigrants in the United States, and compared these migrants to the pool of migrants who chose to reside in other countries (Australia or Canada). The study of the various Censuses revealed that the United States, as a result of major changes in immigration policy, began to attract relatively less skilled persons in the 1970s. In a sense, the United States became less competitive in the international marketplace that determines the migration decision and the sorting of migrants across host countries.

Notes

1. A recent survey of this literature is given in Greenwood and McDowell (1986).

2. Jasso and Rosenzweig (1985) also stress this important technical point in their work.

3. A fourth case where $Q_0 > 0$ and $Q_1 < 0$ is theoretically impossible since it requires $\rho > 1$.

4. Data on international differences in income inequality are published by the World Bank (1986). These data, however, do not correspond directly to the variances that lie at the heart of the Roy model. In particular, σ_0^2 and σ_1^2 measure the dispersion in "opportunities" (for given X) rather than the variance in incomes across households in a given country.

5. There is a slight technical problem that must be taken into account in the derivation of this result. An increase in σ_1 "stretches" the income distribution of the United States and will lead to a different mean wage level in the pool of migrants even if this pool is restricted to include the same persons. A simple solution to this problem is to define quality in terms of "standardized units," or Q_1/σ_1. The prediction in the text can then be easily derived.

6. Because the year of immigration is not precisely available in the U.S. Census, it is approximated as the midpoint of the available intervals. In the 1980 Census, y takes

on a value of 2.5 if the immigrant arrived in 1975–79, 8 if arrived in 1970–74, 13 if arrived in 1965–69, 18 if arrived in 1960–64, 25.5 if arrived in 1950–59, and 45 if arrived prior to 1950. In the 1970 Census, y is 2.5 if arrived in 1965–69, 8 if arrived in 1960–64, 15.3 if arrived in 1950–59, and 39.8 if arrived before 1950. The estimated y for these last two intervals are calculated using the more precise year of migration data available in the 1970 Census.

7. There is an implicit assumption in (25) that is directly responsible for this simple framework. In particular, growth rates for immigrants are independent of the year of migration θ. The model can be generalized to allow for these types of interactions. However, the estimating equations would include higher order polynomials, and the estimation of the underlying structural parameter may become quite sensitive to the very high correlation among the right-hand-side variables.

8. The construction of the data sets is described in detail in Borjas (1987).

9. The enrollment data are available in Unesco (1969, 1980). Enrollments are available for each "level" of education. The data sources also give the number of years of education associated with that "level" for each country. The means presented in table 1.2 are calculated using both these statistics.

10. It is important to note that many of these differences in quality across cohorts from a given country of origin are statistically significant at conventional levels. For some evidence on this point, see Borjas (1987).

11. Since the dependent variable in the "second-stage" regressions is a linear function of regression coefficients, the regressions are weighted to account for heteroskedasticity. For details, see Borjas (1987).

12. These statistics are available in the United Nations (1982, 44). The calculations ignore the large (and presumably) temporary flows from Ethiopia to Somalia in the late 1970s as well as the movement of guest workers to oil-producing countries in the Middle East.

13. A number of previous studies (e.g., Tandon 1978; Chiswick and Miller 1985; and Chiswick 1988) analyze the labor market performance of immigrants in Australia and Canada. These studies, however, do not study the nonrandom sorting of migrants across host countries.

14. This section is based on the excellent descriptions and summaries of immigration policies given by Boyd (1976), Keely (1979), and Keely and Elwell (1981), Kubat (1979), and Price (1979).

15. Throughout this section, the native base in each of the host countries is the *entire* population of native persons (regardless of ethnic or racial origin). This differs from the native baselines chosen in the previous section but makes the comparisons among host countries less arbitrary.

16. There are some differences in the calendar years bracketed by these dummy variables across the countries of destination. The brackets reported in the table are those that apply to U.S. data. The Canadian and Australian brackets are quite similar for post-1960 migrants but differ for pre-1960 migrants.

17. The average was calculated over the two or three source countries that formed the bulk of immigration from that continent to the particular host country.

References

Australian Department of Immigration and Ethnic Affairs. 1982. *Australian Immigration, Consolidated Statistics no. 12*. Canberra: Australian Government Publishing Service.

Birrell, R. 1984. A New Era in Australian Migration Policy. *International Migration Review* 18:65–84.

Borjas, George J. 1985. Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics* 3:463–89.

———. 1987. Self-Selection and the Earnings of Immigrants. *American Economic Review* 77:531–53.

Boyd, Monica. 1976. Immigration Policies and Trends: A Comparison of Canada and the United States. *Demography* 18:83–104.

Carliner, Geoffrey. 1980. Wages, Earnings, and Hours of First, Second, and Third Generation American Males. *Economic Inquiry* 18:87–102.

Chiswick, Barry R. 1978. The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* 86:897–921.

———. 1988. Immigration Policy, Source Countries, and Immigrant Skills: Australia, Canada, and the United States. In *The Economics of Immigration*, ed. Lyle Baker and Paul Miller, 163–206. Canberra: Australian Government Printing Service.

Chiswick, Barry R., and Paul W. Miller. 1985. Immigrant Generation and Income in Australia. *Economic Record* 61:540–53.

Douglas, Paul H. 1919. Is the New Immigration More Unskilled than the Old? *Journal of the American Statistical Association* 16:393–403.

Greenwood, Michael J. 1975. Research on Internal Migration in the United States: A Survey. *Journal of Economic Literature* 13:397–433.

Greenwood, Michael J., and John M. McDowell. 1986. The Factor Market Consequences of U.S. Immigration. *Journal of Economic Literature* 24:1738–72.

Heckman, James J., and Richard Robb. 1983. Using Longitudinal Data to Estimate Age, Period, and Cohort Effects in Earnings Equations. In *Analyzing Longitudinal Data for Age, Period, and Cohort Effects,* ed. H. Winsborough and O. Duncan, 173–50. New York: Academic.

Hicks, John R. 1932. *The Theory of Wages.* London: Macmillan.

Jackson, J. A. 1969. *Migration.* Cambridge: Cambridge University Press.

Jasso, Guillermina, and Mark R. Rosenzweig. 1985. How Well Do U.S. Immigrants Do? Vintage Effects, Emigration Selectivity, and the Occupational Mobility of Immigrants. University of Minnesota. Mimeo.

Keely, Charles B. 1979. The United States of America. In *The Politics of Migration Policies,* ed. D. Kubat, 51–64. New York: Center for Migration Studies.

Keely, Charles B., and Patricia J. Elwell. 1981. International Migration: Canada and the United States. In *Global Trends in Migration,* ed. M. Kritz, C. Keely, and S. Tomasi, 181–207. New York: Center for Migration Studies.

Kubat, Daniel. 1979. Canada. In *The Politics of Migration Policies*, ed. D. Kubat, 19–36. New York: Center for Migration Studies.

Price, Charles. 1979. Australia. In *The Politics of Migration Policies*, ed. D. Kubat, 3–18. New York: Center for Migration Studies.

Psacharopoulos, George. 1973. *Returns to Education: An International Comparison.* Amsterdam: Elsevier.

Roy, A. D. 1951. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers* 3:135–46.

Schwartz, Aba. 1968. Migration and Lifespan Earnings in the U.S. Ph.D. diss., University of Chicago.

Sjaastad, Larry A. 1962. The Costs and Returns of Human Migration. *Journal of Political Economy* 70:80–93.

Tandon, B. B. 1978. Earnings Differentials among Native Born and Foreign Born Residents of Canada. *International Migration Review* 12:406–10.

Unesco. 1969. *Statistical Yearbook, 1968.* Paris: United Nations.

———. 1980. *Statistical Yearbook, 1978–1979*. Paris: United Nations.
United Nations. 1982. *Demographic Indicators of Countries*. New York: United Nations.
U.S. Department of Commerce. Various Issues. *Statistical Abstract of the United States*. Washington, D.C.: U.S. Government Printing Office.
U.S. Immigration and Naturalization Service. Various Issues. *Statistical Yearbook of the Immigration and Naturalization Service*. Washington, D.C.: U.S. Government Printing Office.
World Bank. 1986. *World Development Report*. New York: Oxford University Press.