Self-Selection and the Earnings of Immigrants: Comment

By GUILLERMINA JASSO AND MARK R. ROSENZWEIG*

In a recent article in this *Review* (Borjas, 1987), George Borjas uses a standard model of self-selection to demonstrate that immigrants in the United States may not necessarily be positively selected, as is commonly assumed. Borjas' empirical results, based on U.S. Census micro data merged with country-of-origin characteristics for 41 sending countries, however, do not provide clear or consistent results with respect to the immigration selection issue, although they do confirm earlier findings that the characteristics of the countries of origin of U.S. immigrants explain a substantial proportion of the differences in their economic status (Jasso and Rosenzweig, 1986b).[1] The purpose of this note is to show that Borjas' results are themselves subject to biases due to additional processes of "self-selection." In particular, his inconclusive findings with respect to immigration selectivity arise from the use of a choice-based sample; enlargement of his highly selective sample of countries, based on the number of foreign-born in the United States, yields results that appear to conform much more closely and consistently to conventional wisdom, namely that the United States appears to attract those persons with above-average skills. We also provide evidence that changes in the earnings of aggregate immigrant entry cohorts, used by Borjas to measure the "assimilation" of immigrants, reflect another evidently important selection process, namely selective emigration. The effects of country characteristics on the wages of the foreign-born at entry (immigration selectivity) are thus not comparable to the effects of such variables on the changes in the characteristics of an immigrant cohort over time (emigration selectivity).

*Department of Sociology, University of Iowa and Department of Economics, University of Minnesota, Minneapolis, MN 55455. Partial support for this research was provided by the Russell Sage Foundation.

[1] See also Greenwood and McDowell (1986).

The effects via immigration self-selection of country-of-origin characteristics on the "quality" Q of immigrants who migrate to a destination country is neatly summarized by Borjas in his equation (12). With unobservables determining earnings normally distributed, the expected quality of immigrants in the United States is given by

$$E(Q) = \mu + \gamma\lambda, \tag{1}$$

where μ = earnings of the U.S. native born and $\lambda = \phi(z)/P$, where P is the emigration rate of the origin country and $\phi(z)$ is the density of the standard normal. Immigrants are positively selected if $\gamma > 0$ and negatively selected if $\gamma < 0$. The effect of, say, migration costs on $E(Q)$ is thus

$$\frac{\partial E(Q)}{\partial c} = \gamma \frac{\partial \lambda}{\partial z} \frac{\partial z}{\partial c}. \tag{2}$$

Since $\partial\lambda/\partial z > 0$, and $\partial z/\partial c > 0$ (as migration costs increase fewer persons migrate), increases in migration costs raise (lower) immigrant quality when immigrants are positively (negatively) selected. Similarly, when mean incomes (earnings opportunities) rise in the origin country, emigration declines and the quality of immigrants rises or falls depending on whether selectivity is positive or negative. Whatever the type of selectivity, mean origin-country income and migration costs should affect the quality of immigrants in the same way.

Borjas' empirical results with respect to origin-country mean income and migration costs are inconsistent with the model. While he finds that distance—the measure of migration costs—and mean origin-country income reduce emigration rates, mean origin-country income is positively associated with the initial wages of immigrants in the United States and with his assimilation measure, while distance has no effect on entry wages

TABLE 1—MEAN CHARACTERISTICS OF COUNTRIES IN BORJAS SAMPLE AND ADDITIONAL SAMPLE COUNTRIES

	Borjas Sample ($N=41$) (1)	Additional Countries ($N=66$) (2)	Extended Sample ($N=107$) (1)+(2)
Immigrants, 1975–80	55240	9395	27127
Foreign-Born Males Aged 25–64, Entered 1975–80[a]	600.3	133.2	312.2
Population Size (1970), in Thousands	41831	23122	30291
Distance, Miles	3745	5273	4687
Per Capita, GNP (1978), Dollars	3810	1775	2555
Centrally Planned	0.146	0.0606	0.0934
Literacy Rate (1975), Percent	84.8	67.8	74.3
Western Hemisphere	0.317	0.273	0.290
Proportion of World Population (1970)	0.498	0.421	0.919
Proportion of U.S. Foreign-Born Entered 1975–80 in 1980	0.661	0.226	0.887

[a] From *1980 Public Use Tapes,* 2.5:100 household A sample.

and has a negative or zero effect on assimilation. Borjas selects his sample of countries, however, on the basis of the number of foreign-born from each country residing in the United States in 1970; those countries with numbers of foreign-born below a minimum are selected out in order to maintain a minimum per country sample size, a constraint that is not relevant to the analysis of the initial earnings of the recent-entrant foreign-born in 1980. If this sample selection rule not only excludes those countries with lower absolute numbers of migrants in the United States but also those with lower emigration *rates*, then the effects of country-of-origin characteristics on emigration rates will be biased (to zero) by this sample truncation. And, from (2), if dz/dc, say, is biased to zero, then so will the estimated selectivity effects of changes in c.

In Table 1 we present the mean characteristics of the 41 countries selected by Borjas and an additional 66 countries for which we could obtain comparable characteristics (and used by us in prior work (Jasso and Rosenzweig, 1986a, 1986b, and forthcoming)).[2] As can be seen, Borjas' country sample accounts for less than half of the world's population and only 66 percent of the foreign-born who entered the United States between 1975 and 1980. The addition of the 66 countries extends the coverage to 89 percent of the world's population and to 92 percent of the newly entered U.S. foreign-born in 1980. Most importantly, the excluded countries (accounting for 23 percent of the foreign-born) have a substantially lower average emigration rate in the 1975–80 period (0.040 versus 0.132 percent), are located significantly farther from the United States on average (5273 versus 3745 miles), and are characterized by lower per capita GNP and literacy rates.

In Table 2 we report grouped probit (minimum-chi square) regressions of the emigration rate based on the Borjas and ex-

[2] Among the countries excluded from the Borjas sample but included in the extended sample are 10 countries in Africa, including South Africa, Liberia, Nigeria, and Kenya; 11 countries in South and Central America, including Chile, Costa Rica, El Salvador and Nicaragua; and Turkey, Syria, mainland China, Indonesia, Pakistan, Malaysia, Thailand, Vietnam, Australia and New Zealand. Sources of country characteristics are provided in Jasso and Rosenzweig (1986a).

TABLE 2—PROBIT REGRESSIONS OF EMIGRATION RATE BASED ON U.S. FOREIGN-BORN MEN AGED 25–64 ENTERED 1975–80[a]

Country of Origin Characteristics	Borjas Sample (1)	Borjas Sample (2)	Extended Sample (1)	Extended Sample (2)
Per Capita GNP ($\times 10^{-4}$)	0.0672	0.0160	−0.0328	−0.556
	(0.94)[b]	(0.24)	(0.31)	(3.43)
Distance ($\times 10^{-4}$)	−0.212	0.0481	−0.537	−0.776
	(1.77)	(0.35)	(9.54)	(9.91)
Literacy Rate	–	0.00234	–	0.00245
		(3.10)		(4.10)
Not Centrally Planned	−0.00793	0.0341	−0.116	−1.52
	(0.34)	(0.18)	(11.8)	(12.0)
Western Hemisphere	−0.221	−0.200	−0.211	−0.325
	(6.22)	(6.14)	(6.06)	(7.61)
Constant	−4.37	−4.70	−4.11	−3.96
	(53.2)	(36.0)	(69.2)	(58.7)
Number of Countries	41	41	107	107
χ^2	156.1	156.6	56.1	55.3

[a]Denominator is country-of-origin population in 1970.
[b]t-ratios in parentheses.

tended country samples, where we use the number of foreign-born residing in the United States in 1980 who entered in the period 1975–80 divided by origin-country population size in 1970 as the dependent variable. The number of resident foreign-born, based on 1980 Census data, is used to measure emigration because the number of legal immigrants, used by Borjas in his emigration regressions, does not correspond to the foreign-born population in the Census from which Borjas estimates his wage functions. The latter population includes, particularly among the newly entered foreign-born, persons with temporary visas permitting employment, foreign students, and various types of persons who are "illegal" immigrants. The resident foreign-born and legal immigrant populations are quite different—for example, of the foreign-born in 1980 who entered between 1975 and 1980, 20.2 percent are from Mexico, while only 14.9 percent of the total number of (legal) immigrants came from Mexico in the same period. Our emigration estimates thus more closely correspond to the selection equation underlying Borjas' entry wage function estimates.[3]

Results from two specifications are reported in Table 2, one with and one without the country's literacy rate (excluded by Borjas).[4] With either specification, we see that

[3]There are other reasons why the use of the Immigration and Naturalization Service (INS) legal immigration data is less desirable in this case. First, prior to 1982, the INS data do not distinguish between immigrants from mainland China and Taiwan, two countries with a distinctly different set of characteristics. While it is possible to identify the foreign-born from Taiwan (one of Borjas' set of 41) and mainland China based on the Census data, the "Chinese" immigrants used in Borjas' emigration regressions based on the INS data are from both mainland China and Taiwan. The INS immigration data used by Borjas also include both men and women (and children), while his wage analysis examines only adult men. In our earlier work (Jasso and Rosenzweig, 1986a and forthcoming) we have found the immigration behavior of men and women to be distinctly different; a larger proportion of legal immigrant women compared to men enter the United States via marriage to a U.S. citizen, while a larger percentage of men compared to women enter by dint of their labor market skills. As a consequence, for example, the presence of a U.S. military base in an origin country importantly influences the flow of female but not male immigrants to the United States from that country.

[4]Borjas' list of country characteristics also included two "political" variables and a measure of income inequality. These variables are not available for most of the countries in the extended sample. One political variable and income inequality did have statistically significant effects on emigration. Thus our emigration results may be subject to specification error, but it is not

the use of the truncated country sample results in a severe underestimate of the negative effect of distance on the probability of immigration to the United States, by a factor of over two with literacy rates excluded. When literacy rates are included, neither the effects of per capita GNP nor of distance are statistically significant in the Borjas sample; in contrast, both variables have the expected negative sign and are highly significant in the extended sample. We interpret the results in the last column of Table 2, estimated on the extended sample, as indicating that the United States offers higher returns, on average, to skills compared to other countries. Origin-countries with the same per capita income, but higher literacy rates, are characterized by lower returns to skills; such countries have significantly higher emigration rates. Similarly, countries with the same literacy rates but higher levels of per capita income are likely to have higher returns to skills, and exhibit significantly lower rates of emigration to the United States.

Does sample truncation also account for the absence of an effect of distance on the "entry" wage of the foreign-born in Borjas' results? In Table 3 we present estimates of the determinants of the log of the hourly wage rate of foreign-born males aged 25–64 who entered the United States in 1975–1980. We selected the same sample as Borjas and use the same Census-based regressors.[5] Our specification differs from his in that he allowed the effects of personal characteristics on the wage to differ across countries but assumed the effects of these variables to be the same within countries for different entry cohorts. We look at only one entry cohort, but assume that the effects of personal characteristics do not differ across countries.[6] Results based on the Borjas sample (41 countries) are reported in column 1. Despite our somewhat different specification, we are able to replicate the insignificant effect of distance and the positive and significant effect of per capita GNP on the entry wage rate with the truncated Borjas sample.[7]

In the second column of Table 3, we report the entry wage function estimates using the extended sample. While the effects of the personal characteristics on the wage are not sensitive to country selection, distance becomes statistically significant (at the 0.05 level) and its magnitude almost doubles when the sample is not truncated. The positive effects of both per capita GNP and of distance on the wage rates of the foreign-born are consistent with positive immigrant selection with respect to unobservables that influence wages, given that both variables inhibit immigration to the United States. The marginally significant literacy rate effect may be due to literacy measuring, in part, schooling quality; the sign of its effect does not therefore constitute as clear a test of selectivity as do the signs of the distance and origin-country income coefficients.

Borjas' additional findings that distance significantly reduces or has no effect on "as-

obvious how such specification error is responsible for the changes in estimates across the samples. Moreover, none of the excluded variables were found by Borjas to affect entry wages or assimilation rates. The only skill variable included by Borjas in the emigration equation was a measure of the English proficiency of the U.S. residents from the origin country, which we believe to be a poor measure of the average English proficiency, or average skill level, of the total *origin* population, given immigrant selectivity. We include, as did he, a dummy variable indicating whether or not the country is in the Western Hemisphere to capture, in part, the effects of the differential immigration criteria existing for the two hemispheres in the period between 1965 and 1978.

[5]The sample consists of employed (in 1979) foreign-born men aged 25–64 who were not self-employed or working without pay in 1979 and who were not in the Armed Forces or a resident of group quarters during the survey week. The sample was drawn from the 1:100 *A Sample of the 1980 Census Public Use Tape*.

[6]Borjas estimated separate wage functions based on Census samples from each of the 41 countries including a random sample of comparable U.S. native-born residents. It appears that the specification used assumed that the effects of all personal characteristics (see Table 3) except years in the United States were identical for the native-born and foreign-born populations. Since the reference population (U.S. native-born residents) is the same for all immigrants, we only employ a sample of the U.S. foreign-born.

[7]The variables excluded from our specifications, including income inequality, a variable cogently highlighted by Borjas in his theoretical analysis, did not have a statistically significant effect on either entry wage differentials or assimilation rates in Borjas' specification.

TABLE 3—DETERMINANTS OF LOG OF HOURLY WAGE RATE OF NEW ENTRANT FOREIGN-BORN MEN AGED 25–64 IN 1980

	Borjas Sample (N = 1949, Countries = 41)	Extended Sample (N = 2534, Countries = 91)
Country Characteristics		
Per Capita GNP ($\times 10^{-4}$)	0.699	0.664
	(7.29)[a]	(8.20)
Distance ($\times 10^{-4}$)	0.123	0.242
	(0.76)	(2.21)
Literacy Rate ($\times 10^{-3}$)	0.510	1.18
	(0.44)	(1.37)
Not Centrally Planned	0.229	0.0998
	(3.29)	(2.14)
Western Hemisphere	−0.0696	0.00363
	(0.75)	(0.05)
Individual Characteristics		
No English or English Not Well	−0.202	−0.198
	(4.62)	(5.22)
Age	0.107	0.0919
	(7.34)	(7.16)
Age Squared	−0.00122	−0.00105
	(6.84)	(6.73)
Schooling Attainment	0.0206	0.0231
	(4.75)	(6.31)
Married, Spouse Present	0.109	0.133
	(2.69)	(3.73)
SMSA	0.197	0.133
	(2.62)	(3.73)
Disability	−0.101	−0.142
	(0.79)	(1.26)
Constant	−1.16	−0.770
	(2.74)	(12.13)
R^2	0.211	0.190
F	43.2	49.2

[a] t-ratios in parentheses.

similation" while origin-country GNP increases it would appear to contradict our conclusions about positive immigration selectivity (although they are not consistent with negative selectivity either). However, aside from sample truncation, the change in the earnings of a foreign-born cohort defined by date of and age at entry to the United States, used by Borjas to measure assimilation, does not necessarily correspond to the change in earnings for a randomly chosen "typical" immigrant.[8] The survivors of an age-entry cohort after a period of time has elapsed are not the same people as the original entrants due to deaths and, most importantly, to the emigration or re-migration of the foreign-born.[9] Distance may impede emigration from the United States, as it does to the United States. If foreign-born emigrants (from the United States) are negatively selected (for example, returnees are those who fare less well than they expected) then the progress of an aggregate immigrant cohort may very well be negatively associated with distance even though immigrant selectivity is positive.

[8] Borjas does not actually track or compare each cohort over time, as he pools men of the same age in both 1970 and 1980. However, as he notes, this procedure is equivalent if the appropriate parameters describing the relations of interest are stable over time.

[9] Another reason is a change in the coverage of the Census. If the coverage change is selective by country-of-origin and income, inferences about differential assimilation rates (or the selectivity of emigration) based on aggregate cohort data from adjacent censuses are made even more difficult.

TABLE 4—DECADAL COHORT CHANGE (PERCENT) IN RATIO OF FEMALES TO MALES BY AGE GROUP AND NATIVITY: 1960–70 AND 1970–80

Nativity	Cohort Age in 1960			Cohort Age in 1970		
	20–29	30–39	20–44	20–29	30–39	20–44
Native Born	−0.48	2.90	0.97	−2.90	2.53	−0.97
New Entrants[a]—Eastern Hemisphere	23.9	1.14	13.0	49.7	9.38	24.0
New Entrants—Western Hemisphere	68.4	21.3	43.4	5.60	5.86	−2.36

Source: 1960, 1970, and *1980 Public Use Tapes.*

[a]New entrants in 1960 defined as those born in a foreign country and residing abroad five years prior to the 1960 Census; new entrants in 1970 defined as those born in a foreign country and entering the United States in 1965–70.

Borjas notes the possibility of emigration selectivity, but presents no evidence of the importance of this form of selection. While we have no direct information as yet on the magnitude or direction of re-migration selectivity in terms of earnings, we can adduce evidence from the Census data used by Borjas on the extent to which entry cohorts of the U.S. foreign-born are altered due solely to emigration (or other selection processes) rather than to changes in the characteristics of individuals who stay (assimilation). We need only compare statistics on time-persistent characteristics of an entry cohort over time; any changes in these characteristics must be due to selection out (emigration), since, by definition, no cohort additions occur unless sample coverage rates increase. One characteristic that does not change with time (with few exceptions) for an individual, is not subject to recall error, and is not likely to be altered for reasons of ego, is gender.[10] Changes in the sex ratio of an age-entry cohort is thus likely to signal emigration, although it will surely underestimate it to the extent that emigration rates by sex are similar.

Table 4 reports the percentage change over two ten-year periods in the ratio of females to males for the foreign-born entering in 1955–60 and entering in 1965–70, by age group, with comparisons to the sex-ratio change for the native-born, taken from the 1960, 1970, and 1980 Public Use Tapes. As can be seen, the changes are considerably greater for the foreign-born entry cohorts in each age group (but one) compared to the native-born. While there is essentially no change in the sex ratio for the native-born in the (initial) age group 20–44 over either ten-year period, in both periods the ratios of females to males among the foreign-born cohorts rise considerably, by as much as 43 percent, except for the Western Hemisphere foreign-born who entered 1965–70. And, as is consistent with evidence on age effects on mobility, changes in the sex ratio tend to be greatest among the foreign-born in the youngest age group. Moreover, the changes differ significantly by Hemisphere and by decade, as would be expected given the differential changes in immigration law by Hemisphere in the 1960 to 1980 period. Table 4 thus demonstrates that emigration by the U.S. foreign-born is important and is selective by sex, by age, and by entry cohort. Interpretation of the effects of origin-country characteristics on the changes in the earnings of age-entry cohorts of the U.S. foreign-born thus is likely to require a richer model of self-selection, incorporating decisions to migrate to the United States and to remain there after immigration.

[10]It is well known, for example, that retrospectively ascertained schooling attainment levels drift upward over the life cycle.

REFERENCES

Borjas, George J., "Self-Selection and the Earnings of Immigrants," *American Economic Review*, September 1987, *77*, 531–53.

Greenwood, Michael J. and McDowell, John M.,

"The Factor Market Consequences of U.S. Immigration," *Journal of Economic Literature*, December 1986, *4*, 1738–72.

Jasso, Guillermina and Rosenzweig, Mark R., (1986a) "Family Reunification and the Immigration Multiplier: U.S. Immigration Law, Origin Country Conditions, and the Reproduction of Immigrants," *Demography*, August 1986, *23*, 291–311.

_____ **and** _____, (1986b) "What's in a Name? Country-of-Origin Influences on the Earnings of Immigrants in the United States," *Research in Human Capital and Development*, Vol. 4, Greenwich, CN: JAI Press, 1986, 75–106.

_____ **and** _____, *The New Chosen People: Immigrants in the United States*, New York: Russell Stage, forthcoming.