Economic Assimilation of Foreign-Born Workers in the United States: An Overlapping Rotating Panel Analysis

Seik Kim[1]

Department of Economics

Yale University

seik.kim@yale.edu

http://pantheon.yale.edu/~sk442/

December 6, 2006

Abstract

This paper presents new evidence on whether foreign-born workers assimilate, which we define as the degree to which the wages of foreign-born workers approach those of comparable native-born workers with additional time spent in the United States. The key econometric challenge is to separate wage growth due to assimilation from composition effects. The composition of immigrant population varies over time due to variation in initial skill levels at year of entry and also because of nonrandom outmigration. Progress in measuring assimilation has been inhibited by the absence of panel data sets with adequate numbers of immigrants and by the problems of sample attrition and selective return migration. We develop a unique method for addressing panel attrition and outmigration and apply the method using the Current Population Survey (CPS). Overall, we find little evidence of a narrowing of the foreign-native gap in economic performance. The wages of new immigrants from Europe and Asia exceed those of natives and there is no strong evidence of convergence. New immigrants from Central and South America earn lower wages than natives, and this gap widens with time in the U.S. labor market. We also find that older migrants are more skilled than younger ones conditional on the year of entry. Our results suggest that analyses of immigrant wage growth based on repeated cross-section studies may be biased upward by individual heterogeneity. Controlling for this heterogeneity reverses the conventional result of economic assimilation.

Keywords: Economic Assimilation, Immigration, Outmigration, Overlapping Rotating Panel Data, Panel Attrition, Population Attrition

JEL Classification Number: C13, C23, J31, J61

[1] I am especially grateful to my advisor, Joseph Altonji, for his insightful suggestions, support, and encouragement. I also thank Yuichi Kitamura, Fabian Lange, and Mark Rosenzweig for their constructive advice and guidance. I have also benefited from helpful comments made by Donald Andrews, Nathaniel Baum-Snow, Debopam Bhattacharya, Meta Brown, Kenneth Chay, Mark Duggan, Donghoon Lee, Thomas Lemieux, Adriana Lleras-Muney, Taisuke Otsu, Peter Phillips, Paul Schultz, Aloysius Siow, Emiko Usui, Ebonya Washington, and seminar participants at Yale's Labor-Public Finance Lunch, Econometrics Prospectus Workshop, Graduate Summer Workshop, and Labor-Population Workshop.

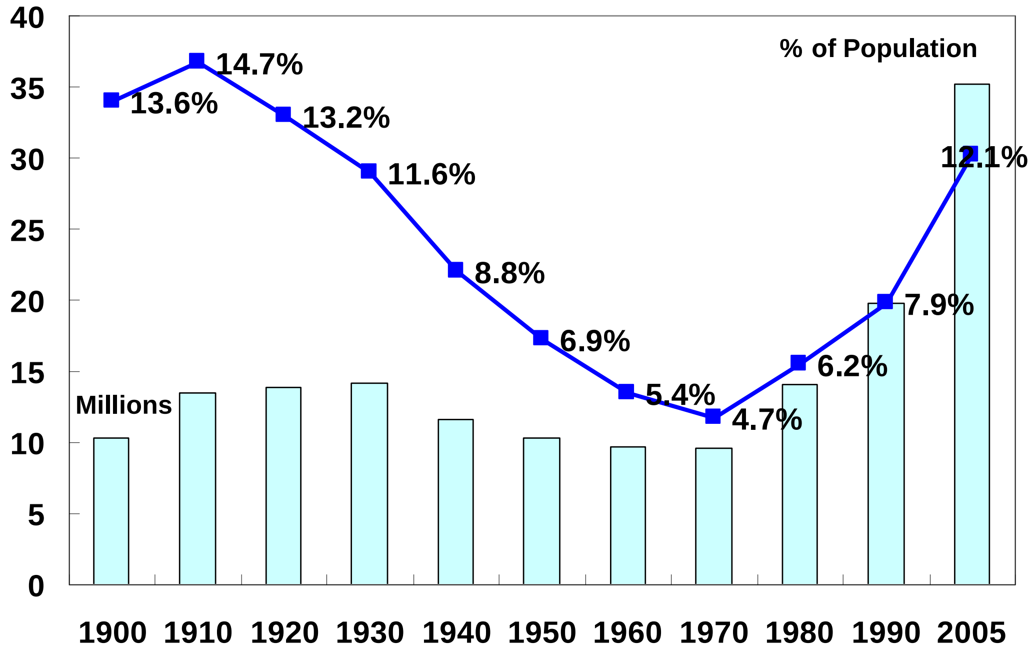

Figure 1: The Number and Share of Foreign-Born Population in the United States (1900-2005)

1 Introduction

The large and growing share of foreign-born workers in the United States has heightened interest in the economic impact of immigration. (See Figure 1.[2]) How immigrants fare as they accumulate experience in the U.S. labor market is the key to many of these effects.[3] First and foremost, the earnings of immigrants will directly effect the level and distribution of per capita income in the United States. Second, the better immigrants do on arrival and over time, the greater the extent to which their contributions as tax payers will outweigh their use of government services. Third, the greater the extent to which immigrants who enter the U.S. in low skill jobs quickly acquire country specific skills and spread into higher skill jobs, the smaller any negative impact on less skilled natives is likely to be.[4]

This paper presents new evidence on whether foreign-born workers assimilate, which we define as the degree to which the wages of foreign-born workers approach those of comparable native-born workers with additional time spent in the United States. Assimilation rates are the net result of several offsetting factors. Upon entry into the U.S. labor market, foreign-born persons may earn lower wages than their native counterparts to the extent that human capital in not perfectly transferable across economies and cultures and because employers are likely to have less knowledge about their productivity. On the other hand, some groups of foreign-born

[2] Figure 1 shows the number of foreign-born persons living in the United States over the last 100 years and their share of the total population. Since 1970 the foreign-born population has more than tripled. At the beginning of 2005, 35.2 million foreign-born persons were residing in the United States, making up 12.1 percent of the total U.S. population.

Source: http://www.cis.org/articles/2005/back1405.html, Center for Immigration Studies (CIS), based on the Decennial Censuses for 1900-1990 and CIS Analysis of March 2005 CPS.

[3] In U.S. immigration law the term "immigrant" or "permanent resident alien" denotes a person admitted to this legal classification. For expositional convenience, we use the terms "foreign-born person" and "immigrant" interchangeably although our sample possibly includes aliens in an illegal status.

[4] See Borjas (1995b) and LaLonde and Topel (1997) for discussions of the effect of immigrants on the labor market outcomes of natives.

workers might outperform natives if they possess superior skill endowments, stronger work ethics, or more powerful incentives. As immigrants stay longer in the United States, their wages might converge to those of natives.

The key obstacle to measuring assimilation rates is how to distinguish growth in earnings of particular immigrants from variation in initial skill levels associated with age at entry, year of entry, country of origin, and other factors. As Borjas (1985) points out, estimates of assimilation based on a single cross-section are biased if the ability and skill endowments of immigrants vary by year of entry. Studies using repeated cross-sections can control for the variation in skill composition by tracking the groups of individuals with same year of entry. However, such studies are vulnerable to bias from heterogeneity within an immigration year cell even when the country of origin is controlled for.[5] Furthermore, outmigration of the immigrants poses a problem for single cross-section and repeated cross-section analyses to the extent that it is systematically related to wage growth. For example, if less skilled and/or unlucky immigrants tend to return to their home country, stayers will on average earn higher wages than return migrants. Consequently, estimates using only stayers will tend to overstate relative labor market performance of immigrants compared to natives.

In principle, longitudinal data on native-born and foreign-born populations, by tracking specific individuals over time, offers the huge advantage of permitting one to control for fixed unobserved heterogeneity. In practice, longitudinal analysis of U.S. immigrants has been limited by two key factors. First, sample sizes of immigrants in U.S. panels such as the Panel Study of Income Dynamics (PSID) or National Longitudinal Survey of Youth 1979 (NLSY79) are too small. Second, the use of panel data gives rise to an additional problem: nonrandom panel attrition. For example, if persons with negative (positive) wage shocks are more likely to drop out of the sample, panel data estimates will overstate (understate) the growth of wages. Unless the attrition rate is low, panel attrition may distort the empirical findings. In addition, outmigration of immigrants that is related to wage growth poses another attrition problem for panel data analyses as well as for single cross-section and repeated cross-section analyses.

This paper provides a longitudinal analysis of assimilation. We address the sample size problem by using Merged Outgoing Rotation Groups (MORG) of the Current Population Survey (CPS) for 1994 to 2004. The CPS MORG is an overlapping rotating panel, which is a collection of panels of length two, and has the crucial advantage of being much larger than alternative panel data sets.[6] However, the panel attrition problem is particularly severe as the survey does not follow households who move. To address the problems of panel attrition as well as outmigration, we draw on recent work by Hirano, Imbens, Ridder, and Rubin (2001) to develop an estimation procedure for use with overlapping rotating panel data which accounts for both problems.[7]

The identification strategy involves using the availability of representative cross-sections as the basis for weighting persons in the matched sample. In the absence of outmigration, the attrition correcting weighting function is given by the inverse of one minus the probability of panel attrition. Panel attrition can depend on both exogenous and endogenous variables. Hirano et al. show that the attrition process can be identified

[5] See Borjas (1985) for a critique of studies based on single cross-sections and for the first application of a synthetic cohort analysis based on repeated cross-sections. Other examples of repeated cross-section analysis include Borjas (1995a), Duleep and Regets (1997b), and LaLonde and Topel (1992).

[6] A rotating panel is a collection of panel data sets which are usually short. If the sample periods of the short panels overlap, the collection of short panels is called the overlapping rotating panel data set.

[7] Panel attrition and outmigration rates for 1994 to 2004 are about 22-40% and 3% per year, respectively, and panel attrition has a larger impact on estimation results than outmigration does.

under fairly flexible (additive non-ignorable) assumptions up to a known link function such as the logit or probit when a panel and representative cross-sections are available. An estimation strategy is developed by Bhattacharya (2006).

When the target population is nonstationary and the model of interest requires a counterfactual situation of what if the population had remained stationary, the attrition correcting technique cannot be used. We develop an estimation scheme that can be used when the target population is nonstationary. In our analysis, the nonstationarity arises from outmigration. Outmigration can be identified nonparametrically when representative cross-sections of the U.S. population are available, although much more restrictive assumptions are needed about the factors that drive it than is necessary to handle panel attrition alone. The key is that outmigration can be identified without knowing who emigrated from the United States. In the presence of outmigration, we show that weighting representative cross-sections by one minus the probability of outmigration prior to estimating the panel attrition correcting weighting function produces consistent estimators.

In contrast to much of the literature based on repeated cross-sections, we find little evidence of a narrowing of the foreign-native gap in economic performances with time since immigration. The wages of new immigrant workers from Europe and Asia exceed those of natives and there is no strong evidence of subsequent convergence. New immigrant workers from Central and South America earn lower wages than natives, and the wage gap widens with time spent in the United States. Another interesting finding relates to heterogeneity of immigrant skills at the time of entry into the United States. It appears that older migrants are more skilled than younger ones conditional on year of entry and other observables. The results suggest that past estimates of immigrant wage growth based on repeated cross-section studies are biased upward by individual heterogeneity.[8] Controlling for this heterogeneity reverses the conventional result of economic assimilation.

The paper proceeds as follows. Section 2 defines economic assimilation, outlines the model for economic performance, and introduces some of the key econometric issues. Section 3 introduces the data set. The summary statistics suggest that panel attrition and possibly outmigration must be taken into account in estimation of economic assimilation. They also provide initial evidence that different ethnic groups may experience assimilation differently. Section 4 derives a consistent estimator in the presence of panel attrition and outmigration. Section 5 presents the main results and Section 6 offers conclusions and a research agenda.

2 Issues in Measuring Economic Assimilation

2.1 Definition of Economic Assimilation

In this paper, economic performance is measured by hourly wages. Economic performance of an immigrant is generated by

$$y_{it} = h_{imm}\left(age_{it}, ysm_{it}, edu_i, \mu_i, t\right), \tag{1}$$

and that of a native by

$$y_{it} = h_{nat}\left(age_{it}, edu_i, \mu_i, t\right), \tag{2}$$

[8] Changes in the composition of natives who are in the labor market can also lead to different estimates of the age profile for natives relative to what one would get in a repeated cross-section. Failure to account for such changes would bias repeated cross-section estimates. There have been changes in labor force participation rates by age, race, and education in the U.S. population. Note that when estimating sample attrition weights, we include labor force participation status. In the CPS MORG, we find that attrition is negatively correlated with wage growth.

for some functions $h_{imm}(\cdot)$ and $h_{nat}(\cdot)$, where y is the logarithm of the hourly wage, age is the worker's age, ysm is the number of years since migration, edu is the number of years of education, μ reflects ability or skill endowment, and t reflects market conditions and economic shocks. Years since migration combined with age reflects an immigrant's gain such as information acquisition, human capital accumulation, and employer learning. Ability or skill endowment is not observed but may be correlated with year of entry, age at migration, and country of origin.

We say that economic assimilation occurs if the economic performance of foreign-born workers approaches those of comparable native-born workers with additional time spent in the United States. There exist at least three alternative definitions of economic assimilation frequently used in the previous literature. The first definition compares wages between typical foreign-born and native-born persons. According to this definition, economic assimilation occurs if the economic performance of a foreign-born person converges to that of a representative native-born person. The second definition compares wages between earlier and later arrivals within the foreign-born population. According to the second definition, economic assimilation occurs if an earlier migrant performs better in the labor market than a recent migrant conditional on age, education, and factors other than years since migration. The third definition compares wages between foreign-born workers and their ethnically similar native-born counterparts. The reference group of the third definition lies in between those of the first and second definitions. We adopt the first definition and focus on the typical foreign-native differentials. So, the reference group consists of white natives. We restrict the analysis to men.

The economic performance of a typical foreign-born worker relative to a representative native-born worker at time t can be measured by

$$EA(age, ysm; t) = \left.\frac{d}{dt}h_{imm}\right|_{(age,ysm,t)} - \left.\frac{d}{dt}h_{nat}\right|_{(age,t)}. \tag{3}$$

Roughly speaking, $EA(age, ysm; t)$ is a difference-in-difference estimator. It reflects the rate of convergence in wages between foreign-born and native-born workers. Many studies find that foreign-born workers initially earn lower wages than average native-born workers. In this case, wage convergence from below toward the higher native mean, $EA(age, ysm; t) > 0$, means economic assimilation.[9] This paper also considers the case where foreign-born workers initially earn higher wages than average native-born workers. In this case, we say that economic assimilation occurs if their wages converge from above toward the lower native mean, $EA(age, ysm; t) < 0$.

Figure 2 illustrates an idea of how economic assimilation can be measured using the repeated cross-sections. The sample is drawn from the CPS. The figure depicts the mean hourly wages of foreign-born and native-born workers of various age groups during 1994-2004. The foreign-born workers in the figure are confined to those who arrived between 1980 and 1991. For the time being, assume that selective return migration is negligibly small. The three thicker lines with larger symbols indicate the mean wages of native-born workers and the three thinner lines with smaller symbols indicate the mean wages of foreign-born workers. The solid lines with squares track the mean wages of those who were 20-24 years old in 1994. The dashed lines with triangles are the mean wages of those who were 30-34 years old in 1994. The dotted lines with circles correspond to the mean wages of those were 40-44 years old in 1994. Therefore, behavior of the gaps between the thicker and

[9] Although we focus on the mean wages, the technique developed later in this paper can be applied to the entire distribution of wages.

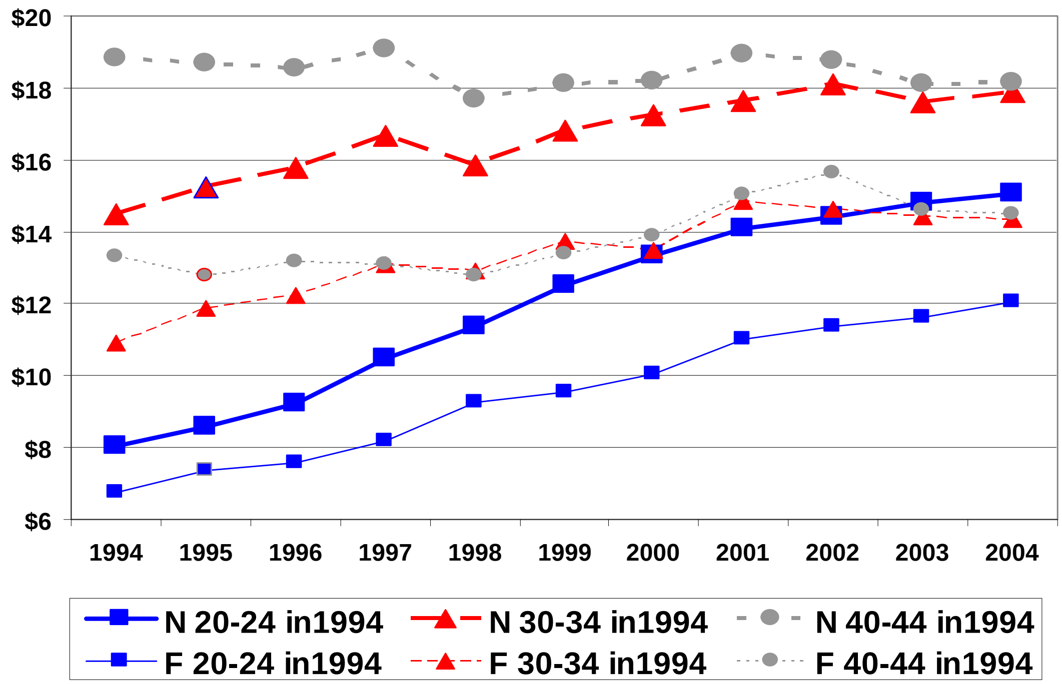

Figure 2: Wage Dynamics of Native-Born and Foreign-Born Workers

the thinner lines of same type with identical symbols measure economic assimilation.[10]

We observe that the wage gap between the immigrants and the natives in the "20-24 in 1994" cohort widens as the foreign-born workers stay longer in the United States. Foreign-born workers who were 20-24 years old in 1994 fail to assimilate economically during the 1994-2004 period. The foreign-born workers in the "30-34 in 1994" cohort also fail to catch up over the 1994-2004 period–the wage gap remains stable. The foreign-born workers in the "40-44 in 1994" cohort experience economic assimilation over the 1994-2004 period as the wage gap narrows. Later in this paper, we show that the findings in Figure 2 are supported by the estimation results.

2.2 Advantages and Disadvantages of the Available Data Sets

The key econometric challenge is to distinguish the growth in earnings of particular immigrants from variation in initial skill levels associated with year of entry, age at migration, country of origin, and other factors. Most empirical studies on economic assimilation rely on cross-section data drawn from the U.S. Census or the CPS. Estimates of assimilation using a single cross-section are biased if the ability of immigrants varies by year of entry. Studies using repeated cross-sections can control for the variation in skill composition by tracking individuals with the same year of entry. However, such studies are vulnerable to bias from heterogeneity within an immigration year cell. This is because the ability, skill endowment, and work motivation of new immigrants vary over time, which can bias estimates of the relative wage growth of immigrant and native workers. When fixed unobserved heterogeneity is correlated with economic performance conditional on age,

[10] To be precise, we need more controls, such as years since migration, year of entry, and country of origin, and need to look at log wage differences rather than wage differences.

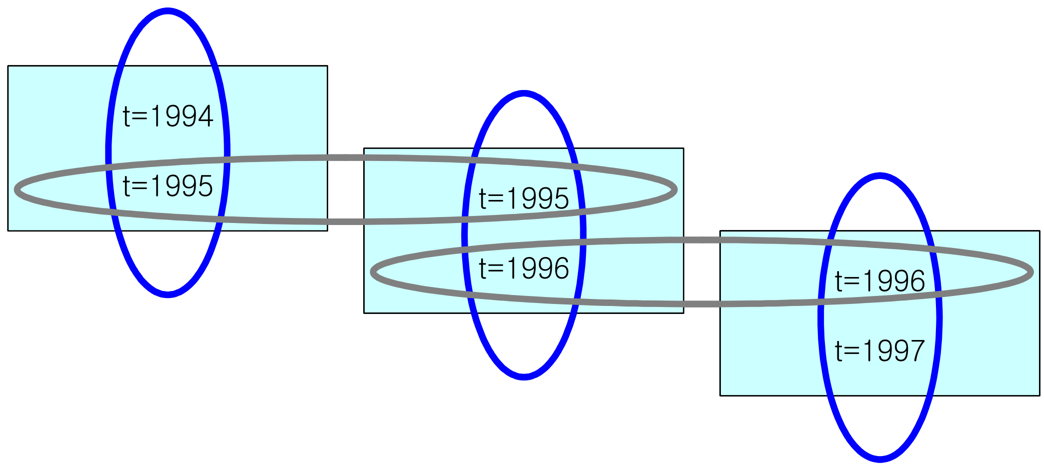

Figure 3: Data Structure of an Overlapping Rotation Panel Data Set

years since migration, and other observables, it is desirable to use fixed effects models with panel data.

Several studies do use longitudinal samples, but most of the panels have few foreign-born workers or are for non-representative samples. For instance, Chiswick (1980) uses the National Longitudinal Survey (with 98 male immigrants who all arrived before 1965) and Borjas (1989) uses a longitudinal survey of scientists and engineers. A representative random sample of permanent residents from the Immigration and Naturalization Service for fiscal year 1971 used by Jasso and Rosenzweig (1988) does not include wage information. More recently, Hu (1999) uses the Health and Retirement Study (HRS) linked to the Social Security Earnings data and Lubotsky (2000) uses the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) linked to the Social Security Earnings data. They collect a sample of individuals with known social security numbers from cross-sections and connect time series of their past social security earnings. These linked data, however, fail to include immigrants who left the United States before the cross-section interview period. In addition, the data may underrepresent immigrants working in the uncovered sector or the underground economy.

An ideal sample for estimating economic assimilation would be a longitudinal data set containing a large representative sample of foreign-born and native-born persons. Longitudinal data on native-born and foreign-born populations permit one to track specific individuals over time and thus control for fixed unobserved heterogeneity. In practice, however, longitudinal analysis gives rise to an additional problem: panel attrition. Furthermore, outmigration of the immigrants poses a fundamental problem for both panel and cross-section analyses to the extent that it is related to wage growth. Although panel analyses are robust to a link between outmigration and individual fixed effects, if the missing individuals are nonrandomly selected, longitudinal analyses must be approached cautiously.

Given that an ideal sample is not available, it is desirable to have a data set which enables us to control for panel attrition and outmigration. As we show, one may do so with an overlapping rotating panel data set, such as the CPS MORG. The data structure of an overlapping rotating panel is depicted in Figure 3, where the short panels are represented by the blocks. Vertical circles symbolize the longitudinal feature of an overlapping rotating panel. Horizontal circles illustrate the overlapping feature of the short panels. As the sampling periods of two adjacent short panel data sets overlap, short panels can mimic a longitudinal sample if combined properly.

An overlapping rotating panel data set shares most of the advantages of usual panel data sets and is superior in some dimensions. First, the sample has a longitudinal feature. This means that usual panel data models, such as the first difference or the fixed effects models, can be used to control for individual specific permanent components. Second, the rotating panel that we use, the CPS MORG, is large, which makes it even more powerful than a usual panel, such as the PSID or the NLSY79. Sample sizes matter in immigration studies because foreign-born persons, after all, are minorities. Third, the sample serves as a representative cross-section of the target population for any given time period. This property is the key in identifying panel attrition and outmigration processes.

3 Data Description

This section introduces the structure of the data set. Then it reports summary statistics. We discuss imputed wages, panel attrition, and outmigration in the sample. We also discuss differences in wage growth by ethnic origin. Finally, the section summarizes the lessons from the summary statistics.

3.1 The Current Population Survey and its Merged Outgoing Rotation Group

The CPS sample is a collection of representative cross-sections. The CPS collects a sample of approximately 56,000 housing units from 792 sample areas. Each month, data are collected from the sample housing units on demographic and labor force characteristics of the civilian non-institutional population 16 years of age and older. Since 1994, the CPS includes information on international migration, such as year of entry to the United States and country of birth along with demographic and labor market information, such as age, schooling, marital status, earnings per hour or week, usual hours of work, and labor market status. Prior to 1994, CPS supplements on immigration were administered to all households participating in the survey in November 1979, April 1983, June 1986, June 1988, and June 1991.

The design of the CPS is as follows. A housing unit is interviewed for 4 consecutive months, is dropped out of the sample for the next 8 months, is brought back in the following 4 months, and then is retired from the sample.[11] If a household is included in either the first or the last 4 months of the interview periods, it is said that the household is in the rotation group. Figure 4 demonstrates the sample design for a housing unit which, for instance, joins the survey on March 1994. This housing unit is interviewed from March to June in 1994 and 1995. The pre-selected housing units are kept unchanged over the interview periods. If the occupants of a dwelling unit move, the new occupants of the unit are interviewed. Although the interviewees may be replaced by new occupants within the sampling periods, the CPS provides a representative cross-section of the target population because the random sample of housing units is kept fixed.

An interesting feature of the CPS sample is its rotation scheme. Selected questions on labor market information, such as usual weekly earnings and usual weekly hours worked, are asked only in the last interview

[11] About 3/4 of the first and fifth interviews are conducted by visiting. In other interview months, almost 90% of the interviews are conducted over the phone. The rotation scheme ensures that in any 1 month, one-eighth of the housing units are interviewed for the first time, another eighth is interviewed for the second time, and so on. That is, after the first month, 6 of the 8 rotation groups will have been in the survey for the previous month; there will always be a 75 percent month-to-month overlap. When the system has been in full operation for 1 year, 4 of the 8 rotation groups in any month will have been in the survey for the same month, 1 year ago; there will always be a 50 percent year-to-year overlap.

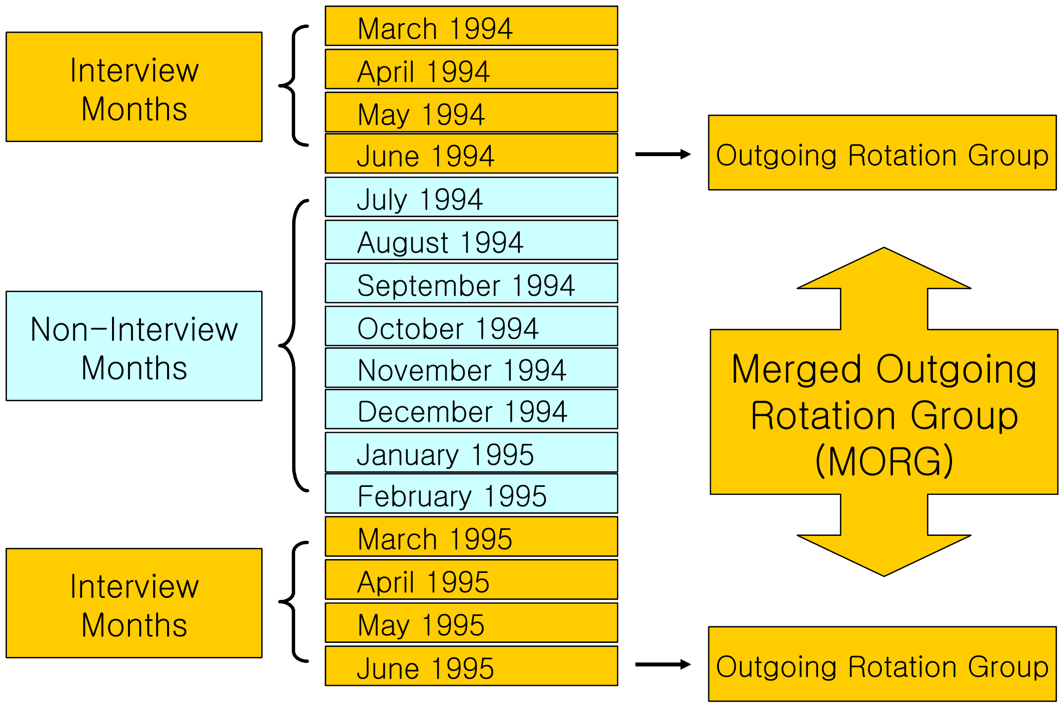

Figure 4: Sample Design of the CPS and its Merged Outgoing Rotation Group

of each 4-month rotation group. The sets of households in the fourth or eighth month are called the outgoing rotation groups. If records from the 4th and 8th interviews are appended, we get repeated observations on the same individuals. The appended sample is called the Merged Outgoing Rotation Group (MORG) data. (See Figure 4.) By construction, an individual appears only once in a year, but may reappear in the following year. Due to the 4-8-4 rotation scheme, the CPS MORG is an overlapping rotating panel data set comprised of multiple panels of length two. The 1994-1995 panel, for instance, contains the individuals in the households which enter the survey scheme between October 1993 and September 1994.

3.2 Summary Statistics

The sample used in this analysis is drawn from the CPS MORG between 1994 and 2004. We take a sample of foreign-born and native-born men of ages 18-64.[12] We define an individual as matched if the individual appears twice in the CPS MORG. In order to examine differences based on ethnic origin, we divide the foreign sample into 4 groups: immigrants from Central and South America, from Europe (including Australia, New Zealand, and Canada), from Asia, and from other countries.[13] The group of the other countries consists of immigrants from Africa, Oceania, and unclassified ones. The last group is of little interest due to its small

[12] The foreign sample includes foreign-born men who were not U.S. citizens at the time of birth. Following Warren and Peck (1980), our foreign sample consists of persons born outside the United States, the Commonwealth of Puerto Rico, and the outlying areas of the United States. Foreign-born persons may have acquired U.S. citizenship by naturalization or may be in illegal status. The reference group consists of native-born white men. The native sample includes persons born in the Unites States, but excludes persons born in the Puerto Rico and the outlying areas.

[13] We combine Australia, New Zealand, and Canada with Europe because of sample size considerations and so that immigrants from countries that are predominantly white and are at a similar stage of political and economic development are grouped together. We refer to the group as Europe. The data do not identify mother tongue. The impact of language proficiency has been studied in a large literature. LaLonde and Topel (1997) provide a survey.

sample size and heterogeneity. Details on how the data are processed are explained in the Appendix. This section provides a general picture.

Table 1 reports summary statistics for cross-section/matched and all/reported wages samples. In this section we focus on the cross-section sample with all individuals. Years of education provides a rough measure of skill endowment. Foreign-born persons have lower mean and a much larger standard deviation of education. In the cross-section sample with all the individuals, the average education level is 13.6 years for native-born persons and is 12.0 years for foreign-born persons. Immigrants from Central and South America have 10.0 years of average education, those from Europe 13.7 years, those from Asia 14.2 years, and those from the other countries 13.7 years.

The average hourly wage of native-born workers is $16.0, in 1994 dollars, while the average foreign-born worker earns $13.0. Immigrants from Central and South America make $9.8 per hour, those from Europe $18.4, those from Asia $16.5, and from the other countries $14.7. The estimates also indicate that foreign-born persons are about 2 years younger than native-born persons on average. Immigrant workers work 1.3-1.4 more hours per week than native workers. 79.0% and 78.7% of the foreign-born and native-born populations are full-time workers, 5.4% and 5.8% are part-time workers, respectively. Although not reported in the table, the proportions of full-time and part-time workers are relatively stable over the sampling period: 75-82% and 5-7% of foreign-born population and 76-80% and 5-6% of native-born population are full-time and part-time workers, respectively. We also find that a larger proportion of the foreign-born population is married.

Table 1. Summary Statistics

	Cross-Section Sample				Matched Sample			
	All		Reported Wage		All		Reported Wage	
	Nat.	Imm.	Nat.	Imm.	Nat.	Imm.	Nat.	Imm.
Age	41.1	39.4	41.4	39.4	42.5	40.8	42.8	40.8
	(12.1)	(11.6)	(12.3)	(11.7)	(11.3)	(11.2)	(11.4)	(11.3)
Education	13.6	12.0	13.7	11.9	13.7	12.1	13.7	11.9
	(2.4)	(4.3)	(2.4)	(4.3)	(2.4)	(4.3)	(2.5)	(4.4)
C.S.America		10.0		9.9		10.1		9.9
		(4.1)		(4.3)		(4.2)		(4.2)
Europe		13.7		13.8		13.7		13.7
		(3.3)		(3.3)		(3.3)		(3.4)
Asia		14.2		14.2		14.3		14.3
		(3.4)		(3.4)		(3.4)		(3.4)
Others		13.7		13.6		13.7		13.5
		(3.5)		(3.6)		(3.6)		(3.7)
Wage	16.0	13.0	16.2	12.8	16.5	13.5	16.6	13.5
	(15.5)	(12.9)	(15.2)	(13.1)	(15.3)	(13.5)	(15.4)	(14.4)
C.S.America		9.8		9.4		10.2		9.8
		(7.2)		(6.8)		(7.3)		(7.2)
Europe		18.4		19.6		18.9		20.4
		(18.6)		(19.8)		(19.6)		(21.3)
Asia		16.5		17.0		17.0		17.8
		(15.5)		(16.9)		(16.3)		(18.3)
Others		14.7		13.9		14.6		14.7
		(15.9)		(13.8)		(15.0)		(15.2)
Hours	43.4	42.0	43.6	42.3	43.8	42.3	44.2	42.9
	(10.5)	(9.5)	(10.9)	(9.8)	(10.3)	(9.6)	(10.9)	(10.3)
Full Time	0.787	0.790	0.746	0.750	0.814	0.810	0.767	0.760
Part Time	0.058	0.054	0.058	0.052	0.049	0.050	0.050	0.050
Marital Status	0.640	0.680	0.639	0.682	0.696	0.730	0.699	0.739
U.S. Citizen	1.000	0.385	1.000	0.387	1.000	0.440	1.000	0.434
C.S.America		0.513		0.529		0.497		0.508
Europe		0.163		0.161		0.179		0.181
Asia		0.256		0.254		0.265		0.262
Others		0.068		0.056		0.059		0.049
N	872598	126240	578519	82630	254837	34018	167981	20718

Standard deviations are reported in parentheses. N: sample size

All: reported & imputed wages; Reported Wage: reported wages only; Nat.: native sample; Imm.: foreign sample

Wage: hourly rate of pay; Hours: usual hours worked per week

Marital Status: 1 if married; U.S. Citizen: 1 if U.S. citizen

C.S.America: Central and South America; Europe: Europe, Australia, New Zealand, and Canada;

Asia: Asia; Others: Africa, Oceania, and other countries

3.2.1 Imputed Wages, Panel Attrition, and Outmigration

The wage information in the CPS sample is mostly self-reported, but also involves imputed wages. Throughout the sample period, an increasing fraction of workers do not answer questions about wages. When a person is working but does not report the wage, the Census Bureau assigns values for the missing wages using an allocation rule which is known as the cell hot deck match criteria.[14] The native imputation rates are about 17-23% with an increasing trend from September 1995 through 2004. The foreign imputation rates are higher than the native ones by 2-4% points. The imputation rates are homogeneous across different ethnic groups.

In Table 1, we observe that mean characteristics of persons with reported wages are different from those in the entire sample, especially among foreign-born workers. For instance, the imputed wages for those from Central and South America are higher than the reported wages and those from Europe and Asia are lower. As the imputation rule does not account for the country of origin, the imputed wages of immigrant workers tend to be biased toward the wages of native workers. Consequently, our preferred way to handle the imputed wages is simply dropping them.[15]

Matching is directly related to residential mobility and outmigration as the housing units in the sample are kept fixed over the interview periods, provided that the non-interview rate is low.[16] Between 1994 and 2004, the attrition rates are 28-40% among the immigrant samples and 22-32% among the native samples. In practice, matching is not possible between June 1994 - August 1995 and June 1995 - August 1996 due to sample redesign. If samples in 1994-1995 and 1995-1996 are excluded, the attrition rates are 28-35% among the immigrant samples and 22-29% of the native samples. The gaps between the foreign and native attrition rates are stable in these periods ranging 6-8% points. A part of the gap in the attrition rates may be due to outmigration. Attrition rates are reported in Table A4 of the Appendix.

The summary statistics for the matched sample in Table 1 summarize the first year observations. We observe that persons in the matched samples, regardless the ethnic origins, tend to earn more, work longer, and participate more in the labor market than those in the cross-section samples. It implies that more successful workers are more likely to be matched than unsuccessful ones. Foreign-born persons from Central and South America tend to attrite more than those from Europe and Asia. The consequence of nonrandom attrition, however, has not been addressed in immigration studies using the matched CPS.[17] We find substantial panel attrition bias.

The United States stopped collecting information on return migrants in 1957. To estimate outmigration rates, we exploit the structure of the CPS MORG. As housing units in the sample are kept fixed over the

[14] According to the imputation rule, a value of the wage is allocated based on the cell of same gender, age, race, education, occupation, hours worked and receipt of tips, commissions, or overtime. (The numbers of cells are 14976 in 1994-2002 and 11520 in 2003-2004.)

[15] Hirsch and Schumacher (2004) raise the problem of imputed wages. They find that regression estimates including variables not used in imputation rules, such as union status, are biased. As country of origin is not used as imputation criteria, using the whole sample may bias the results. Bollinger and Hirsch (2006) propose a weighting scheme to correct for the bias.

[16] The average yearly non-interview rates for the CPS in the early 1990's are as low as 4-7%. This non-interview rate is comparable with the initial non-response rate of the National Longitudinal Survey of Youth 1979 (NLSY79), which is 10%. The Census Bureau classifies the noninterviews into three types. Type A noninterviews are for household members that refuse, are absent during the interviewing period, or are unavailable for other reasons. Type B noninterviews include a vacant housing unit (either for sale or rent), a unit occupied entirely by individuals who are not eligible for a CPS labor force interview, or other reasons why a housing unit is temporarily not occupied. Type C noninterviews are for addresses that may have been converted to a permanent business, condemned or demolished, or fall outside the boundaries of the segment for which it was selected.

[17] While many papers have used the matched CPS, only two that we are aware of focus on immigration: Duleep and Regets (1997a) and Bratsberg, Barth, and Raaum (2006).

sampling period, the decrease in the sample size of immigrants will imply outmigration. Using the panels prior to trimming individuals with extreme wages or negative experience, Table 2 provides the ratios of persons staying in the United States (one minus the outmigration rates) by year of entry. For instance, the cell in the first row and first column indicates that in the 1st year of the 1994-1995 panel, there were 5329 foreign-born persons in the United States. Then we count the number of foreign-born persons in the 2nd year of the 1994-1995 panel, which is 5331. We take the ratio between these numbers and get 1.00 (=5331/5329). This roughly means that little outmigration occurred during this period. Similarly in 1995-1996, the numbers of the foreign-born persons in the first and the second years are 5417 and 4605, respectively. It implies that about 15% (=1–4605/5417) of the foreign-born population in 1995 left the United States in 1996.

Conceptually, it is impossible to have the stay rate exceed unity (or the outmigration rate below zero). Estimates above unity could arise from sampling error and/or if the reentering foreign-born persons report their previous entry years. In the sample, values greater than unity are observed frequently, implying that sampling errors and measurement errors are relatively large. Taking this into account, the last column reports the stay probability over the entire sample period. The last column of the first row reports that 25.2% (=1–0.768) of the foreign-born population who arrived in the United States in 1994 or before left the country in 2004.[18] On average, 2.6% (=1–0.974) of the foreign-born population outmigrates. The stay probability by ethnic origin is reported in Table A1 of the Appendix.

3.2.2 Wage Growth Differential

Table 3A presents the wage growth differentials of foreign-born workers relative to native-born workers using the matched sample of all wage information. For instance, in 1994-1995, the native mean wage increased by \$0.01 or 0.0% (from \$17.23 to \$17.24) and the foreign mean wage increased by \$0.50 or 3.6% (from \$14.08 to \$14.58). The column DD (difference-in-difference) reports the difference between the wage increment of the two groups, \$0.49 or 3.6% points. Roughly speaking, in 1994-1995, foreign-born workers experienced more rapid wage growth than native-born workers. In 1995-1996, native-born workers have steeper wage growth than foreign-born workers by \$0.39 or 2.2% points. However, most of these estimates are noisy. In general, we find no significant pattern with respect to which group is doing better than the other in terms of wage increments.[19] Table 3B reports the relative wage growth estimates by simply excluding the imputed wages. Again, most of the DD estimates are noisy. Our summary statistics do not support the hypothesis of economic assimilation for 1994-2004 because the first DD columns in Tables 3A and 3B suggest that the sign of relative wage growth is rather volatile and unpredictable.

[18] This estimate is consistent to other empirical findings. For instance, Warren and Peck (1980) estimate that more than 1/6 of total immigrants admitted during the 1960s emigrated by the end of the decade.

[19] Duleep and Regets (1997a) provide a similar table for 1987-1988. The mean wages of foreign-born workers are \$9.51 in 1987 and \$10.48 in 1988. The mean wages of native-born workers are \$11.27 in 1987 and \$11.88 in 1988. Consequently, their estimate of the wage growth of foreign-born workers exceeds that of native-born workers by \$0.36 or 5.4% points. Hence their result supports the economic assimilation hypothesis for 1987-1988.

Table 2. Stay Probability (One minus the Outmigration Rate) by Arrival Year

	1994 –1995	1995 –1996	1996 –1997	1997 –1998	1998 –1999	1999 –2000	2000 –2001	2001 –2002	2002 –2003	2003 –2004	[row in total] row average
all foreign persons											
# in 2nd year	5331	4605	5011	5070	5398	5578	6299	6293	6831	6090	
# in 1st year	5329	5417	5121	5220	5527	5435	6060	6021	7001	6811	[0.768]
stay probability	1.000	0.850	0.979	0.971	0.977	1.026	1.039	1.045	0.976	0.894	0.974
before 1980 arrivals											
# in 1st year	2524	2417	2158	2078	2090	1936	1893	1793	1904	1745	[0.644]
stay probability	0.998	0.826	0.966	0.949	0.944	0.999	1.031	1.022	0.957	0.896	0.957
1980-1981 arrivals											
# in 1st year	517	615	511	520	467	474	458	534	457	483	[0.931]
stay probability	0.965	0.862	0.971	0.952	1.000	1.108	1.083	1.060	1.039	0.917	0.993
1982-1983 arrivals											
# in 1st year	323	343	282	317	329	294	321	313	349	338	[0.844]
stay probability	0.947	0.930	1.035	0.987	0.936	0.959	1.078	1.099	1.003	0.879	0.983
1984-1985 arrivals											
# in 1st year	456	521	411	451	401	389	429	395	444	447	[1.041]
stay probability	1.042	0.904	1.010	0.940	0.983	1.103	1.061	1.104	0.977	0.940	1.004
1986-1987 arrivals											
# in 1st year	400	433	421	405	353	357	375	353	426	409	[1.062]
stay probability	1.055	0.885	0.964	1.007	1.057	1.050	1.053	1.125	1.035	0.861	1.006
1988-1989 arrivals											
# in 1st year	567	545	473	529	528	596	502	497	498	527	[0.672]
stay probability	0.984	0.809	0.981	1.000	0.992	0.938	0.982	1.012	1.044	0.890	0.961
1990-1991 arrivals											
# in 1st year	542	543	491	437	478	476	536	587	588	542	[0.827]
stay probability	1.018	0.855	0.994	1.078	0.912	1.053	1.076	1.019	0.927	0.910	0.981
1992-1993 arrivals											
# in 1st year			374	483	424	442	458	450	481	477	[1.028]
stay probability			0.976	0.948	1.087	1.068	1.020	1.096	0.977	0.876	1.003
1994-1995 arrivals											
# in 1st year					457	471	542	572	520	488	[1.153]
stay probability					1.011	1.064	1.068	1.038	0.981	0.986	1.024
1996-1997 arrivals											
# in 1st year							546	527	566	575	[0.829]
stay probability							0.987	1.004	0.952	0.878	0.954
1998-1999 arrivals											
# in 1st year									768	780	[0.785]
stay probability									0.944	0.832	0.886

in 1st (2nd) year: the number of foreign-born persons in the 1st (2nd) year

stay probability: the (unconditional) ratio between the numbers of foreign-born persons in the 2nd and in the 1st years

this value is less than or equal to unity in the population.

Table 3A. Mean Wage and Wage Growth Differential by Year & Origin: Reported & Imputed Wages

	Natives	Immigrants									
		All		C.S.America		Europe		Asia		Others	
	wage	wage	DD	wage	DD	wage	DD	wage	DD	wage	DD
1994	17.23	14.08		9.50		17.33		16.40		19.26	
–1995	17.24	14.58	0.49	9.45	–0.06	18.00	0.66	17.81	1.40	19.56	0.29
	[7355]	[828]	(0.63)	[348]	(0.94)	[148]	(1.44)	[213]	(1.21)	[119]	(1.63)
1995	16.10	13.83		10.67		19.39		15.61		17.36	
–1996	16.68	14.02	–0.39	10.31	–0.94	20.74	0.77	16.77	0.58	13.36	–4.58
	[5852]	[613]	(0.67)	[324]	(0.90)	[123]	(1.48)	[141]	(1.36)	[25]	(3.23)
1996	16.31	12.99		9.42		17.95		17.01		12.99	
–1997	17.38	13.52	–0.54	9.83	–0.66	17.83	–1.19	17.88	–0.20	15.17	1.11
	[17599]	[2128]	(0.38)	[1116]	(0.51)	[351]	(0.91)	[558]	(0.73)	[103]	(1.67)
1997	16.58	13.75		9.76		19.99		16.66		19.50	
–1998	16.02	13.20	0.01	10.04	0.84	17.26	–2.17	15.91	–0.19	18.87	–0.07
	[17568]	[2154]	(0.37)	[1111]	(0.49)	[368]	(0.87)	[611]	(0.67)	[64]	(2.06)
1998	15.72	13.14		10.01		18.41		15.65		16.86	
–1999	16.29	13.30	–0.41	10.18	–0.40	18.11	–0.87	16.34	0.12	15.14	–2.29
	[17773]	[2377]	(0.28)	[1277]	(0.36)	[415]	(0.64)	[602]	(0.53)	[83]	(1.42)
1999	16.26	13.34		10.39		19.51		16.43		12.12	
–2000	16.63	14.02	0.31	10.63	–0.13	20.63	0.75	17.77	0.97	13.08	0.59
	[17640]	[2357]	(0.28)	[1291]	(0.37)	[350]	(0.71)	[586]	(0.55)	[130]	(1.15)
2000	16.06	13.32		10.20		19.55		16.45		13.45	
–2001	16.72	13.54	–0.44	10.38	–0.48	18.78	–1.43	17.09	–0.02	14.73	0.62
	[17174]	[2722]	(0.26)	[1491]	(0.33)	[408]	(0.64)	[671]	(0.50)	[152]	(1.03)
2001	16.49	14.03		10.40		19.70		19.12		13.37	
–2002	16.95	14.38	–0.11	10.76	–0.10	20.65	0.49	19.00	–0.58	14.20	0.37
	[18187]	[2605]	(0.28)	[1433]	(0.36)	[363]	(0.73)	[638]	(0.55)	[171]	(1.04)
2002	17.09	13.50		10.49		19.16		17.13		13.38	
–2003	16.67	13.51	0.43	10.57	0.50	18.34	–0.40	17.50	0.79	13.46	0.50
	[19466]	[2939]	(0.27)	[1656]	(0.35)	[456]	(0.67)	[667]	(0.55)	[160]	(1.12)
2003	16.52	13.85		10.48		19.28		18.19		14.17	
–2004	16.89	14.26	0.04	10.80	–0.05	19.43	–0.22	18.87	0.31	14.99	0.45
	[17627]	[2731]	(0.18)	[1513]	(0.24)	[410]	(0.45)	[653]	(0.36)	[155]	(0.73)
Simple Average:			–0.06		–0.15		–0.36		0.32		–0.30

Mean Hourly Wages in 1994 dollar; Sample sizes are in square brackets and standard errors are in parentheses.

DD: difference in difference (difference between wage increments)

DD is the wage growth of an ethnic group relative to that of natives

C.S.America: Central and South America; Europe: Europe, Australia, New Zealand, and Canada;

Asia: Asia; Others: Africa, Oceania, and other countries

Table 3B. Mean Wage and Wage Growth Differential by Year & Origin: Reported Wages Only

	Natives	Immigrants									
		All		C.S.America		Europe		Asia		Others	
	wage	wage	DD	wage	DD	wage	DD	wage	DD	wage	DD
1994											
–1995	N/A	N/A		N/A		N/A		N/A		N/A	
1995	16.17	14.08		10.08		20.87		16.26		25.39	
–1996	17.09	14.45	–0.55	9.65	–1.35	20.15	–1.64	16.17	–1.01	15.34	–10.97
	[3880]	[362]	(0.71)	[199]	(0.92)	[76]	(1.53)	[77]	(1.48)	[10]	(4.12)
1996	16.53	13.15		8.99		20.06		17.33		15.36	
–1997	17.43	13.99	–0.06	9.52	–0.37	20.89	–0.07	18.63	0.40	17.28	1.02
	[11786]	[1277]	(0.40)	[691]	(0.52)	[198]	(0.97)	[331]	(0.76)	[57]	(1.80)
1997	16.58	13.59		9.36		20.66		17.96		15.90	
–1998	16.16	13.02	–0.15	9.66	0.72	17.85	–2.39	16.86	–0.68	15.91	0.43
	[11554]	[1306]	(0.42)	[720]	(0.54)	[205]	(1.03)	[347]	(0.79)	[34]	(2.47)
1998	15.70	13.01		9.64		19.75		15.60		20.09	
–1999	16.51	13.39	–0.43	9.97	–0.48	20.65	0.09	16.23	–0.18	16.59	–4.31
	[11056]	[1340]	(0.28)	[765]	(0.36)	[219]	(0.67)	[315]	(0.55)	[41]	(1.55)
1999	16.25	13.60		9.88		22.07		17.72		13.02	
–2000	16.80	14.03	–0.12	10.21	–0.22	22.10	–0.52	18.73	0.46	13.11	–0.46
	[10280]	[1316]	(0.26)	[762]	(0.34)	[191]	(0.68)	[302]	(0.53)	[61]	(1.18)
2000	16.38	13.19		9.78		21.29		16.94		14.10	
–2001	17.08	13.76	–0.13	10.05	–0.43	22.29	0.30	17.83	0.19	15.53	0.73
	[9587]	[1416]	(0.27)	[830]	(0.34)	[197]	(0.70)	[310]	(0.56)	[79]	(1.08)
2001	16.82	14.28		10.33		21.56		20.38		13.17	
–2002	17.42	14.60	–0.28	10.39	–0.54	22.88	0.72	20.88	–0.10	13.25	–0.52
	[10183]	[1346]	(0.32)	[781]	(0.40)	[182]	(0.84)	[302]	(0.65)	[81]	(1.23)
2002	17.52	13.42		10.20		19.43		18.12		12.92	
–2003	16.90	13.22	0.42	10.09	0.51	19.33	0.52	17.49	–0.01	13.41	1.11
	[10987]	[1496]	(0.34)	[876]	(0.43)	[207]	(0.89)	[343]	(0.69)	[70]	(1.51)
2003	16.93	14.01		9.97		20.97		19.46		14.07	
–2004	17.23	14.24	–0.07	10.30	0.03	20.58	–0.69	19.67	–0.09	15.22	0.85
	[9804]	[1427]	(0.16)	[814]	(0.21)	[214]	(0.39)	[330]	(0.32)	[59]	(0.74)
Simple Average:			–0.15		–0.24		–0.41		–0.11		–1.35

Mean Hourly Wages in 1994 dollar; Sample sizes are in square brackets and standard errors are in parentheses.

DD: difference in difference (difference between wage increments)

DD is the wage growth of an ethnic group relative to that of natives

C.S.America: Central and South America; Europe: Europe, Australia, New Zealand, and Canada;

Asia: Asia; Others: Africa, Oceania, and other countries

3.2.3 Lessons from Summary Statistics

The findings in Tables 1–3B motivate the present analysis. First, there are systematic differences between imputed wages and reported wages, especially in levels. As imputed wages comprises about 17-23% of the sample, the impact can be large. In estimation of economic assimilation, we employ the reported wage sample as our preferred sample. We provide results using the entire sample as well as using weights which is suggested by Bollinger and Hirsch (2006) as a robustness check. Second, the persons in the matched sample are a nonrandom subset of the cross-section sample. About 22-40% of the interviewees drop out of the sample in the second period. Nevertheless, the matched sample is useful in order to control for individual specific attributes. We develop an estimation strategy which accounts for both panel attrition and outmigration by assigning weights to persons in the matched sample. Third, the impact of outmigration is likely to be small in the first difference analysis because it is only 2.6% per year. Fourth, different ethnic groups experience assimilation differently.

4 Estimation of an Overlapping Rotating Panel Model

This part has two sections. Section 4.1 introduces the methodology to correct for panel attrition, which is based on Hirano, Imbens, Ridder, and Rubin (2001). The estimation strategy consists of two steps. In the first step, one estimates the panel attrition process and obtain the weights for individuals in the matched longitudinal sample. In the second step, one estimates the main model using the matched sample along with the weights. Section 4.2 develops an estimation strategy when panel attrition and outmigration (or "population attrition") occur at the same time. The estimation strategy consists of three steps. In the first step, one estimates the outmigration process and weight the second period cross-section. Next, one applies the two-step panel attrition correcting method introduced in Section 4.1.

4.1 Correcting for Panel Attrition

Assume that there is no outmigration. Consider a two-period panel data set where all the interviewees respond in the first period but some do not respond in the second period. Denote $D_S = 1$ when an individual is in the sample (or responds) in the second period and $D_S = 0$ when an individual is not in the sample (or does not respond) in the second period. Now it is possible to construct a balanced longitudinal sample by collecting all the individuals with $D_S = 1$: we call the sample the matched sample.

Suppose the model of interest is identified by a conditional moment restriction

$$E\left[m\left(y_1, y_2, x_1, x_2, \theta\right)|x_1, x_2\right] = 0, \quad \text{w.p.1}, \tag{4}$$

uniquely when $\theta = \theta_0$, where y is the endogenous variable, x is a vector of exogenous variables, θ is a parameter vector, $m(\cdot)$ is a known function, and the subscripts denote the period. We do not observe the joint distribution of (y_1, y_2, x_1, x_2) due to nonresponse. Instead we observe the joint distribution of the matched sample, $(y_1, y_2, x_1, x_2)|D_S = 1$. However,

$$E\left[m\left(y_1, y_2, x_1, x_2, \theta_0\right)|x_1, x_2\right] \neq E\left[m\left(y_1, y_2, x_1, x_2, \theta_0\right)|x_1, x_2, D_S = 1\right]. \tag{5}$$

Therefore, just using the matched sample will result in an inconsistent estimator of θ_0.

Now assume that in addition to the two period panel there is a representative cross-section available in the second period. This cross-section is called the refreshment sample. Hirano, Imbens, Ridder, and Rubin (2001) specify the attrition process using the second period cross-section allowing the attrition to depend on the endogeneous variables in the second period. This is a substantially more general attrition model than standard attrition correction which only allows dependence on the variables in the first period. An overlapping rotating panel data set, by construction, includes representative cross-sections for both periods. This is because a new short panel is activated from the target population in each period. The CPS MORG, for instance, includes representative cross-sections.

In order to specify the attrition process, assume that attrition is a function of u_1, u_2, and v, where u_1 and u_2 are vectors of time-varying variables in periods 1 and 2, respectively, and v is a vector of time invariant variables. For instance, u_1 (or u_2) is a vector of the endogeneous variable, y_1 (or y_2), and time-varying exogenous variables in x_1 (or x_2) and is v is a vector of time-invariant exogenous variables.[20] It is worth noticing that u_2 is observed because the second period cross-section is available. This fact is crucial to the method.

We need to know $f(u_1, u_2, v)$ to calculate the LHS of (5), but it is not observed when there is attrition. What we do know, however, is the joint density of non-attriting individuals, $f(u_1, u_2, v|D_S = 1)$, along with marginal densities, $f(u_1, v)$ and $f(u_2, v)$. Notice that $f(u_1, v)$ is known from the first period cross-section. Notice also that $f(u_2, v)$ is known from the second period cross-section. A fact that is crucial to the method is

$$f(u_1, u_2, v) = \frac{f(u_1, u_2, v|D_S = 1)\Pr(D_S = 1)}{\Pr(D_S = 1|u_1, u_2, v)}.$$

So if we come up with a candidate for $\Pr(D_S = 1|u_1, u_2, v)$, we can obtain a consistent estimator of θ_0.

Hirano et al. prove that $\Pr(D_S = 1|u_1, u_2, v)$ is nonparametrically just-identified up to a known link function, $g(\cdot)$, if it takes an additive non-ignorable form:

$$\Pr(D_S = 1|U_1 = u_1, U_2 = u_2, V = v) = g(k_0(v) + k_1(u_1, v) + k_2(u_2, v)),$$

where $k_\cdot(\cdot)$ are unknown functions with the normalization of $k_1(0, v) = k_2(0, v) = 0$ and the known link function $g(\cdot)$ is a bounded strictly increasing function such that $\lim_{r\to-\infty} g(r) = 0$ and $\lim_{r\to\infty} g(r) = 1$. Identification stems from the fact that we observe two marginal densities, $f(u_1, v)$ from the year-one cross-section and $f(u_2, v)$ from the year-two cross-section, and $f(u_1, v)$ and $f(u_2, v)$ obey

$$\begin{aligned} f(u_1, v) &= \int \frac{\Pr(D_S = 1)}{g(k_0(v) + k_1(u_1, v) + k_2(u_2, v))} f(u_1, u_2, v|D_S = 1)\, du_2, \\ f(u_2, v) &= \int \frac{\Pr(D_S = 1)}{g(k_0(v) + k_1(u_1, v) + k_2(u_2, v))} f(u_1, u_2, v|D_S = 1)\, du_1, \end{aligned} \tag{6}$$

for almost all (u_1, u_2, v).

In estimation of (6), the standard semiparametric methods cannot be applied because the attrition function is defined implicitly by nonlinear integral equations. Bhattacharya (2006) shows that the identification

[20] The attrition function does not have to be determined by the same variables in the main model (4). The variables in (u_1, u_2, v) may or may not include the variables in (y_1, y_2, x_1, x_2).

conditions in (6) can be transformed into conditional moment restrictions:

$$
\begin{aligned}
1 &= E\left[\frac{D_S}{g\left(k_0\left(v\right)+k_1\left(u_1,v\right)+k_2\left(u_2,v\right)\right)}|u_1,v\right] \quad \text{w.p.1,} \\
1 &= E\left[\frac{D_S}{g\left(k_0\left(v\right)+k_1\left(u_1,v\right)+k_2\left(u_2,v\right)\right)}|u_2,v\right] \quad \text{w.p.1.}
\end{aligned} \tag{7}
$$

The transformed identification conditions in (7) can be estimated, for instance, by the sieve minimum distance (SMD) developed by Ai and Chen (2003). When we specify a parametric attrition process, one can use the smoothed empirical log-likelihood (SEL) developed by Kitamura, Tripathi, and Ahn (2004). As $g\left(k_0\left(v\right)+k_1\left(u_1,v\right)+k_2\left(u_2,v\right)\right)$ and $\Pr\left(D_S=1\right)$ are estimable, we can construct the attrition correcting weighting function

$$
C\left(u_1,u_2,v\right)=\frac{\Pr\left(D_S=1\right)}{g\left(k_0\left(v\right)+k_1\left(u_1,v\right)+k_2\left(u_2,v\right)\right)}. \tag{8}
$$

Then, we weight the matched sample by (8) and estimate

$$
E\left[m\left(y_1,y_2,x_1,x_2,\theta_0\right)\cdot C\left(u_1,u_2,v\right)|x_1,x_2,D_S=1\right]=0, \quad \text{w.p.1,} \tag{9}
$$

to obtain a consistent estimator of θ_0. In sum, the model with attrition can be estimated consistently by taking the individuals in the matched sample and weighting them with corresponding attrition correcting weighting function values. The weighting function is proportional to the inverse of one minus the attrition process. Intuitively, the LHS of (7) is equivalent to weighting the sample with the inverse of one minus the probability of attrition, $1/g\left(k_0\left(v\right)+k_1\left(u_1,v\right)+k_2\left(u_2,v\right)\right)$.

The attrition correcting method has at least four attractive features. First, the panel attrition process for a longitudinal sample is identified nonparametrically under relatively weak conditions. In particular, it is identified provided the attrition function is additive non-ignorable with a known link function such as the logit or probit and representative cross-sections are available. The constraint of additive non-ignorable assumption reduces the dimension of the attrition function of our interest.[21] Second, different from the Heckman's self-selection model, no exclusion restriction is needed to estimate the attrition function. Heckman's solution requires at least one exogenous variable affecting selection that does not appear in the structural equation. The key to the approach used here is the availability of additional information than the self-selection setup. In consequence, there is no need of making assumptions on unobservables.

Third, the correction is robust to individual fixed effects. This is because each individual gets his unique weight which is a function of the characteristics in the first and second periods. Therefore, the usual fixed effects strategies for panel data models can be used to control individual heterogeneity. Fourth, the weighting function estimates do not have to be interpreted as causal effects. They simply describe the state. For instance, the wage may affect residential mobility, but the latter may affect the former, too. Therefore, even if there is reverse causality problem from mobility to wages, the weighting function estimates successfully describe the

[21] As an additive non-ignorable attrition model includes the first and the second period variables, it nests models of the selection on observables and the selection on unobservables. Therefore, two models can be distinguished by use of the second period cross-section. The data provides testable restrictions on those models. An additive non-ignorable model, however, rules out interactions between the variables in the first and the second periods. For instance, consider $wage_1$ and $wage_2$. Panel attrition can depend on $\log wage_2 - \log wage_1$ but not on $\left(wage_2 - wage_1\right)/wage_1$, although both measure wage growth. In the Appendix, models of the selection on observables and the selection on unobservables are introduced.

attrition process in a statistical sense.

4.2 Correcting for Panel Attrition in the presence of Outmigration

When the target population is nonstationary and the model of interest requires a counterfactual situation of what if the population had remained stationary, the attrition correcting technique has to be modified. Consider a pair of representative cross-section data sets where some of the interviewees drop out of the population in the second period. Denote $D_P = 1$ when an individual is in the population (or stays in the United States) in the second period and $D_P = 0$ when an individual is not in the population (or leaves the United States) in the second period. Now an individual is in the matched sample if $D_P = 1$ and $D_S = 1$. Similarly, an individual stays in the U.S. but does not respond in the second period if $D_P = 1$ and $D_S = 0$. An individual who leaves the U.S. in the second period is denoted by $D_P = 0$. A combination of $D_P = 0$ and $D_S = 1$, where an individual leaves the country and responds in the second period, is not possible. As a result, being in the matched sample, $D_S = 1$, also implies residing in the U.S. at the same time, $D_P \cdot D_S = 1$.

Again, the model of interest is identified by a conditional moment restriction (4). An available data set is a matched sample of $(y_1, y_2, x_1, x_2)\,|D_P \cdot D_S = 1$. Similar to (5), simply using the balanced part will lead to an inconsistent estimator. Specify the panel attrition model by

$$\Pr(D_S = 1|U_1 = u_1, U_2 = u_2, V = v) = g\left(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)\right),$$

where $k_.'(\cdot)$ and $g(\cdot)$ are defined as before. In the presence of outmigration, the LHS of the second identification condition in (6), $f(u_2, v)$, is unobservable: we observe $f(u_2, v|D_P = 1)$ instead from the second period cross-section. But, we know that

$$f(u_2, v) = \frac{f(u_2, v|D_P = 1)\Pr(D_P = 1)}{\Pr(D_P = 1|u_2, v)}.$$

Therefore, the identification condition becomes

$$\begin{aligned} f(u_1, v) &= \int \frac{\Pr(D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} f(u_1, u_2, v|D_S = 1)\,du_2, \\ \frac{f(u_2, v|D_P = 1)\Pr(D_P = 1)}{\Pr(D_P = 1|u_2, v)} &= \int \frac{\Pr(D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} f(u_1, u_2, v|D_S = 1)\,du_1, \end{aligned} \tag{10}$$

for almost all (u_1, u_2, v). So if we come up with a candidate for $\Pr(D_P = 1|u_2, v)$, the LHS of the second equation in (10) is known, and we can obtain a consistent estimator of θ_0 by the panel attrition correcting technique in the previous section.

Now a remaining question is how we specify $\Pr(D_P = 1|u_2, v)$. When $\Pr(D_P = 1|u_2, v)$ is a function of variables of known transition probability, it can be nonparametrically identified when representative cross-sections are available. Assume that the transition probability is given by $P(Z_2 = z_2|Z_1 = z_1)$, where z is a vector of variables of known transition probability.[22] For instance, if z is year of entry, the transition probability is given by $P(z_2|z_1) = 1(z_2 = z_1)$, where $1(\cdot)$ is the indicator function. If z is age, the transition probability is given by $P(z_2|z_1) = 1(z_2 = z_1 + 1)$. The assumption requires that the population attrition is solely determined by variables of known transition probability. Selection on variables of known transition

[22] The variables in z_2 must be included in (u_2, v).

probability implies that one minus the outmigration probability is given by

$$\begin{aligned} \Pr\left(D_P=1|u_2,v\right) &= \Pr\left(D_P=1|z_2\right) \\ &\equiv k\left(z_2\right), \end{aligned} \tag{11}$$

where $k(\cdot)$ is some unknown function. The outmigration process, $k(z_2)$, is nonparametrically identified from

$$\begin{aligned} k\left(z_2\right) &= \frac{f\left(z_2|D_P=1\right)\Pr\left(D_P=1\right)}{f_2\left(z_2\right)} \\ &= \frac{f\left(z_2|D_P=1\right)\Pr\left(D_P=1\right)}{\int f_1\left(z_1\right)f\left(z_2|z_1\right)dz_1}. \end{aligned}$$

The key estimation strategy can be described in two steps. First, the second period cross-section with outmigration is identical to the counterfactual second period distribution where there is no outmigration adjusted by $k(z_2)$. Second, counterfactual second period distribution where there is no outmigration is available from the first period cross-section and the known transition probability. These steps can be written in the following way. For simplicity, assume that $k(z_2)$ is given by a parametric form, $k(z_2'\psi)$. Consider the following:

$$\begin{aligned} \Pr\left(D_P=1\right)E_{Z_2}\left[Z_2|D_P=1\right] &= E_{Z_2}\left[k\left(Z_2'\psi\right)Z_2\right] \\ &= E_{Z_1}\left[\int k\left(z'\psi\right)zP\left(dz|Z_1\right)\right], \end{aligned}$$

which is implied by the two steps. Now, we can apply a GMM type estimation by employing the sample analog of these equations.[23] The LHS is the average over the variables in the second period population (after outmigration has taken place) adjusted by the probability of outmigration. The RHS is the average over the variables in the first period population (prior to outmigration) transformed into the second period variables by the transition probability. Therefore, the sample analog is given by

$$\begin{aligned} \frac{1}{n_2}\Pr\left(D_P=1\right)\textstyle\sum_{j=1}^{n_2}z_{2j} &= \frac{1}{n_1}\textstyle\sum_{i=1}^{n_1}\left[\int k\left(z'\psi\right)zP\left(dz|z_{1i}\right)\right] \\ &= \frac{1}{n_1}\textstyle\sum_{i=1}^{n_1}\sum_{z\in S_2}k\left(z'\psi\right)z\Pr\left(z|z_{1i}\right), \end{aligned}$$

where n_1 and n_2 are the sample sizes of the first and the second period cross-sections. The second equation holds if z is a vector of discrete variables, where S_2 is the support of Z_2. In the Appendix, we illustrate the estimation strategy in the analysis.

Again in the Appendix, we show that the identification conditions in (10) under assumption (11) can be transformed into conditional moment restrictions given by

$$\begin{aligned} 1 &= E\left[\frac{D_S}{g\left(k_0'\left(v\right)+k_1'\left(u_1,v\right)+k_2'\left(u_2,v\right)\right)}|u_1,v\right] \quad \text{w.p.1}, \\ \frac{1}{k\left(z_2\right)} &= E\left[\frac{D_S}{g\left(k_0'\left(v\right)+k_1'\left(u_1,v\right)+k_2'\left(u_2,v\right)\right)}|u_2,v,D_P=1\right] \quad \text{w.p.1}. \end{aligned} \tag{12}$$

[23] Technically, this method is similar to the method developed by Guell and Hu (2006). Both methods require cross-sections for two periods and use individual level information, but their method only allows time-invariant variables to enter the process. The two methods are developed for conceptually different purposes. Our method targets the attrition in the population or the duration of staying in the United States, whereas their method focuses on the duration of unemployment.

In sum, once the attrition-outmigration correcting weighting function

$$C(u_1, u_2, v) = \frac{\Pr(D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} \tag{13}$$

is estimated, we weight the matched sample by (13) and estimate (9) to obtain a consistent estimator of θ_0. Intuitively, the RHS in the second period is equivalent to weighting the individuals in the population (or more precisely the cross-section) with the inverse of one minus the probability of outmigration, $1/k(z_2)$, and the LHS of (12) is equivalent to weighting the individuals in the matched sample with the inverse of one minus the probability of panel attrition, $1/g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))$.

In practice, the vector z_t includes age, years since migration, education (assuming that no additional schooling is obtained), country of origin, and year of entry. These variables have deterministic time paths and satisfy the known transition probability assumption. The assumption, however, is more restrictive than the model of the selection on observables, for instance, because observable variables with unknown transition probability, such as the wage, cannot enter in the selection function. The assumption can be problematic as the transition probabilities of labor market performance variables are usually not known. Intuitively labor market performance will affect outmigration decision. If the assumption is indeed a serious problem in practice, it is required to develop an alternative way of handling outmigration.

Despite its limitation, the attrition-outmigration correcting method has at least four advantages. First, the outmigration process is identified nonparametrically under selection on variables of known transition probability when repeated cross-sections are available. It allows stochastic transition, which is given by transition probability, so it is more flexible than assuming a deterministic mapping from one period to the other. Second, the method identifies the panel attrition and the outmigration processes separately. This is a very useful result because we are using a data set which does not provide information on who left the country. Notice that Heckman's self-selection correction cannot be applied, as it is not possible to distinguish those who migrate internally and who outmigrate. Third, the method is robust against fixed effects. Finally, the weighting function estimates need not have a causal interpretation.

5 Empirical Evidence of Economic Assimilation

This part provides empirical specifications and the identification conditions. Then it reports empirical findings. As our results are qualitatively different from the findings in the previous studies that use repeated cross-sections, we explore why they are different. Finally, this section presents assimilation estimates by ethnic origin.

5.1 Identification of Empirical Specifications

Based on the model given in (1) and (2), an empirical specification is given by

$$y_{it} = (\alpha_{nat} + \alpha_1)\, age_{it} + \delta_1 ysm_{it} + (\beta_{nat} + \beta)\, edu_i + \mu_i + \gamma_{imm,t} + \varepsilon_{it}, \tag{14}$$

and

$$y_{it} = \alpha_{nat} age_{it} + \beta_{nat} edu_i + \mu_i + \gamma_{nat,t} + \varepsilon_{it}, \tag{15}$$

where μ_i involves ability or skill endowment, and γ_t reflects business cycles, and ε captures idiosyncratic shocks.[24] We name the model given in (14) and (15) the individual heterogeneity (IH) model as it allows fixed unobserved heterogeneity such as variation in skill endowments within the groups of individuals of same years of entry. Empirical findings of this paper suggest positive correlation between ability and age at migration. Estimation of the IH model requires longitudinal sample.

Another empirical specification is given by

$$y_{it} = (\alpha_{nat} + \alpha_1)\, age_{it} + \delta_1 ysm_{it} + (\beta_{nat} + \beta)\, edu_i + \mu_c + \lambda_a am_i + bc_i' \lambda_b + \gamma_{imm,t} + \varepsilon_{it}, \tag{16}$$

and

$$y_{it} = \alpha_{nat} age_{it} + \beta_{nat} edu_i + \gamma_{nat,t} + \varepsilon_{it}, \tag{17}$$

where μ_c is arrival year cohort effects, am is the age at migration, bc is a vector of birth country indicators. In this specification, year of entry, age at entry, and country of origin control for fixed unobserved heterogeneity. As the individual heterogeneity within an immigration year cell is neglected, we name the model given in (16) and (17) the cohort heterogeneity (CH) model. Estimation of the CH model requires repeated cross-sections.

The empirical measure of economic assimilation in (3) for both the IH and the CH models is given by

$$\begin{aligned} EA\,(age, ysm; t) &\approx \left(\alpha_{nat} + \alpha_1 + \delta_1 + \gamma_{imm,t+\triangle t} - \gamma_{imm,t}\right) - \left(\alpha_{nat} + \gamma_{nat,t+\triangle t} - \gamma_{nat,t}\right) \\ &= \alpha_1 + \delta_1 + \left(\gamma_{imm,t+\triangle t} - \gamma_{imm,t}\right) - \left(\gamma_{nat,t+\triangle t} - \gamma_{nat,t}\right). \end{aligned} \tag{18}$$

We address three issues regarding the identification of (18). First, we assume that the vector of the coefficients for the calendar year dummy variables common to foreign-born and native-born workers: $\gamma_{imm,t} = \gamma_{nat,t} = \gamma_t$ for all t. This restriction is proposed by Borjas (1985). With the restriction,

$$EA\,(age, ysm; t) = \alpha_1 + \delta_1.$$

It is crucial for identification as the number of years since migration, the arrival year, and the calendar year are perfectly correlated. So we need some restrictions for identification, although different restrictions lead to different estimates of the underlying parameters of interest. Under the common calendar year effects identification restrictions, aggregate economic shocks affect the wages by the same percentage amount to foreign-born and native-born workers.[25]

Second, the IH model is estimated by taking the first difference:

$$\begin{aligned} \Delta y_{it} &= \alpha_{nat} + \alpha_1 + \delta_1 + \Delta\gamma_t + \Delta\varepsilon_{it}, \\ \Delta y_{it} &= \alpha_{nat} + \Delta\gamma_t + \Delta\varepsilon_{it}. \end{aligned}$$

Therefore, $\alpha_1 + \delta_1$ is identified by $(\alpha_{nat} + \alpha_1 + \delta_1 + \Delta\gamma_t) - (\alpha_{nat} + \Delta\gamma_t)$. Notice that $\alpha_{nat} + \alpha_1 + \delta_1$ and α_{nat} are not identified due to $\Delta\gamma_t$. The assumption of common γ plays a key role in identification of $\alpha_1 + \delta_1$.

The third issue is the age at migration, am, in the CH model. The coefficients for age, years since migration,

[24] A more general model is estimated in this analysis allowing for nonlinearities in the age and the number of years since migration. These generalizations do not affect the discussion of identification issues.

[25] As Baker and Benjamin (1997) point out, the shocks will not be common, if immigrant and native workers differ in their sensitivity to the business cycle.

and age at migration are not separately identified: the three variables perfectly correlated, $am = age - ysm$. To identify these coefficients, Borjas (1995a), for instance, restrict the age coefficient for immigrants and natives to be common. However, we claim that restricting age coefficients does more harm than good if the parameter of interest is $\alpha_1 + \delta_1$. Consider a CH model dropping age at migration. Now we have the omitted variable bias: the probability limits of the coefficients for the age and the years since migration are $\alpha_1 + \lambda_a$ and $\delta_1 - \lambda_a$, respectively. In consequence, α_1 and δ_1 are not identified. It is worth noticing, however, that even if there is an omitted variable, the sum of the coefficients, $\alpha_1 + \delta_1$, is identified because $\alpha_1 + \delta_1 = (\alpha_1 + \lambda_1) + (\delta_1 - \lambda_1)$.[26]

5.2 Estimates of Economic Assimilation

Table 4 reports the estimated measures of economic assimilation for the IH and the CH models. The economic assimilation estimates are based on the wage equation estimates in Appendix tables. The wage equations are specified by linear, quadratic, and cubic polynomials in age and years since migration. We estimate these equations by (1) the attrition correcting method and (2) the attrition-outmigration correcting method as well as (3) without adjustment.[27] These results are reported in Table A3 of the Appendix. We assume that a representative foreign-born worker arrives in the United States at age 20 as many other studies do. It is a reasonable assumption as the mean age is about 40 and the mean years since migration is about 20 in our analysis.

In the upper panel of Table 4, we use the IH models and find that the estimates are mostly negative.[28] As the mean wage of foreign-born workers is below the native mean, we conclude that there is little evidence of economic assimilation. For instance, the linear specification using the sample of reported wages gives insignificant negative values: –0.22 (attrition-adjusted), –0.25 (attrition-outmigration-adjusted), and –0.18 (unadjusted). An estimate of –0.22 implies that the wage growth of a foreign-born worker is slower than that of a native-born worker by 0.22% points per year. The attrition-outmigration-adjusted estimates from the quadratic specification suggest that the mean wage of a foreign-born worker grows significantly slower than that of a native-born worker by 1.17% points at age 24 and by 0.75% points at age 32. From the cubic specification, we find that the mean wage of foreign-born workers grows slower than that of native-born workers by 1.49% points at age 24 and by 0.55% points at age 32. The nonlinear specification results reveal that young foreign-born workers fall behind rather than catch up.

In general, the attrition-adjusted estimates are smaller than the unadjusted ones. The wage equation results reveal that the wage growth rates of natives are more affected by the attrition correcting method than those of immigrants. More specifically, among the natives the slower the wage growth, the higher the residential mobility. But this correlation is not as strong among immigrants. Therefore, the assimilation measure is

[26] We may interact the cohort fixed effects with age at migration, $\mu_{c,am}$, but perfect correlation is still a problem. The point is no need of further restrictions such as common age coefficients for immigrants and natives. Later we show that $\mu_{c,am}$ eliminates the bias in estimation of economic assimilation.

[27] The main (wage) equations use the matched longitudinal sample of workers with positive wages. In this step, we exclude individuals with too high or too low wages and negative potential experience. In estimation of the matching functions, we use the matched longitudinal sample of individuals and cross-sections of all individuals including those not working, but we exclude extreme wage observations. Not-working individuals are included in this step in order to reflect market level changes, such as in the composition of natives, between consequtive years. In estimation of outmigration, we use the (unbalanced) panel of all individuals including extreme wage observations. In estimation of the outmigration process, labor market outcomes are not used as the variables must have known transition probabilities. To ensure enough sample size of the foreign sample, we keep the largest available sample.

[28] To be precise, one-sided test should be used instead of a two-sided test, as the alternative hypothesis is given by either $EA\,(age, ysm) > 0$ or $EA\,(age, ysm) < 0$.

smaller when we correct for panel attrition. The impact of outmigration seems negligible as attrition-adjusted and attrition-outmigration-adjusted estimates are similar. There are two possible reasons. One is that the outmigration is not large between years and the other is that the outmigration correcting method does not fully correct for outmigration.

Our findings are strikingly different from the results in the previous literature. For instance, using the 1970, 1980, and 1990 Census cross-sections, Borjas (1999) reports that the relative wage growth of immigrants is 0.60-0.76% points higher per year during the first 10 years and 0.38-0.50% points higher per year during the first 20 years based on CH models. To replicate the CH models, we drop the second period observations from the longitudinal samples and construct cross-section data. The estimates are reported in the lower panel of Table 4. Now, the empirical findings from the level models suggest significant economic assimilation. For instance, the linear specification using reported wages gives significant positive values: 0.99 (attrition-adjusted and attrition-outmigration-adjusted) and 0.95 (unadjusted). The quadratic specification suggests that the mean wage of a foreign-born worker grows significantly faster than that of a native-born worker by 0.91-1.05% points at age 24, by 0.73-0.83% points at age 32, and by 0.56-0.60% points at age 40. Using the cubic specification, we find significant convergence in mid-ages, 32-40.

5.3 Analyzing the Difference between IH and CH Estimates

Why are the estimates from the IH and the CH models different? A possible explanation is that the cohort fixed effects fail to fully reflect individual heterogeneity. In order to see this point, it is useful to discuss how IH and CH models control for heterogeneity. Assume that the true data generating process is given by

$$\begin{aligned} y_{it} &= \alpha_1 age_{it} + \delta_1 ysm_{it} + \beta edu_i + \mu_i + \varepsilon_{it} \\ &= \alpha_1 age_{it} + \delta_1 (t - c) + \beta edu_i + \mu_i + \varepsilon_{it}, \end{aligned} \tag{19}$$

for an individual i in an arrival year c. We drop year fixed effects as they are identified from the native equation. For the time being, assume that there is no panel attrition nor outmigration. Further, assume that all immigrants are from a common source country.

First of all, single cross-section analyses fail to identify $\alpha_1+\delta_1$ when the skill composition of new immigrants change over time. Assume that the individual fixed effects, μ_i, in (19) can be replaced with the arrival year cohort fixed effects, μ_c. Then we have

$$\begin{aligned} E[y_{it}|c_1, t, age_{it}, edu_i] &= \alpha_1 age_{it} + \delta_1 (t - c_1) + \beta edu_i + \mu_{c1}, \\ E[y_{jt}|c_2, t, age_{jt}, edu_j] &= \alpha_1 age_{jt} + \delta_1 (t - c_2) + \beta edu_j + \mu_{c2}. \end{aligned}$$

In this case $\alpha_1 + \delta_1$ is not identified unless $\mu_{c1} = \mu_{c2}$, which is exactly the same argument made by Borjas (1985).

Table 4. Economic Assimilation Estimates in %

	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	linear	quadra.	cubic	linear	quadra.	cubic	linear	quadra.	cubic
Individual Hetero.									
All Wages									
age=24, ysm=4	–0.03	–1.23**	–1.23*	–0.06	–1.33**	–1.32*	0.20	–0.96	–0.91
	(0.34)	(0.59)	(0.73)	(0.34)	(0.59)	(0.73)	(0.34)	(0.59)	(0.73)
age=32, ysm=12		–0.68*	–0.58		–0.73*	–0.64		–0.50	–0.49
		(0.38)	(0.43)		(0.38)	(0.43)		(0.39)	(0.43)
age=40, ysm=20		–0.13	–0.05		–0.14	–0.07		–0.04	–0.10
		(0.36)	(0.53)		(0.36)	(0.53)		(0.35)	(0.54)
age=48, ysm=28		0.42	0.37		0.45	0.39		0.42	0.27
		(0.53)	(0.58)		(0.53)	(0.58)		(0.54)	(0.58)
Reported Wages									
age=24, ysm=4	–0.22	–1.13**	–1.46**	–0.25	–1.17**	–1.49**	–0.18	–1.15**	–1.44**
	(0.31)	(0.55)	(0.68)	(0.31)	(0.55)	(0.68)	(0.30)	(0.54)	(0.68)
age=32, ysm=12		–0.72**	–0.52		–0.75**	–0.55		–0.78**	–0.70*
		(0.36)	(0.38)		(0.35)	(0.39)		(0.35)	(0.38)
age=40, ysm=20		–0.31	0.08		–0.33	0.05		–0.40	–0.16
		(0.32)	(0.47)		(0.32)	(0.47)		(0.32)	(0.47)
age=48, ysm=28		0.10	0.35		0.08	0.33		–0.03	0.18
		(0.48)	(0.53)		(0.48)	(0.53)		(0.47)	(0.52)
Cohort Hetero.									
All Wages									
age=24, ysm=4	0.82***	1.02***	1.30**	0.81***	1.01***	1.30**	0.94***	1.32***	1.71***
	(0.15)	(0.25)	(0.38)	(0.15)	(0.25)	(0.37)	(0.15)	(0.27)	(0.39)
age=32, ysm=12		0.70***	0.67***		0.70***	0.66***		0.97***	0.96***
		(0.17)	(0.17)		(0.17)	(0.17)		(0.17)	(0.17)
age=40, ysm=20		0.39***	0.24		0.38***	0.23		0.63***	0.45**
		(0.15)	(0.19)		(0.14)	(0.19)		(0.15)	(0.19)
age=48, ysm=28		0.07	0.02		0.07	0.02		0.29	0.18
		(0.20)	(0.22)		(0.20)	(0.22)		(0.21)	(0.22)
Reported Wages									
age=24, ysm=4	0.99***	0.91**	0.68	0.99***	0.93**	0.70	0.95***	1.05***	0.77
	(0.21)	(0.36)	(0.52)	(0.21)	(0.36)	(0.52)	(0.21)	(0.37)	(0.55)
age=32, ysm=12		0.73***	0.69***		0.74***	0.69***		0.83***	0.76***
		(0.24)	(0.24)		(0.24)	(0.24)		(0.25)	(0.25)
age=40, ysm=20		0.56***	0.65**		0.56***	0.64**		0.60***	0.69**
		(0.21)	(0.27)		(0.21)	(0.27)		(0.21)	(0.27)
age=48, ysm=28		0.38	0.57*		0.37	0.56*		0.37	0.56
		(0.30)	(0.33)		(0.30)	(0.33)		(0.30)	(0.33)

Standard errors are reported in parentheses. Confidence levels: 99% (***), 95% (**), 90% (*).

ATT-Adjusted: Attrition-Adjusted; ATT-OUT-Adjusted: Attrition-Outmigration-Adjusted

Standard errors for adjusted estimates do not account for sampling error in weighting functions estimation.

Estimates represent foreign-born workers' annual percentage wage growth relative to the natives' percentage wage growth.

Second, assume that repeated cross-sections are available: we have 2 periods, and i and j are in the same arrival year cohort but in different cross-sections. From (19), we have

$$\begin{aligned} E\left[y_{it}|c,t,age_{it},edu_i\right] &= \alpha_1 age_{it} + \delta_1\left(t-c\right) + \beta edu_i + E\left[\mu_i|c,t,age_{it},edu_i\right], \\ E\left[y_{jt'}|c,t',age_{jt'},edu_j\right] &= \alpha_1 age_{jt'} + \delta_1\left(t'-c\right) + \beta edu_j + E\left[\mu_j|c,t',age_{jt'},edu_j\right], \end{aligned}$$

where $t'=t+1$. Now $\alpha_1+\delta_1$ is identified under, for instance, the cohort heterogeneity assumption

$$E\left[\mu_i|c,t,age_{it},edu_i\right]=\mu_c \quad \text{w.p.1 for all } i\in c \text{ and for } t,t'.$$

This identification restriction is employed in almost all the repeated cross-section studies. The constraint, however, is not likely to hold if age at migration, $age_{it}-(t-c)$, is correlated with ability conditional on the year of entry and other observable variables. For instance, suppose the correlation between ability and age at migration is given by

$$E\left[\mu_i|c,t,age_{it},edu_i\right]=\mu_c+\eta_a\left(age_{it}-(t-c)\right). \tag{20}$$

Under (19), assume that we follow a group of persons of the same age at migration:

$$age_{it}-(t-c)=age_{jt'}-\left(t'-c\right).$$

Then, we have

$$\begin{aligned} E\left[y_{it}|c,t,age_{it},edu_i\right] &= \alpha_1 age_{it} + \delta_1\left(t-c\right) + \beta edu_i + \left[\mu_c+\eta_a\left(age_{it}-(t-c)\right)\right], \\ E\left[y_{jt'}|c,t',age_{jt'},edu_j\right] &= \alpha_1 age_{jt'} + \delta_1\left(t'-c\right) + \beta edu_j + \left[\mu_c+\eta_a\left(age_{jt'}-\left(t'-c\right)\right)\right]. \end{aligned}$$

Therefore, $\alpha_1+\delta_1$ is identified.

We make two points. If ability and age at migration are correlated, the cohort heterogeneity assumption made in most repeated cross-section studies leads biased estimates. Also, the true correlation structure is likely to be much more complicated than (20). If this is the case, it is safer to use a longitudinal analysis than a repeated cross-section approach. When a longitudinal sample is available, $\alpha_1+\delta_1$ is identified by the difference of the following equations:

$$\begin{aligned} E\left[y_{it}|c,t,age_{it},edu_i,\mu_i\right] &= \alpha_1 age_{it} + \delta_1\left(t-c\right) + \beta edu_i + \mu_i, \\ E\left[y_{it'}|c,t',age_{it'},edu_i,\mu_i\right] &= \alpha_1 age_{it'} + \delta_1\left(t'-c\right) + \beta edu_i + \mu_i. \end{aligned}$$

The advantage of using longitudinal sample is that no additional identifying restrictions for the individual fixed effects are needed.

In consequence, the empirical findings in Table 4 indicate a positive correlation between ability and age at migration conditional on age, year of entry, and other observables. It implies that among the immigrants of the same year of entry, older persons are more skilled or motivated than younger ones. It does not imply, however, that the existence of convergence in the previous literature is completely wrong. Recall that our results are based on much recent sample. Other papers applying IH type models for earlier periods report existence of economic assimilation. For instance, Lubotsky (2000) uses time series linked to cross-section data

1951-1997 and accounts for possible measurement errors in years since migration. He finds that the earnings of immigrants have grown 0.50-0.65% points per year during the first twenty years relative to the earnings of native-born workers of similar characteristics. Therefore, the implication of our findings is that assimilation estimates based on CH models appear to be biased upward for 1994-2004.

5.4 Economic Assimilation by Ethnic Origins

Given that there is little evidence of economic assimilation in general for 1994-2004, a natural and interesting question is whether some ethnic groups do assimilate economically while others do not. Table 5 reports estimates of economic assimilation using reported wages by ethnic origin. In this stage, we use the previously calculated weights instead of estimating them from each ethnic group.

The first panel presents economic assimilation of immigrants from Central and South America. From the attrition-outmigration-adjusted estimates of nonlinear specifications, we learn that they annually lose 1.41-2.23% points relative to the native-born workers at age 24 and 0.39-0.76% points at age 32. As they become more experienced, there is no significant difference in relative wage growth compared with native-born workers. Immigrants from Europe and Asia insignificant assimilation measure estimates. It implies that their higher mean wage do not deviate from the mean wage of native-born workers. As European and Asian immigrant workers tend to have higher wage levels than native-born workers, we conclude that empirical findings do not strongly support the hypothesis of economic assimilation.

As a robustness check, Table A2 in the Appendix provides assimilation estimates using different samples and methods. The table reports estimates using all the individuals. In addition, following Bollinger and Hirsch (2006), it reports estimates when individuals with reported wages are weighted by the inverse probability of reporting wages. The weights correct for nonrandom selection of not reporting wages and are obtained from linear index logit models by country of origin, using age, years since migration, education, citizenship status, and marital status. The table also reports estimates when we drop foreign-born persons who immigrated before age 18. Dropping these persons significantly diminishes the sample sizes, but the results are qualitatively the same.

A caveat is that the findings of economic assimilation might be sensitive to the choice of the definition of economic assimilation. Especially for immigrants from Central and South America, perhaps we would obtain different results if we compare them to the natives within their ethnic groups using the CPS MORG. Immigrants from Europe and Asia are of little interest because they perform better than natives. This is an important point as there exists difference in economic performance among natives by country of ancestry and race, but it is difficult to precisely measure assimilation. Borjas (1995a) finds that Mexicans fail to catch up to natives of similar ancestry. However, he does not make a strong argument out of it because the composition of non-white natives has changed over time. In addition to this, we address the importance of measurement errors in the survey of ethnic origin among natives. If there is a systematic pattern of misreporting ancestry, the measurement errors pose another difficulty in the approach of comparison with natives by country of ancestry and race.

Table 5. Economic Assimilation Estimates in % (by Origin): Reported Wages Only

Individual Hetero.	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
C.S.America									
age=24, ysm=4	0.11	–1.40**	–2.22***	0.10	–1.41**	–2.23***	0.12	–1.33**	–2.36**
	(0.37)	(0.64)	(0.78)	(0.37)	(0.64)	(0.78)	(0.37)	(0.63)	(0.77)
age=32, ysm=12		–0.74*	–0.38		–0.76*	–0.39		–0.82**	–0.57
		(0.41)	(0.48)		(0.41)	(0.47)		(0.41)	(0.46)
age=40, ysm=20		–0.09	0.68		–0.11	0.66		–0.31	0.44
		(0.41)	(0.59)		(0.41)	(0.59)		(0.41)	(0.58)
age=48, ysm=28		0.57	0.96		0.55	0.94		0.20	0.66
		(0.64)	(0.66)		(0.64)	(0.66)		(0.63)	(0.66)
Europe									
age=24, ysm=4	–1.17	–0.88	1.86	–1.18	–0.96	1.80	–1.09	–1.16	2.54
	(0.86)	(1.74)	(2.49)	(0.86)	(1.74)	(2.49)	(0.84)	(1.77)	(2.63)
age=32, ysm=12		–0.79	–1.14		–0.85	–1.21		–0.95	–1.00
		(1.20)	(1.20)		(1.20)	(1.19)		(1.23)	(1.23)
age=40, ysm=20		–0.71	–2.59**		–0.73	–2.64**		–0.74	–2.68**
		(0.85)	(1.29)		(0.86)	(1.29)		(0.87)	(1.23)
age=48, ysm=28		–0.62	–2.48*		–0.62	–2.50*		–0.54	–2.49
		(0.93)	(1.36)		(0.94)	(1.35)		(0.91)	(1.29)*
Asia									
age=24, ysm=4	–0.48	–0.76	–0.17	–0.51	–0.84	–0.27	–0.36	–1.12	–0.51
	(0.64)	(1.37)	(1.83)	(0.64)	(1.37)	(1.84)	(0.62)	(1.30)	(1.72)
age=32, ysm=12		–0.46	–0.32		–0.52	–0.38		–0.60	–0.47
		(0.82)	(0.87)		(0.82)	(0.87)		(0.79)	(0.85)
age=40, ysm=20		–0.17	–0.27		–0.19	–0.29		–0.08	–0.19
		(0.76)	(1.05)		(0.76)	(1.05)		(0.75)	(1.04)
age=48, ysm=28		0.12	–0.02		0.13	–0.00		0.45	0.32
		(1.28)	(1.31)		(1.27)	(1.30)		(1.24)	(1.25)
Others									
age=24, ysm=4	–0.50	–1.30	–2.92	–0.66	–0.58	–3.18	–0.04	–0.11	–1.73
	(1.72)	(3.15)	(3.90)	(1.75)	(3.26)	(4.05)	(1.61)	(2.92)	(3.70)
age=32, ysm=12		0.25	1.62		–0.03	1.35		0.34	0.98
		(1.96)	(2.22)		(2.03)	(2.28)		(1.83)	(2.00)
age=40, ysm=20		0.80	3.21		0.52	2.93		0.79	2.17
		(1.88)	(2.65)		(1.94)	(2.68)		(1.88)	(2.55)
age=48, ysm=28		1.35	1.86		1.07	1.57		1.25	1.83
		(2.99)	(3.01)		(3.08)	(3.10)		(3.02)	(3.07)

Standard errors are reported in parentheses. Confidence levels: 99% (***), 95% (**), 90% (*).

Sample sizes: Native (89117), C.S.America (6438), Europe (1689), Asia (2657), Others (492)

Estimates represent immigrants' annual percentage wage growth relative to the natives' percentage wage growth.

6 Conclusion and Further Research Agenda

This study reexamines the evidence of wage convergence of immigrants using a novel research design. The existing literature on immigrant wage convergence suffers from a lack of representative longitudinal data on immigrant population with sufficient sample size. Longitudinal data on native-born and foreign-born populations, by tracking specific individuals over time, offers the huge advantage of permitting one to control for fixed unobserved heterogeneity. However, the use of panel data gives rise to an additional problem: nonrandom panel attrition. In addition, outmigration of immigrants that is related to wage growth poses another attrition problem for panel data analyses as well as for single cross-section and repeated cross-section analyses.

We address the sample size problem by using Merged Outgoing Rotation Groups (MORG) of the Current Population Survey (CPS) which forms an overlapping rotating panel data set. To address the problems of panel attrition as well as outmigration, we develop an estimation procedure for use with overlapping rotating panel data which accounts for both problems. The key estimation strategy is to use the availability of representative cross-sections as the basis for weighting the persons in the matched sample. In order to obtain the weights, identification of panel attrition and outmigration is necessary. Panel attrition can depend on both exogenous and endogenous variables. Outmigration can be identified without knowing who emigrated from the United States.

The empirical findings suggest little evidence of economic assimilation. The growth rate of hourly wages of immigrants from Central and South America at age 24 is 1.41-2.23% slower than that of native-born workers. At age 32 the gap in growth rates is between 0.39-0.76% points. New immigrants from Central and South America earn lower wages than natives, and this gap widens with time in the U.S. labor market. Foreign-born workers from Europe and Asia earn higher wages than native-born workers but there is no strong evidence of convergence. These results are qualitatively different from the findings in the previous studies that use repeated cross-sections. Our results suggest that analyses of immigrant wage growth based on repeated cross-section studies may be biased upward by individual heterogeneity. We find that the ability or skill endowment of individual workers is positively correlated with the age at migration. Controlling for this heterogeneity reverses the conventional result of economic assimilation.

A large research agenda remains. The first one is a more extensive analysis of how both the repeated cross-sections and first difference estimates are affected by changes in labor force participation rates over time. Changes in labor force participation and market condition over time mean that the population for whom wages are available changes systematically. Second, this paper looks at foreign-born men, but foreign-born women are also of interest. Baker and Benjamin (1997), for instance, study the role of the family in immigrants' labor market activity. They support a family investment model, in which wives take on low skilled jobs to finance their husbands' investments in human capital. If this is the case, we need to model men and women at the same time and include labor market status and hours worked. Finally, the method developed in this paper can be applied to quantile analyses. A related study is conducted by Butcher and DiNardo (2002). They divide the native sample into deciles and assign immigrants to the decile cells. Following the cells over time, one can observe the changes in the cell sizes and see how immigrants fare as they stay longer in the United States. More generally, we are interested not only in the mean but of the entire earnings distributions of immigrants and natives.

7 References

Ai, Chunrong and Xiaohong Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions containing Unknown Functions," *Econometrica*, 71 (6), 1795-1843.

Baker, Michael and Dwayne Benjamin (1997): "The Role of the Family in Immigrants' Labor-Market Activity: An Evaluation of Alternative Explanations," *American Economic Review*, 87 (4), 705-727.

Bhattacharya, Debopam (2006): "Inference in Panel Data Models under Attrition Caused by Unobservables," Working Paper.

Bollinger, Christopher R. and Barry T. Hirsch (2006): "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching," *Journal of Labor Economics*, 24 (3), 483-519.

Borjas, George J. (1985): "Assimilation, Changes in Cohort Quality and the Earnings of Immigrants," *Journal of Labor Economics*, 3, 463-489.

Borjas, George J. (1989): "Immigrant and Emigrant Earnings: a Longitudinal Study," *Economics Inquiry*, 27 (1), 21-37.

Borjas, George J. (1995a): "Assimilation and Changes in Cohort Quality Revisited: What happened to Immigrant Earnings in the 1980s?" *Journal of Labor Economics*, 13 (2), 201-245.

Borjas, George J. (1995b): "The Economic Benefits from Immigration," *Journal of Economic Perspectives*, 9 (2), 3-22.

Borjas, George J. (1999): "The Economic Analysis of Immigration," in Ashenfelter, Orley C. and David Card, eds., *Handbook of Labor Economics*, Vol 2A, Ch28.

Borjas, George J. and Bernt Bratsberg (1996): "Who Leaves? The Outmigration of the Foreign-Born," *Review of Economics and Statistics*, 78, 165-176.

Bratsberg, Bernt, Erling Barth, and Oddbjorn Raaum (2006): "Local Unemployment and the Relative Wages of Immigrants: Evidence from the Current Population Surveys," *Review of Economics and Statistics*, 88 (2), 243-263.

Butcher, Kristin F. and John DiNardo (2002): "The Immigrant and Native-Born Wage Distributions: Evidence from United States Censuses," *Industrial and Labor Relations Review*, 56 (1), 97-121.

Card, David and John E. DiNardo (2002): "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, 20 (4), 733-783

Chen, Xiaohong, Han Hong, and Elie Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343-366.

Chiswick, Barry R. (1978): "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy*, 86 (5), 897-922.

Chiswick, Barry R. (1980): *An Analysis of the Economic Progress and Impact of Immigrants*, U.S. Department of Labor, National Technical Information Service, PB80-200454.

Dieleman, Frans M. (2001): "Modelling Residential Mobility; A Review of Recent Trends in Research," *Journal of Housing and the Built Environment*, 16, 249-265.

Duleep, Harriet O. and Mark C. Regets (1997a): "Measuring Immigrant Wage Growth using Matched CPS Files," *Demography*, 34, 239-249.

Duleep, Harriet O. and Mark C. Regets (1997b): "Immigrant Entry Earnings and Human Capital Growth: Evidence from 1960-1980 Censuses," in Soloman W. Polachek, ed., Vol 16, *Research in Labor Economics*, JAI Press, 297–317.

Gittleman, Maury and Mary Joyce (1996): "Earnings Mobility and Long-Run Inequality: An Analysis using Matched CPS Data," *Industrial Relations*, 35 (2), 180-196.

Guell, Maia and Luojia Hu (2006): "Estimating the Probability of Leaving Unemployment using Uncompleted Spells from Repeated Cross-Section Data," *Journal of Econometrics*, 133 (1), 307-341.

Hirsch, Barry T. and Edward J. Schumacher (2004): "Match Bias in Wage Gap Estimates Due to Earnings Imputation," *Journal of Labor Economics*, 22 (3), 689-722.

Hirano, Keisuke, Guido W. Imbens, Geert Ridder, and Donald B. Rubin (2001): "Combining Panel Data Sets with Attrition and Refreshment Samples," *Econometrica*, 69, 1645-1659.

Hu, Wei-Yin (1999): "Immigrant Earnings Assimilation: Estimates from Longitudinal Data," *American Economic Review Papers and Proceedings*, 90, 368-372.

Jasso, Guillermina and Mark R. Rosenzweig (1982): "Estimating the Emigration Rates of Legal Immigrants using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States," *Demography*, 19 (3), 279-290.

Jasso, Guillermina and Mark R. Rosenzweig (1988): "How Well do U.S. Immigrants do? Vintage Effects, Emigration Selectivity, and Occupational Mobility," in T. Paul Schultz, ed., *Research in Population Economics*, Vol 6. 229-253.

Jasso, Guillermina and Mark R. Rosenzweig (1990): *The New Chosen People: Immigrants in the United States*, New York: Russell Sage Foundation.

Jasso, Guillermina, Mark R. Rosenzweig, and James P. Smith (1998): "The Changing Skills of New Immigrants to the United States: Recent Trends and Their Determinants," Working Paper 6764, National Bureau of Economic Research.

Kitamura, Yuichi, Gautam Tripathi, and Hyungtaik Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667-1714.

LaLonde, Robert J. and Robert H. Topel (1992): "The Assimilation of Immigrants in the U.S. Labor Market," in George J. Borjas and Richard B. Freeman, eds., *Immigration and the Work Force*, 67-92.

LaLonde, Robert J. and Robert H. Topel (1997): "Economic Impact of International Migration and The Economic Performance of Migrants," in Mark R. Rosenzweig and Oded Stark, eds., *Handbook of Population and Family Economics*, Vol 3B, Ch 14.

Lemieux, Thomas (2006): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, 96 (3), 461-498.

Little, Roderick J. A., and Donald B. Rubin (2002): *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics.

Long, Larry (1988): *Migration and Residential Mobility in the United States*, Russell Sage Foundation.

Lubotsky, Darren H. (2000): "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings," Working paper.

Lubotsky, Darren H. (2001): "The Effect of Changes in the U.S. Wage Structure on Recent Immigrants' Earnings," Working paper.

Neumark, David and William Wascher (1996): "The Effects of Minimum Wages on Teenage Employment and Enrollment: Evidence from Matched CPS Surveys," in S.W. Polachek, ed., Vol 15, *Research in Labor Economics*, JAI Press, 25–63.

Peracchi, Franco and Finis Welch (1995): "How Representative are Matched Cross-Sections? Evidence from the Current Population Survey," *Journal of Econometrics*, 68 (1), 153-179.

Ridder, Geert, and Robert Moffitt (2003): "The Econometrics of Data Combination," Working Paper.

Schmitt, John (2003): "Creating a consistent hourly wage series from the Current Population Survey's Outgoing Rotation Group, 1979-2002," Working Paper.

U.S. Census Bureau (2002): *Technical Paper 63RV: Current Population Survey - Design and Methodology.*

Warren, Robert and Jennifer M. Peck (1980): "Foreign-Born Emigration from the United States: 1960 to 1970," *Demography*, 17, 71-84.

Wooldridge, Jeffrey M. (2002): "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification," *Portuguese Economic Journal*, 1, 117-139.

8 Appendix

8.1 Variables used in the Analysis

This section explains in detail how the CPS MORG are processed to generate the sample used in the analysis. The wage measure used in the analysis is the hourly rate of pay. The wage measure is the hourly wage for the hourly workers and the weekly payments divided by the usual weekly hours of work for non-hourly workers. We clean the wage measure by following steps which are similar to those in Lemieux (2006). Both the hourly and the weekly wages are topcoded. For workers paid by the hour, the topcode remains between $99.00-99.99 and only a small fraction of workers have their wage censored at this value. On the other hand, a substantial number of non-hourly workers have topcoded wages. The weekly wage is topcoded by $1923 in 1994-1997 and by $2884 in 1998-2004. Topcoded wages are adjusted by a factor of 1.4.[29] Workers with extreme wages (less than $2 and more than $200 in 1994 dollars) are trimmed. In addition, the sample drops persons with negative potential experience. As a result, 998 out of 35,016 foreign-born and 11,791 out of 266,628 native-born persons are dropped. These trimmed samples are used throughout the paper unless otherwise indicated.

The year of arrival information provided by the CPS MORG let us identify those who arrived in the United States before 1950, 1950-1959, 1960-1964, 1965-1969, 1970-1974, 1975-1979, 1980-1981, 1982-1983, and so on. The most recent entrants, however, are coded in a inconsistent way. For instance, the arrival year code 13 in the 1994 sample includes the 1992-1994 arrivals, the code 13 in the 1995 sample includes the 1992-1995 arrivals, and the code 13 in the 1996 sample and afterwards include the 1992-1993 arrivals. Therefore foreign-born persons who arrived in the United States in 1992-1993 and are in the 1994-1995 or the 1995-1996 panels cannot be matched. In consequence, we drop immigrants with the arrival year code 13 in the 1994-1995 or the 1995-1996 panels. So, the most recent immigrants in the 1994-1995 and the 1995-1996 panels are those who entered the U.S. in 1990-1991 with the arrival year code 12. Accordingly in the panels of the subsequent years, we keep immigrants with the arrival year code numbers of the followings:

1994-1995 panel: codes 1-12 (1990-1991)
1995-1996 panel: codes 1-12 (1990-1991)
1996-1997 panel: codes 1-13 (1992-1993)
1997-1998 panel: codes 1-13 (1992-1993)
1998-1999 panel: codes 1-14 (1994-1995)
1999-2000 panel: codes 1-14 (1994-1995)
2000-2001 panel: codes 1-15 (1996-1997)
2001-2002 panel: codes 1-15 (1996-1997)
2002-2003 panel: codes 1-16 (1998-1999)
2003-2004 panel: codes 1-16 (1998-1999)

where the years in the parentheses indicate the entry years of the most recent immigrants.

Some variables in the CPS MORG are given by intervals. One example is the arrival year. It is given by periods rather than years. In the analysis, the arrival year variable is defined by the mid-point of each period.

[29] The simplest way of handling topcoded values is to adjust censored values by a factor that approximates the mean for those above the censoring point (typically, a factor like 1.33 or 1.4). According to Schmitt (2003), a more sophisticated way is estimating the mean above the topcode using the pareto distribution. As the pareto distribution has two parameters, what is mostly done is to fit the pareto distribution through a point high in the observed distribution.

Immigrants who arrived in the United States before 1950 are coded as 1940. The education measure needs adjustment, too. The values for the education measure are assigned by the following rule:

0 if less than 1st grade

2.5 if 1st-4th grade

5.5 if 5th-6th

7.5 if 7th-8th

10 if 9th, 10th, 11th, or 12th grades with no diploma

12 if high school graduate including GED

14 if some college but no degree or Associate degree

16 if Bachelor's degree

18 if Master's degree, Professional school degree, or Doctorate degree

The estimation results are not very sensitive to the ways of coding year of entry and education.

8.2 Equivalence of Identification Conditions

Equivalence of the identification conditions (6) and the conditional moment restrictions (7) when the population is stationary is proved by Bhattacharya (2006). We show equivalence of (10) and (12) with outmigration. The first identification condition in (10) is

$$f(u_1, v) = \int \frac{\Pr(D_S = 1)}{\Pr(D_S = 1 | u_1, u_2, v)} f(u_1, u_2, v | D_S = 1)\, du_2.$$

Dividing both sides with $f(u_1, v)$, we have

$$\begin{aligned}
1 &= \int \frac{\Pr(D_S = 1) f(u_1, u_2, v | D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)) f(u_1, v)} du_2 \\
&= \int \frac{\Pr(D_S = 1) f(u_1, u_2, v | D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)) f(u_1, v)} du_2 \\
&= \int \frac{P(u_2, D_S = 1 | u_1, v)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} du_2 \\
&= \sum_{s=0,1} \int \frac{s \cdot P(u_2, D_S = s | u_1, v)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} du_2 \\
&= E\left[\frac{D_S}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} | u_1, v\right] \quad \text{for all } u_1, v,
\end{aligned}$$

which is the first condition in (12). The second identification condition in (10) is

$$\begin{aligned}
f(u_2, v) &= \frac{f(u_2, v | D_P = 1) \Pr(D_P = 1)}{\Pr(D_P = 1 | u_2, v)} \\
&= \int \frac{\Pr(D_S = 1)}{\Pr(D_S = 1 | u_1, u_2, v)} f(u_1, u_2, v | D_S = 1)\, du_1.
\end{aligned}$$

Notice that, we do not observe $f(u_2, v)$, but do $f(u_2, v|D_P = 1)$. Thus, dividing both sides with $f(u_2, v)$, we have

$$\begin{aligned}
1 &= \int \frac{\Pr(D_S = 1) f(u_1, u_2, v|D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)) f(u_2, v)} du_1 \\
&= \int \frac{\Pr(D_S = 1) f(u_1, u_2, v|D_S = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)) f(u_2, v|D_P = 1)} \cdot \frac{\Pr(D_P = 1|u_2, v)}{\Pr(D_P = 1)} du_1 \\
&= \int \frac{\Pr(D_S = 1|D_P = 1) f(u_1, u_2, v|D_S \cdot D_P = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)) f(u_2, v|D_P = 1)} \cdot \Pr(D_P = 1|u_2, v) du_1 \\
&= \int \frac{P(u_1, u_2, v, D_S = 1|D_P = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v)) f(u_2, v|D_P = 1)} \cdot \Pr(D_P = 1|u_2, v) du_1 \\
&= \int \frac{P(u_1, D_S = 1|u_2, v, D_P = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} \cdot \Pr(D_P = 1|u_2, v) du_1 \\
&= \sum_{s=0,1} \int \frac{s \cdot P(u_1, D_S = s|u_2, v, D_P = 1)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} \cdot \Pr(D_P = 1|u_2, v) du_1 \\
&= E\left[\frac{D_S \cdot \Pr(D_P = 1|u_2, v)}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} | u_2, v, D_P = 1\right] \quad \text{for all } u_2, v,
\end{aligned}$$

where the second equation uses

$$f(u_2, v) = \frac{f(u_2, v|D_P = 1) \Pr(D_P = 1)}{\Pr(D_P = 1|u_2, v)}$$

and the third equation uses

$$\begin{aligned}
\Pr(D_S = 1) &= \Pr(D_S = 1|D_P = 1) \cdot \frac{\Pr(D_P = 1)}{\Pr(D_P = 1|D_S = 1)} \\
&= \Pr(D_S = 1|D_P = 1) \cdot \Pr(D_P = 1),
\end{aligned}$$

and

$$f(u_1, u_2, v|D_S = 1) = f(u_1, u_2, v|D_S \cdot D_P = 1),$$

as $D_S = 1$ implies $D_S \cdot D_P = 1$. Finally, by the assumption (11), we have

$$\frac{1}{k(z_2)} = E\left[\frac{D_S}{g(k_0'(v) + k_1'(u_1, v) + k_2'(u_2, v))} | u_2, v, D_P = 1\right] \quad \text{for all } u_2, v.$$

This is the second condition in (12).

8.3 Estimation of Weighting Functions

This section discusses technical details in the weighting function estimation. Empirical specification of the model is as follows. Let

$$\Pr(D_S = 1|U_1 = u_1, U_2 = u_2, V = v) = g\left(v'\phi_0 + u_1'\phi_1 + u_2'\phi_2\right),$$

where v is a vector of constant, age, education, and dummy variables (marital status, years in the United States, citizenship status, country of birth), u_1 and u_2 are vectors of log real hourly dollar wages and indicators of not usually working, and $g(r) = e^r/(1+e^r)$. Assume that (11) holds and z_2 is a vector of age, years since migration, education (assuming that no additional schooling is obtained), country of origin, and year of entry.

The identification conditions in (12) can be transformed to

$$\begin{aligned} E\left[\frac{D_S \cdot a(u_1,v)}{g(v'\phi_0 + u_1'\phi_1 + u_2'\phi_2)}\right] &= E[a(u_1,v)], \\ E\left[\frac{D_S \cdot a(u_2,v)}{g(v'\phi_0 + u_1'\phi_1 + u_2'\phi_2)}\right] &= E\left[\frac{a(u_2,v)}{k(z_2)}\right], \end{aligned} \tag{21}$$

for an arbitrary function $a(\cdot)$. Let n be the sample size of the incoming sample and n_m be the sample size of the matched sample. In addition, let n_1 and n_2 be the sample sizes of the representative cross-section samples in the incoming and the outgoing years. For simplicity, write

$$g(u_1,u_2,v,\phi) \equiv g\left(v'\phi_0 + u_1'\phi_1 + u_2'\phi_2\right).$$

Then, the LHS of (21) can be estimated by

$$\begin{aligned} \frac{1}{n}\sum_{m=1}^{n}\frac{D_{Sm}\cdot a(u_{tm},v_m)}{g(u_{1m},u_{2m},v_m,\theta)} &= \frac{1}{n}\sum_{l=1}^{n_m}\frac{1\cdot a(u_{tm},v_m)}{g(u_{1m},u_{2m},v_m,\theta)} + \frac{1}{n}\sum_{m=n_m+1}^{n}\frac{0\cdot a(u_{tm},v_m)}{g(u_{1m},u_{2m},v_m,\theta)} \\ &= \frac{1}{n}\sum_{l=1}^{n_m}\frac{a(u_{tm},v_m)}{g(u_{1m},u_{2m},v_m,\theta)}, \end{aligned}$$

for $t = 1, 2$, and the RHS of (21) can be estimated by

$$\frac{1}{n_1}\sum_{i=1}^{n_1} a(u_{1i},v_i) = 0,$$

for $t = 1$ and

$$\frac{1}{n_2}\sum_{i=1}^{n_2}\frac{a(u_{2i},v_i)}{k(z_2)} = 0,$$

for $t = 2$. In estimation, the LHS uses the matched longitudinal sample and the RHS uses the representative cross-sections. The GMM can be used, where function $a(\cdot)$ is a vector of *age*, *age*^{2}, *age*^{3}, *educ*, *educ*^{2}, *educ*^{3}, a marital status dummy, $\log wage$, $\log wage^2$, $\log wage^3$, and a dummy of not working for period $t = 1, 2$. For the foreign sample, we add *ysm*, *ysm*^{2}, *ysm*^{3}, a citizenship dummy, and continent of origin (Europe, Asia, and Africa-Oceania) dummies.

The outmigration process, $k(z_2)$, can be estimated as follows. Notice that all the variables in z_1 have deterministic time paths and map to z_2 one-to-one. Without loss of generality we replace (11) with

$$\Pr(D_P = 1 | Z_1 = z_1) = k(z_1).$$

When we have z_1, we do not need to worry about the transition probability $P(Z_2 = z_2 | Z_1 = z_1)$. Assume that $k(z_1)$ is given by a parametric form: $k(z_1'\psi)$. Now consider the following transformation:

$$p(z_1'\psi) \equiv \frac{k(z_1'\psi)}{1 + k(z_1'\psi)}.$$

The estimation strategy is estimate $p(z_1'\psi)$ and transform it to $k(z_1'\psi)$. We use discrete choice model instead of applying GMM. Generate an indicator variable that is set to unity for observations in the second period cross-section. To fix idea, suppose there is no outmigration and assume that the sample sizes are the same. Then there will equal number of 0's and 1's. So $p(z_1'\psi) = 1/2$ for all z_1. If outmigration occurs to individuals with $z_1 = \widetilde{z}_1$, we expect $p(\widetilde{z}_1'\psi) < 1/2$. We use a logit model

$$p(z_1'\psi) = \frac{e^{z_1'\psi}}{1 + e^{z_1'\psi}}$$

and obtain $p(z_1'\psi)$.

Tables A4-1 and A4-2 report the estimated ϕ coefficients for the native and the foreign samples under the stationary population (or the negligible outmigration) assumption. The weighting function estimates in 1994-1995 and 1995-1996 are less stable than in other years because of the smaller sample sizes. Again, the estimates do not necessarily have causal interpretation. For instance, labor market outcome and residential mobility may affect each other. Over the matching periods from 1996-1997 through 2003-2004, the matching rates are higher among older, married, not recently arrived citizens. Among those who work, the current wage is positively correlated with the matching rate. In addition, those who are not usually working are matched with higher probability.

Table A4-3 reports the ψ estimates, where a positive coefficient implies that the stay probability in the United States is positively correlated with the variable. We observed that the outmigration process is rather poorly estimated. Just looking at the signs of estimates, only the education has stable coefficients. More educated foreign-born persons have higher stay probability than less educated. The other variables, age, years since migration, country of origin, and the arrival year are not significant. The estimation results do not support the hypothesis that outmigration rates decline with time spent in the United States. However, this may not be very surprising because outmigration is not very large between years. Table A4-4 reports the attrition-outmigration correcting weighting function estimates.

8.4 Comparison with Other Attrition-Correction Approaches

There are two general approaches are often used to deal with attrition in panel data sets: selection on observables and selection on unobservables. Models of selection on observables make the assumption that the probability of attrition depends only on U_1 and V, which are observed. In this case, U_2 is missing at random in the panel, i.e.,

$$D_S \perp U_2 | (U_1, V),$$

and the attrition probability can be written as

$$\Pr(D_S = 1 | U_1, U_2, V) = \Pr(D_S = 1 | U_1, V).$$

With this structure $\Pr(D_S = 1 | u_1, u_2, v) = \Pr(D_S = 1 | u_1, v)$ can be observed and therefore $f(u_1, u_2, v)$ is identified. An example where the assumption of selection on observables fails is when an individual does not respond in the second period if the individual experiences an unexpected negative wage shock in the second period. Under selection on observable assumption, Jeffrey Wooldridge (2002) proposes an inverse probability weighted (IPW) M-estimator for two-period panel data models. It is known that if selection is ignorable, an inverse probability weighting scheme generally identifies the population parameters.

Models of selection on unobservables make the assumption that the probability of attrition depends on U_2, which may not be observed. In this case, the attrition process does not depend on the first period variables U_1, i.e.,

$$D_S \perp U_1 | \left(U_2, V\right),$$

and the attrition probability can be written as

$$\Pr\left(D_S = 1 | U_1, U_2, V\right) = \Pr\left(D_S = 1 | U_2, V\right).$$

This assumption fails if individuals do not respond in the second period if there was a negative wage shock in the first period. For estimation, the probability of attrition can be specified to depend on arbitrary functions of U_2. A special case of the method of selection on unobservables is the standard sample selection model by Heckman (1976, 1979). Heckman's solution requires at least one exogenous variable affecting selection that does not appear in the structural equation.

Table A1. Stay Probability (One minus the Outmigration Rate) by Ethnic Origin

	1994 –1995	1995 –1996	1996 –1997	1997 –1998	1998 –1999	1999 –2000	2000 –2001	2001 –2002	2002 –2003	2003 –2004	[row in total] row average
before 1980 arrivals											
C.S.America	1016	1014	1019	964	979	891	880	798	845	812	[0.748]
	1.024	0.858	0.952	0.979	0.941	1.027	1.007	1.014	0.998	0.929	0.971
Europe	611	674	537	540	520	461	486	445	498	424	[0.537]
	1.005	0.785	0.944	0.930	0.940	0.978	0.998	1.054	0.940	0.854	0.940
Asia	518	575	495	502	510	488	430	455	464	430	[0.651]
	1.041	0.854	1.032	0.896	0.961	0.945	1.100	0.991	0.905	0.884	0.958
1980-1991 arrivals											
C.S.America	1399	1439	1396	1448	1456	1498	1490	1484	1597	1541	[0.965]
	1.065	0.941	0.983	0.966	0.955	1.012	1.060	1.065	1.009	0.921	0.996
Europe	279	385	264	295	268	247	241	307	289	330	[0.961]
	1.018	0.875	1.038	0.980	1.045	1.117	1.083	1.010	0.972	0.855	0.996
Asia	680	965	805	829	720	969	706	712	703	717	[0.652]
	1.059	0.802	0.970	1.037	1.001	0.752	1.067	1.098	0.980	0.881	0.958
1992-1993 arrivals											
C.S.America			173	237	200	209	253	218	265	266	[1.030]
			1.000	0.920	1.130	1.139	0.953	1.110	0.940	0.876	1.004
Europe			63	73	89	83	58	63	64	56	[0.876]
			0.937	0.973	1.101	0.916	1.000	1.032	1.016	0.911	0.984
Asia			117	152	120	114	119	125	122	124	[1.034]
			0.957	0.967	0.925	1.088	1.092	1.168	1.008	0.863	1.004
1994-1995 arrivals											
C.S.America					218	253	302	320	309	277	[1.149]
					1.055	1.059	1.017	1.072	0.961	0.982	1.023
Europe					78	69	80	59	69	69	[1.232]
					0.962	1.072	1.050	1.017	1.043	1.072	1.035
Asia					141	111	126	143	102	114	[1.066]
					0.894	1.117	1.111	1.014	1.010	0.939	1.011
1996-1997 arrivals											
C.S.America							251	287	301	308	[0.790]
							1.068	0.913	0.960	0.844	0.943
Europe							87	58	85	75	[0.979]
							0.908	1.207	0.918	0.973	0.995
Asia							152	127	138	129	[0.807]
							0.875	1.150	0.855	0.938	0.948
1998-1999 arrivals											
C.S.America									411	462	[0.805]
									0.973	0.827	0.897
Europe									118	99	[0.779]
									0.941	0.828	0.883
Asia									158	154	[0.771]
									0.949	0.812	0.854

1st row: #in 1st year; 2nd row: stay probability

Table A2-1. Economic Assimilation Estimates in % (by Origin): Reported & Imputed Wages

Individual Hetero.	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	linear	quadra.	cubic	linear	quadra.	cubic	linear	quadra.	cubic
C.S.America									
age=24, ysm=4	0.01	–1.46**	–2.03**	–0.01	–1.52**	–2.09**	0.17	–1.25*	–1.98**
	(0.41)	(0.70)	(0.85)	(0.41)	(0.70)	(0.85)	(0.41)	(0.71)	(0.86)
age=32, ysm=12		–0.88*	–0.63		–0.91**	–0.66		–0.78	–0.63
		(0.45)	(0.52)		(0.45)	(0.52)		(0.45)	(0.52)
age=40, ysm=20		–0.30	0.24		–0.30	0.24		–0.31	0.19
		(0.46)	(0.65)		(0.46)	(0.65)		(0.46)	(0.65)
age=48, ysm=28		0.28	0.59		0.31	0.62		0.16	0.49
		(0.72)	(0.74)		(0.72)	(0.73)		(0.72)	(0.74)
Europe									
age=24, ysm=4	–1.68	–3.09*	–1.92	–1.69	–3.17*	–1.94	–1.39	–3.14*	–2.21
	(0.90)	(1.77)	(2.31)	(0.90)	(1.76)	(2.30)	(0.90)	(1.79)	(2.38)
age=32, ysm=12		–2.20*	–2.23*		–2.24*	–2.29*		–2.19*	–2.09*
		(1.23)	(1.27)		(1.22)	(1.26)		(1.24)	(1.26)
age=40, ysm=20		–1.30	–2.06		–1.32	–2.13		–1.25	–1.68
		(0.91)	(1.48)		(0.90)	(1.47)		(0.91)	(1.47)
age=48, ysm=28		–0.40	–1.42		–0.39	–1.45		–0.30	–0.99
		(1.04)	(1.51)		(1.03)	(1.50)		(1.03)	(1.50)
Asia									
age=24, ysm=4	0.57	0.15	1.15	0.55	–0.01	0.96	0.91	0.79	2.47*
	(0.69)	(1.40)	(1.82)	(0.69)	(1.40)	(1.81)	(0.69)	(1.38)	(1.74)
age=32, ysm=12		0.43	0.13		0.36	0.07		0.86	0.41
		(0.86)	(0.92)		(0.85)	(0.92)		(0.84)	(0.91)
age=40, ysm=20		0.71	0.04		0.72	0.08		0.92	–0.23
		(0.80)	(1.14)		(0.80)	(1.14)		(0.80)	(1.14)
age=48, ysm=28		0.98	0.86		1.08	0.97		0.99	0.54
		(1.29)	(1.32)		(1.29)	(1.32)		(1.29)	(1.31)
Others									
age=24, ysm=4	1.17	–0.73	–0.09	0.93	–1.15	–0.42	1.80	–0.29	0.87
	(1.58)	(2.84)	(3.43)	(1.58)	(2.84)	(3.44)	(1.54)	(2.83)	(3.54)
age=32, ysm=12		0.34	0.01		–0.02	–0.43		0.74	0.26
		(1.83)	(2.02)		(1.83)	(2.01)		(1.81)	(1.95)
age=40, ysm=20		1.42	0.71		1.11	0.27		1.77	0.59
		(1.70)	(2.51)		(1.70)	(2.51)		(1.69)	(2.46)
age=48, ysm=28		2.49	1.99		2.23	1.68		2.80	1.85
		(2.60)	(2.82)		(2.61)	(2.83)		(2.60)	(2.83)

Standard errors are reported in parentheses. Confidence levels: 99% (***), 95% (**), 90% (*).

Sample sizes: Native (156241), C.S.America (11560), Europe (3392), Asia (5340), Others (1162)

Estimates represent immigrants' annual percentage wage growth relative to the natives' percentage wage growth.

Table A2-2. Economic Assimilation Estimates in % (by Origin): Weighted Reported Wages

Individual Hetero.	ATT-OUT-Adjusted			Not Adjusted			A-O-Adjusted, enter ≥18		
	linear	quadra.	cubic	linear	quadra.	cubic	linear	quadra.	cubic
C.S.America									
age=24, ysm=4	0.16	–1.29**	–2.13***	0.17	–1.21*	–2.25**	–0.71	–2.77***	–3.80***
	(0.37)	(0.64)	(0.77)	(0.37)	(0.63)	(0.77)	(0.50)	(0.99)	(1.48)
age=32, ysm=12		–0.71*	–0.36		–0.76**	–0.54		–2.05***	–1.43*
		(0.41)	(0.47)		(0.41)	(0.46)		(0.70)	(0.84)
age=40, ysm=20		–0.12	0.64		–0.32	0.41		–1.32*	–0.36
		(0.41)	(0.58)		(0.41)	(0.57)		(0.73)	(0.92)
age=48, ysm=28		0.46	0.86		0.12	0.59		–0.60	–0.57
		(0.64)	(0.66)		(0.63)	(0.66)		(1.06)	(1.14)
Europe									
age=24, ysm=4	–1.28	–1.25	1.76	–1.15	–1.50	2.47	–1.39	0.41	7.98**
	(0.85)	(1.71)	(2.49)	(0.82)	(1.74)	(2.62)	(1.18)	(2.45)	(3.69)
age=32, ysm=12		–1.04	–1.34		–1.18	–1.13		–0.94	–2.63
		(1.19)	(1.19)		(1.22)	(1.22)		(1.91)	(2.23)
age=40, ysm=20		–0.83	–2.80**		–0.85	–2.80**		–2.29	–7.53**
		(0.85)	(1.29)		(0.86)	(1.22)		(1.72)	(2.61)
age=48, ysm=28		–0.62	–2.60*		–0.53	–2.56**		–3.64*	–6.71***
		(0.92)	(1.35)		(0.90)	(1.29)		(1.99)	(2.50)
Asia									
age=24, ysm=4	–0.49	–0.75	–0.16	–0.33	–1.04	0.38	–1.53	–3.07*	–2.51
	(0.63)	(1.36)	(1.83)	(0.62)	(1.30)	(1.71)	(0.73)	(1.68)	(2.65)
age=32, ysm=12		–0.49	–0.37		–0.57	–0.46		–2.00*	–2.10
		(0.81)	(0.86)		(0.78)	(0.85)		(1.20)	(1.50)
age=40, ysm=20		–0.23	–0.35		–0.10	–0.26		–0.93	–0.99
		(0.77)	(1.04)		(0.76)	(1.03)		(1.16)	(1.39)
age=48, ysm=28		0.03	–0.11		0.37	0.22		0.14	0.82
		(1.29)	(1.31)		(1.25)	(1.25)		(1.60)	(1.69)
Others									
age=24, ysm=4	–0.82	–0.43	–3.40	–0.15	–0.04	–1.88	–0.63	2.78	–5.48
	(1.81)	(3.30)	(4.16)	(1.66)	(2.93)	(3.80)	(2.05)	(4.83)	(6.28)
age=32, ysm=12		0.10	1.47		0.44	1.06		4.25	0.17
		(2.08)	(2.34)		(1.85)	(2.00)		(3.77)	(3.33)
age=40, ysm=20		0.64	3.25		0.93	2.39		5.72	6.39
		(1.91)	(2.80)		(1.85)	(2.62)		(3.78)	(4.72)
age=48, ysm=28		1.17	1.94		1.41	2.12		7.19	13.19**
		(2.98)	(3.03)		(2.92)	(3.02)		(4.83)	(6.28)

Standard errors are reported in parentheses. Confidence levels: 99% (***), 95% (**), 90% (*).

Sample sizes: Native (89117), C.S.America (6438), Europe (1689), Asia (2657), Others (492)

Sample sizes of the last column: Native (89117), C.S.America (3530), Europe (979), Asia (1922), Others (355)

Estimates represent immigrants' annual percentage wage growth relative to the natives' percentage wage growth.

Table A3-1. Wage Equation (in First Differenced) Estimates: Reported & Imputed Wages

IH	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	linear	quadratic	cubic	linear	quadratic	cubic	linear	quadratic	cubic
Constant	0.028***	0.101***	0.187***	0.028***	0.101***	0.187***	0.016***	0.091***	0.179***
	(0.005)	(0.006)	(0.015)	(0.005)	(0.006)	(0.015)	(0.005)	(0.007)	(0.015)
$\frac{1}{10}$Age		–0.019***	–0.066***		–0.019***	–0.066***		–0.019***	–0.065***
		(0.001)	(0.007)		(0.001)	(0.007)		(0.001)	(0.007)
$\frac{1}{100}$Age^2			0.006***			0.006***			0.006***
			(0.001)			(0.001)			(0.001)
Imm.	–0.000	–0.030**	–0.062	–0.001	–0.032***	–0.065	0.002	–0.026**	–0.058
	(0.003)	(0.012)	(0.040)	(0.003)	(0.012)	(0.040)	(0.003)	(0.013)	(0.041)
$\frac{1}{10}$Age$_i$		0.007**	0.028		0.008**	0.029		0.007*	0.029
		(0.004)	(0.022)		(0.004)	(0.022)		(0.004)	(0.022)
$\frac{1}{100}$Age$_i^2$			–0.003			–0.003			–0.003
			(0.003)			(0.003)			(0.003)
$\frac{1}{10}$Ysm		–0.000	–0.008		–0.000	–0.008		–0.001	–0.012
		(0.004)	(0.011)		(0.004)	(0.010)		(0.004)	(0.011)
$\frac{1}{100}$Ysm^2			0.002			0.002			0.002
			(0.002)			(0.002)			(0.002)
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are reported in parentheses. N: sample size = 177695

Imm.: indicator variable of a foreign-born person; Age$_i$: age of foreign-born persons × Imm.; Ysm: years since migration

Year: calendar year fixed effects

Table A3-2. Wage Equation (in First Differenced) Estimates: Reported Wages Only

IH-R	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	linear	quadratic	cubic	linear	quadratic	cubic	linear	quadratic	cubic
Constant	0.024^{***}	0.096^{***}	0.178^{***}	0.024^{***}	0.096^{***}	0.178^{***}	0.034^{***}	0.108^{***}	0.194^{***}
	(0.005)	(0.006)	(0.014)	(0.005)	(0.006)	(0.014)	(0.005)	(0.006)	(0.014)
$\frac{1}{10}$Age		-0.019^{***}	-0.064^{***}		-0.019^{***}	-0.064^{***}		-0.018^{***}	-0.064^{***}
		(0.001)	(0.007)		(0.001)	(0.007)		(0.001)	(0.007)
$\frac{1}{100}$Age^2			0.006^{***}			0.006^{***}			0.006^{***}
			(0.001)			(0.001)			(0.001)
Imm.	–0.002	–0.018	–0.034	–0.002	-0.019^{*}	–0.036	–0.002	-0.021^{*}	–0.028
	(0.003)	(0.012)	(0.039)	(0.003)	(0.012)	(0.039)	(0.003)	(0.012)	(0.038)
$\frac{1}{10}$Age$_i$		0.002	0.008		0.003	0.008		0.003	0.005
		(0.003)	(0.021)		(0.003)	(0.021)		(0.003)	(0.020)
$\frac{1}{100}$Age$_i^2$			–0.001			–0.001			–0.000
			(0.003)			(0.003)			(0.002)
$\frac{1}{10}$Ysm		0.003	0.011		0.002	0.010		0.001	0.008
		(0.003)	(0.009)		(0.003)	(0.009)		(0.003)	(0.009)
$\frac{1}{100}$Ysm^2			–0.002			–0.002			–0.001
			(0.002)			(0.002)			(0.002)
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are reported in parentheses. N: sample size = 100393

Imm.: indicator variable of a foreign-born person; Age$_i$: age of foreign-born persons × Imm.; Ysm: years since migration

Year: calendar year fixed effects

Table A3-3. Wage Equation (in Level) Estimates: Reported & Imputed Wages

CH	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	linear	quadratic	cubic	linear	quadratic	cubic	linear	quadratic	cubic
Constant	0.658***	–0.441***	–1.092***	0.658***	–0.441***	–1.092***	0.772***	–0.380***	–1.143***
	(0.010)	(0.016)	(0.043)	(0.010)	(0.016)	(0.043)	(0.011)	(0.016)	(0.046)
Age	0.014***	0.078***	0.134***	0.014***	0.078***	0.134***	0.012***	0.076***	0.139***
	(0.000)	(0.001)	(0.004)	(0.000)	(0.001)	(0.004)	(0.000)	(0.001)	(0.004)
$\frac{1}{100}\text{Age}^2$		–0.081***	–0.228***		–0.081***	–0.228***		–0.079***	–0.242***
		(0.001)	(0.009)		(0.001)	(0.009)		(0.001)	(0.010)
$\frac{1}{1000}\text{Age}^3$			0.012***			0.012***			0.013***
			(0.001)			(0.001)			(0.001)
Imm.	0.464***	0.562***	0.320**	0.468***	0.566***	0.313**	0.404***	0.516***	0.281**
	(0.023)	(0.046)	(0.129)	(0.023)	(0.046)	(0.129)	(0.024)	(0.048)	(0.135)
Age_i	–0.007***	–0.019***	–0.000	–0.007***	–0.019***	0.001	–0.006***	–0.019***	–0.001
	(0.000)	(0.002)	(0.010)	(0.000)	(0.002)	(0.010)	(0.000)	(0.002)	(0.011)
$\frac{1}{100}\text{Age}_i^2$		0.015***	–0.035		0.015***	–0.037		0.015***	–0.030
		(0.003)	(0.027)		(0.003)	(0.027)		(0.003)	(0.028)
$\frac{1}{1000}\text{Age}_i^3$			0.004*			0.004*			0.004
			(0.002)			(0.002)			(0.002)
Ysm	0.015***	0.025***	0.026***	0.015***	0.025***	0.026***	0.016***	0.027***	0.030***
	(0.001)	(0.003)	(0.005)	(0.001)	(0.003)	(0.005)	(0.001)	(0.003)	(0.006)
$\frac{1}{100}\text{Ysm}^2$		–0.035***	–0.043*		–0.035***	–0.043*		–0.037***	–0.053**
		(0.007)	(0.026)		(0.007)	(0.023)		(0.007)	(0.027)
$\frac{1}{1000}\text{Ysm}^3$			0.001			0.001			0.002
			(0.004)			(0.004)			(0.004)
Educ	0.098***	0.092***	0.092***	0.098***	0.092***	0.092***	0.100***	0.095***	0.094***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Educ_i	–0.042***	–0.037***	–0.036***	–0.042***	–0.037***	–0.036***	–0.040***	–0.036	–0.036***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B.Cntry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Arr.Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are reported in parentheses. N: sample size = 177695

Imm.: indicator of a Central or South American who immigrated between 1998-1999

Age_i: age × Imm; Ysm: years since migration; Educ: years of schooling; Educ_i: years of schooling × Imm.

B.Cntry: birth country fixed effects; Arr.Year: arrival year fixed effects; Year: calendar year fixed effects

Table A3-4. Wage Equation (in Level) Estimates: Reported Wages Only

CH-R	ATT-Adjusted			ATT-OUT-Adjusted			Not Adjusted		
	linear	quadratic	cubic	linear	quadratic	cubic	linear	quadratic	cubic
Constant	0.641***	–0.493***	–1.088***	0.641***	–0.493***	–1.087***	0.731***	–0.468***	–1.134***
	(0.013)	(0.020)	(0.055)	(0.013)	(0.020)	(0.055)	(0.014)	(0.021)	(0.059)
Age	0.014***	0.081***	0.132***	0.014***	0.081***	0.132***	0.012***	0.078***	0.134***
	(0.000)	(0.001)	(0.005)	(0.000)	(0.001)	(0.005)	(0.000)	(0.001)	(0.005)
$\frac{1}{100}\text{Age}^2$		–0.085***	–0.219***		–0.085***	–0.219***		–0.083***	–0.225***
		(0.001)	(0.012)		(0.001)	(0.012)		(0.001)	(0.013)
$\frac{1}{1000}\text{Age}^3$			0.011***			0.011***			0.012***
			(0.001)			(0.001)			(0.001)
Imm.	0.499***	0.590***	0.288*	0.501***	0.589***	0.281	0.457***	0.580***	0.337*
	(0.032)	(0.063)	(0.172)	(0.032)	(0.063)	(0.172)	(0.033)	(0.065)	(0.182)
Age_i	–0.009***	–0.019***	0.009	–0.009***	–0.019***	0.010	–0.008***	–0.020***	0.003
	(0.001)	(0.003)	(0.014)	(0.001)	(0.003)	(0.014)	(0.001)	(0.003)	(0.015)
$\frac{1}{100}\text{Age}_i^2$		0.013***	–0.063*		0.013***	–0.064*		0.014***	–0.047
		(0.004)	(0.037)		(0.004)	(0.037)		(0.004)	(0.039)
$\frac{1}{1000}\text{Age}_i^3$			0.006**			0.006**			0.005
			(0.003)			(0.003)			(0.003)
Ysm	0.019***	0.024***	0.015**	0.019***	0.024***	0.015**	0.018***	0.026***	0.017**
	(0.002)	(0.004)	(0.007)	(0.002)	(0.004)	(0.007)	(0.002)	(0.004)	(0.008)
$\frac{1}{100}\text{Ysm}^2$		–0.024**	0.028		–0.024**	0.027		–0.028***	0.020
		(0.010)	(0.035)		(0.010)	(0.035)		(0.011)	(0.037)
$\frac{1}{1000}\text{Ysm}^3$			–0.007			–0.007			–0.006
			(0.005)			(0.005)			(0.005)
Educ	0.100***	0.094***	0.093***	0.100***	0.094***	0.093***	0.101***	0.096***	0.096***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Educ_i	–0.043***	–0.038***	–0.038***	–0.043***	–0.038***	–0.038***	–0.41***	–0.037***	–0.037***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B.Cntry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Arr.Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are reported in parentheses. N: sample size = 100393

Imm.: indicator of a Central or South American who immigrated between 1998-1999

Age_i: age × Imm.; Ysm: years since migration; Educ: years of schooling; Educ_i: years of schooling × Imm.

B.Cntry: birth country fixed effects; Arr.Year: arrival year fixed effects; Year: calendar year fixed effects

Table A4-1. Attrition Correcting Weighting Function Estimates (Natives)

	1994 –1995	1995 –1996	1996 –1997	1997 –1998	1998 –1999	1999 –2000	2000 –2001	2001 –2002	2002 –2003	2003 –2004
Age	0.027	0.045	0.054	0.052	0.057	0.053	0.054	0.056	0.049	0.039
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Education	0.024	0.002	–0.019	–0.031	–0.033	–0.013	–0.027	–0.031	–0.031	–0.015
	(0.005)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mari.Stat.	0.404	0.536	0.576	0.611	0.467	0.615	0.666	0.548	0.577	0.503
	(0.027)	(0.030)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)	(0.016)
LogWage1	0.372	–0.283	0.109	–0.015	0.196	0.148	0.027	–0.046	0.174	0.057
	(0.029)	(0.034)	(0.019)	(0.018)	(0.020)	(0.019)	(0.019)	(0.018)	(0.017)	(0.020)
LogWage2	0.084	0.499	0.277	0.252	0.094	0.167	0.068	0.306	0.221	0.226
	(0.030)	(0.034)	(0.018)	(0.020)	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)	(0.020)
NoWork1	0.960	–0.621	0.310	–0.026	0.392	0.459	0.057	–0.134	0.523	0.253
	(0.082)	(0.094)	(0.052)	(0.051)	(0.055)	(0.054)	(0.055)	(0.052)	(0.049)	(0.056)
NoWork2	0.160	1.059	0.465	0.573	0.159	0.363	0.055	0.562	0.314	0.391
	(0.084)	(0.095)	(0.050)	(0.055)	(0.055)	(0.054)	(0.054)	(0.052)	(0.052)	(0.055)
Constant	–2.021	–1.706	–1.742	–1.299	–1.473	–1.817	–1.014	–1.398	–1.582	–1.463
	(0.085)	(0.096)	(0.052)	(0.054)	(0.055)	(0.055)	(0.056)	(0.055)	(0.053)	(0.057)
N	18555	14220	38305	38588	38608	38664	36974	39902	44266	43822
Mat.Rate	67.9%	69.9%	77.7%	76.8%	77.1%	77.5%	78.3%	77.8%	76.7%	70.7%

Standard errors are reported in parentheses. N: sample size, Mat.Rate: matching rate

The LHS variable is the odds of staying in the same address.

Mari.Stat.: 1 if married; LogWage: log of hourly rate of pay (yrs 1&2); NoWork: no reported wage (yrs 1&2)

Table A4-2. Attrition Correcting Weighting Function Estimates (Immigrants)

	1994 –1995	1995 –1996	1996 –1997	1997 –1998	1998 –1999	1999 –2000	2000 –2001	2001 –2002	2002 –2003	2003 –2004
Age	0.032	0.036	0.021	0.037	0.027	0.026	0.031	0.026	0.026	0.030
	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Education	0.021	0.019	0.008	0.015	0.003	0.043	–0.037	0.031	0.001	–0.003
	(0.010)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
Mari.Stat.	0.227	0.360	0.497	0.344	0.620	0.457	0.119	0.605	0.313	0.257
	(0.089)	(0.110)	(0.052)	(0.054)	(0.050)	(0.050)	(0.045)	(0.047)	(0.042)	(0.048)
LogWage1	–0.018	0.267	–0.341	–0.027	–0.022	0.238	0.120	–0.081	–0.108	0.222
	(0.097)	(0.124)	(0.055)	(0.054)	(0.053)	(0.056)	(0.048)	(0.049)	(0.047)	(0.053)
LogWage2	0.054	–0.008	0.452	0.208	0.033	–0.062	0.180	0.280	0.062	0.103
	(0.089)	(0.108)	(0.057)	(0.061)	(0.055)	(0.056)	(0.051)	(0.050)	(0.048)	(0.052)
NoWork1	0.190	0.443	–0.774	–0.329	0.049	0.573	0.138	–0.064	–0.336	0.356
	(0.238)	(0.303)	(0.141)	(0.142)	(0.136)	(0.145)	(0.127)	(0.129)	(0.123)	(0.141)
NoWork2	–0.221	0.224	1.177	0.736	–0.266	0.052	0.433	0.643	0.223	0.238
	(0.229)	(0.276)	(0.147)	(0.157)	(0.144)	(0.144)	(0.133)	(0.133)	(0.126)	(0.139)
Ysm	0.052	0.054	0.048	0.050	0.028	0.092	0.023	0.076	0.033	0.037
	(0.004)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Citizen	–0.421	–0.086	0.045	0.250	0.152	–0.341	0.154	0.134	0.166	0.236
	(0.090)	(0.109)	(0.048)	(0.050)	(0.048)	(0.048)	(0.044)	(0.044)	(0.042)	(0.046)
Europe	0.391	0.849	0.097	0.313	0.346	–0.025	0.260	0.010	0.285	0.024
	(0.102)	(0.120)	(0.051)	(0.063)	(0.059)	(0.062)	(0.059)	(0.060)	(0.054)	(0.061)
Asia	0.362	–0.178	0.059	0.195	0.105	0.002	0.202	–0.016	–0.154	–0.024
	(0.100)	(0.013)	(0.063)	(0.057)	(0.054)	(0.056)	(0.052)	(0.053)	(0.050)	(0.055)
Africa	2.107	–0.857	0.077	–0.858	–0.242	0.071	0.062	–0.121	–0.148	–0.244
	(0.107)	(0.218)	(0.058)	(0.133)	(0.109)	(0.092)	(0.082)	(0.082)	(0.078)	(0.089)
Constant	–1.896	–2.581	–1.420	–2.171	–1.100	–2.417	–1.093	–2.265	–0.802	–1.950
	(0.213)	(0.276)	(0.129)	(0.138)	(0.130)	(0.135)	(0.124)	(0.127)	(0.120)	(0.130)
N	2323	1840	5355	5406	5705	5683	6277	6257	7049	7063
Mat.Rate	66.2%	60.0%	70.1%	68.8%	70.0%	70.6%	71.3%	71.6%	70.2%	64.6%

Standard errors are reported in parentheses. N: sample size, Mat.Rate: matching rate

The LHS variable is the odds of staying in the same address.

Mari.Stat.: 1 if married; LogWage: log of hourly rate of pay (yrs 1&2); NoWork: no reported wage (yrs 1&2)

Ysm: years since migration; Citizen: 1 if U.S. citizen

Constant: immigrants from Central & South America; Continent Dummies are Deviations from the Constant:

Europe: Europe, Australia, New Zealand, and Canada; Africa: Africa and other countries

Table A4-3. Outmigration Process Estimates

	1994 –1995	1995 –1996	1996 –1997	1997 –1998	1998 –1999	1999 –2000	2000 –2001	2001 –2002	2002 –2003	2003 –2004
Age/10	0.017	–0.003	0.023	0.010	–0.002	0.016	–0.009	0.001	–0.009	–0.001
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)	(0.018)	(0.019)	(0.018)	(0.018)
Ysm/10	0.008	–0.001	–0.010	–0.004	–0.011	–0.007	0.022	0.008	0.024	0.024
	(0.022)	(0.023)	(0.023)	(0.023)	(0.021)	(0.022)	(0.020)	(0.021)	(0.018)	(0.019)
Education	0.004	–0.006	0.004	0.001	0.007	0.003	0.007	0.003	0.008	0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Europe	–0.042	–0.063	–0.001	–0.016	0.014	–0.009	–0.030	–0.006	–0.045	–0.035
	(0.061)	(0.060)	(0.061)	(0.060)	(0.058)	(0.059)	(0.057)	(0.057)	(0.053)	(0.056)
Asia	–0.014	–0.071	0.001	0.037	–0.014	–0.024	0.000	0.023	–0.063	–0.029
	(0.055)	(0.053)	(0.053)	(0.051)	(0.051)	(0.051)	(0.049)	(0.049)	(0.047)	(0.049)
Others	–0.309	–0.240	0.115	0.071	0.065	–0.002	0.002	–0.065	–0.004	–0.036
	(0.064)	(0.091)	(0.097)	(0.110)	(0.099)	(0.085)	(0.076)	(0.076)	(0.072)	(0.078)
Constant	–0.073	–0.031	–0.149	–0.083	–0.081	–0.055	–0.049	–0.011	–0.106	–0.188
	(0.088)	(0.091)	(0.090)	(0.089)	(0.083)	(0.086)	(0.082)	(0.083)	(0.077)	(0.082)
N	10534	9920	10010	10184	10801	10892	12212	12186	13681	12749

Standard errors are reported in parentheses. N: sample size

The LHS variable is the odds of staying in the United States.

Ysm: years since migration

Constant: immigrants from Central & South America; Continent Dummies are Deviations from the Constant:

Europe: Europe, Australia, New Zealand, and Canada; Africa: Africa and other countries

Table A4-4. Attrition-Outmigration Correcting Weighting Function Estimates (Immigrants)

	1994 –1995	1995 –1996	1996 –1997	1997 –1998	1998 –1999	1999 –2000	2000 –2001	2001 –2002	2002 –2003	2003 –2004
Age	0.031	0.032	0.023	0.036	0.026	0.026	0.030	0.026	0.024	0.027
	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Education	0.020	0.010	0.013	0.016	0.012	0.044	–0.033	0.031	0.010	0.001
	(0.010)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
Mari.Stat.	0.219	0.368	0.480	0.339	0.606	0.458	0.123	0.604	0.308	0.256
	(0.088)	(0.108)	(0.052)	(0.053)	(0.050)	(0.050)	(0.045)	(0.047)	(0.042)	(0.047)
LogWage1	–0.004	0.243	–0.326	–0.025	–0.021	0.237	0.120	–0.080	–0.104	0.195
	(0.096)	(0.121)	(0.055)	(0.054)	(0.053)	(0.056)	(0.048)	(0.049)	(0.047)	(0.053)
LogWage2	0.055	–0.038	0.431	0.199	0.033	–0.062	0.175	0.279	0.064	0.105
	(0.089)	(0.106)	(0.057)	(0.061)	(0.055)	(0.056)	(0.051)	(0.050)	(0.048)	(0.052)
NoWork1	0.225	0.402	–0.738	–0.309	0.051	0.571	0.137	–0.063	–0.323	0.300
	(0.236)	(0.299)	(0.141)	(0.141)	(0.136)	(0.145)	(0.127)	(0.129)	(0.123)	(0.141)
NoWork2	–0.207	0.093	1.125	0.695	–0.259	0.051	0.423	0.640	0.223	0.250
	(0.229)	(0.272)	(0.146)	(0.156)	(0.144)	(0.144)	(0.133)	(0.133)	(0.126)	(0.140)
Ysm	0.048	0.025	0.045	0.044	0.025	0.091	0.024	0.076	0.034	0.029
	(0.004)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Citizen	–0.372	0.048	0.110	0.248	0.159	–0.340	0.153	0.134	0.171	0.242
	(0.089)	(0.107)	(0.051)	(0.049)	(0.048)	(0.048)	(0.044)	(0.044)	(0.042)	(0.046)
Europe	0.398	0.497	0.049	0.256	0.349	–0.028	0.253	0.010	0.223	–0.037
	(0.102)	(0.120)	(0.063)	(0.063)	(0.059)	(0.062)	(0.059)	(0.060)	(0.054)	(0.061)
Asia	0.352	–0.226	0.074	0.195	0.083	–0.003	0.198	–0.016	–0.227	–0.044
	(0.100)	(0.121)	(0.058)	(0.057)	(0.054)	(0.056)	(0.052)	(0.053)	(0.050)	(0.054)
Africa	0.962	–1.105	0.374	–0.807	–0.220	0.069	0.061	–0.156	–0.158	–0.269
	(0.107)	(0.214)	(0.101)	(0.132)	(0.109)	(0.092)	(0.082)	(0.082)	(0.078)	(0.089)
Constant	–1.853	–2.040	–1.532	–2.132	–1.146	–2.424	–1.111	–2.260	–0.897	–1.861
	(0.212)	(0.229)	(0.129)	(0.138)	(0.130)	(0.135)	(0.124)	(0.127)	(0.120)	(0.129)
N	2323	1840	5355	5406	5705	5683	6277	6257	7049	7063
Mat.Rate	66.2%	60.0%	70.1%	68.8%	70.0%	70.6%	71.3%	71.6%	70.2%	64.6%

Standard errors are reported in parentheses. N: sample size, Mat.Rate: matching rate

The LHS variable is the odds of staying in the same address.

Mari.Stat.: 1 if married; LogWage: log of hourly rate of pay (yrs 1&2); NoWork: no reported wage (yrs 1&2)

Ysm: years since migration; Citizen: 1 if U.S. citizen

Constant: immigrants from Central & South America; Continent Dummies are Deviations from the Constant:

Europe: Europe, Australia, New Zealand, and Canada; Africa: Africa and other countries