IBS
RESEARCH PROGRAM
ON POPULATION PROCESSES
INSTITUTE OF BEHAVIORAL SCIENCE
UNIVERSITY OF COLORADO AT BOULDER

WORKING PAPER

Inferring Directional Migration Flows From Net Migration Data: Mexico

Andrei Rogers
Junwei Liu
Virgilio Partida
Meg Tilton

February 2006

Population Program POP2006-01

Inferring Directional Migration Flows From Net Migration Data: Mexico

Andrei Rogers
Junwei Lu
Virgilio Partida
Meg Tilton

February 2006

Population Program, Institute of Behavioral Sciences, University of Colorado, Boulder, CO 80309-0484, United States.

Acknowledgements: This paper has been prepared for presentation at the February 22-25, 2006 meeting of the Western Regional Science Association in Santa Fe, New Mexico. The authors are grateful for the contributions of Dr. James Raymer, who worked on earlier stages of this body of work and to Nancy Mann for a superb job of editing the manuscript. They also wish to acknowledge the useful comments of Professor Richard Rogers. This research is being supported by a grant from the National Institute of Child Health and Human Development (R03 HD048561-02).

ABSTRACT

This paper outlines a method for indirectly estimating directional migration streams from net migration. Assuming the stability of age profiles in historical migration schedules, the method uses a regression model to predict the migration levels that, together with the assumed age profiles, best fit an observed data set. An application to Mexican census data illustrates the method.

Contents

1. INTRODUCTION

When confronted with the absence of any data on migration streams, population analysts have turned to residual methods of estimating net migration (e.g., Bogue, Hinze, and White 1982). Such methods attribute differences in population totals between two dates to natural increase and to net migration. By subtracting the known or estimated contribution of the former, they obtain the latter as a residual. The estimates produced by such techniques can be subject to considerable error where the two censuses undercount at different rates. And even where undercount levels are similar, residually estimated net migration totals accumulate many other possible errors, such as age misreporting. Nevertheless, residual methods of estimating net migration continue to be used today, generally with minor refinements in technique or data.

Imagine that the standard residual method has produced a schedule of age-specific net migration rates for each region in a multiregional system. Underlying these regional net migration rate schedules are the corresponding in- and outmigration schedules. Each schedule has both a particular level and an age profile. The corresponding net migration rate schedule then results as a function of these. If we only knew the in- and outmigration schedules, we could derive the net migration schedule they implied. Unfortunately, we are faced with the inverse problem: we can describe the net migration schedule, but not its underlying in- and out-schedules.

Castro (1985) offers a particularly appealing solution. Consider a residually estimated age-specific net migration rate schedule for region i; i(x) and o(x) are the standardized (to unit area under the curve) age-specific in- and outmigration schedules. Given the residually estimated net migration rates, and a pair of assumed model in- and outmigration schedules, one can estimate the associated in- and outmigration levels using a regression model.

How can we assume an appropriate pair of standard schedules? In this paper we borrow the age profiles of historical outmigration rate schedules. We begin by describing the data used and the context. We then explain the method, apply it to the Mexican data, and assess its performance. We conclude with an analysis of possible refinements of the method and their contribution to improved indirect estimates.

Finally before we move on to the details, the work by Plane (1981) should be mentioned. Plane introduced a net migration constrained model, based on a Minimum Information Principle, to update a migration flow table using current net migration estimates and a known migration matrix. The model relies on the overall pattern of a previous migration flow table (containing all of the regions of a census system) and seeks to minimize overall error in estimation of the current migration matrix. But it does not focus on age-specific patterns. In our research, we rely on the age profiles of the migration schedules to estimate their associated levels by assuming that the age profiles of the migration schedules, especially the underlying in- and outmigration schedules associated with the net migration schedule, are relatively stable over a certain period. Only when we expand our model to estimate a full multiregional census system, do we use the spatial structure of a past migration flow table to allocate the already estimated total out-migration flow into its destination-specific sub-flows. Our model uses age-specific net migration data and estimates age- and origin-destination-specific migration flows.

2. DATA AND CONTEXT

The data we use in this paper come from the 1990 and 2000 Censuses of Mexico. The data sets from these years are more reliable than those in previous censuses; the 1980 census had serious undercount problems (Partida 1993), and the 1960-1980 censuses featured no tables that included age, place of previous residence, and place of current residence.

As there is no official regional system in Mexico, we have divided the country into four regions on the basis of economics and history (see Figure 1). The Border region has the most formal employment. The North-Central region is an area of medium-level development, with an economy focused on manufacturing and export agriculture. The Central region, formerly the most dynamic area in Mexico, remains the country's financial and political hub. The South region, historically the country's poorest, currently has an economy based on tourism and petroleum.

Figure 1. Four Regions of Mexico

After the Mexican Revolution (1910-1921), urbanization and development occurred in two stages. In the first, which lasted until the 1960s, the country's economic policy focused on domestic consumption, and Mexico's main cities enjoyed comparative advantages that attracted rural outmigrants. In the second stage, over the last thirty years, globalization has promoted a bigger range of destinations for rural outmigrants and a more even spatial distribution of the population. Mexico City has shown a negative net migration during the last 25 years, and the city of Guadalajara has begun to give signs of following suit. Despite these changes, rural outmigration remains prevalent and proportionally significant, and urbanization continues. The socioeconomic modernization of Mexico has increased the proportion of urban residents from one in ten inhabitants at the beginning of the twentieth century to two in three at the century's end.

Migration has increasingly influenced the spatial distribution of Mexico's population as fertility has become more stable. (Although regional fertility rates once differed dramatically, an antinatalist policy adopted in 1974 has led to a remarkable decline in fertility and mortality in all regions and all social groups.) Even though internal migration levels are relatively small, their long-term impact is significant. Table 1 compares the enumerated populations of our four regions with those that would have been reached in 2000 if no interregional migration had happened since 1950 or since 1970 (that is, if each region's population had experienced only natural increase and *international* migration.)

Table 1. Mexico: Census and Expected Population Without Interregional Migration Since 1950 and Since 1970, by Region of Residence in 2000

	Population				Population distribution	
Region of Residence	Without interregional migration	Census	Difference	Proportional change	Without interregional migration	Census
			Since 1950			
Mexico	97 483 412	97 483 412		1.000	100.0	100.0
Border	12 667 484	17 066 717	4 399 233	1.347	13.0	17.5
North-Central	30 714 558	25 016 273	-5 698 285	0.814	31.5	25.7
Central	35 131 266	39 845 425	4 714 159	1.134	36.0	40.9
South	18 970 104	15 554 997	-3 415 107	0.820	19.5	16.0
			Since 1970			
Mexico	97 483 412	97 483 412		1.000	100.0	100.0
Border	14 852 269	17 066 717	2 214 448	1.149	15.2	17.5
North-Central	25 932 150	25 016 273	- 915 877	0.965	26.6	25.7
Central	39 756 978	39 845 425	88 447	1.002	40.8	40.9
South	16 942 015	15 554 997	-1 387 018	0.918	17.4	16.0

Source: Mexican population census data 1950-2000 and estimations of natural increase and net international migration of the National Population Council of Mexico (INEGI 1986, 19912, 2001)

In both periods, the Border and Central regions show net gains due to interregional migration and the North-Central and South regions net losses—especially between 1950 and 1970. Changes since 1970 have been much more moderate, reflecting the diversification of attraction poles during the past 30 years. Changes in migration flows are presented in Table 2.

Table 2. Gross Interregional Population Flows for Four Periods, 1955-2000

		Period			
From	To	1955-1960	1965-1970	1985-1990	1990-2000
Border	North-Central	51 301	79 507	140 188	151 166
Border	Central	62 268	66 813	75 626	89 325
Border	South	5 471	4 994	16 019	24 100
North-Central	Border	264 615	245 705	378 170	349 543
North-Central	Central	287 925	362 028	168 975	155 289
North-Central	South	12 953	14 276	26 926	33 822
Central	Border	51 413	52 036	200 965	303 621
Central	North-Central	56 429	80 013	310 114	249 379
Central	South	29 983	45 669	220 415	228 632
South	Border	13 461	9 634	44 513	99 145
South	North-Central	14 414	14 780	77 653	102 919
South	Central	146 435	237 841	243 877	226 277
		Classified by size ranges			
Less than 25,000		4	4	1	1
25,000 to 99,999		5	5	4	3
100,000 and over		3	3	7	8

Source: Mexican population census data 1950-2000 and estimations of natural increase and net international migration of the National Population Council of Mexico (INEGI 1986, 19912, 2001)

The first two periods fall into the era of production for domestic consumption. During each of these periods, only three flows surpassed a hundred thousand people, and two of these were directed to the Central Region, where Mexico City was experiencing high economic growth. The two more recent periods reflect a globalized, export-oriented economy. Flows were much larger absolutely, and many large flows now went to regions other than the Central one.

Changes in interregional migration can also be seen in the outmigration rates in Table 3. Two patterns are evident: the reduction of migration from the North-Central region to both the Border and Central regions, and the rising outmigration from the Central region to all three other regions.

Table 3. Crude Interregional Outmigration Rates by Region: 1955-2000

		Period				
From	To	1955-1960	1965-1970	1975-1980	1985-1990	1990-2000
Border	North-Central	1.991	2.104	3.973	2.204	1.921
Border	Central	2.343	1.71	2.216	1.178	1.128
Border	South	0.207	0.129	0.305	0.248	0.303
North-Central	Border	5.117	3.648	3.603	3.605	2.893
North-Central	Central	5.513	5.339	3.379	1.603	1.280
North-Central	South	0.243	0.207	0.401	0.251	0.276
Central	Border	0.786	0.541	1.028	1.209	1.574
Central	North-Central	0.884	0.852	1.864	1.893	1.295
Central	South	0.466	0.485	1.295	1.352	1.200
South	Border	0.479	0.261	0.492	0.692	1.295
South	North-Central	0.526	0.409	0.836	1.230	1.355
South	Central	5.361	6.712	6.527	3.917	3.002

Source: Mexican population census of 1960 (1955-1960), 1970 (1965-1970), 1990 (1985-1990) and 2000 (1995-2000) (INEGI 1986, 19912, 2001) and Partida (1993) for 1975-1980.

Mexican cities have shaped the country's migration patterns, and the changes illustrated in Table 3 result largely from Mexico City's evolving role. After being the main destination for migrants during the first 70 years of the twentieth century, the city has seen a decreased investment in industry and commerce, which in turn has slowed job creation and decreased inmigration (Table 4). Twenty years later a similar trend occurred in Guadalajara, Mexico's second largest city, which is located in the North-Central region. The city's outmigration declined slowly in 1975-1980 and 1985-1990, yet net migration during this period was still positive. During the final three periods of the twentieth century, however, inmigration declined more steeply and the net migration was negative. But such patterns are by no means universal. Although Monterrey, the country's third largest city, shows inmigration and outmigration slowly converging, it is by no means clear that net migration will ultimately turn negative (Table 4).

Internal migration has increased regional socioeconomic disparities. Instead of training the available labor force to promote development at home, the poorest states export labor to the more developed states, receiving as compensation only low remittances that merely promote further migration and perpetuate the conditions of poverty. Critical to the understanding of internal migration are these widespread socioeconomic repercussions.

Table 4. Mexico: Crude Inmigration and Outmigration Rates for Three Main Cities, 1955-2000

	Period				
Flow	1955-1960	1965-1970	1975-1980	1985-1990	1995-2000
Mexico City Inmigration	25.438	20.861	11.096	7.055	5.787
Mexico City Outmigration	3.98	3.614	12.44	10.488	7.221
Guadalajara Inmigration	20.177	27.464	21.201	19.207	7.079
Guadalajara Outmigration	6.317	10.601	9.123	8.999	7.194
Monterrey Inmigration	26.911	24.844	17.722	10.602	8.036
Monterrey Outmigration	9.957	10.500	8.867	5.692	4.175
Mexico City Net Migration	21.459	17.247	-1.344	-3.433	-1.434
Guadalajara Net Migration	13.859	16.862	12.078	10.208	-0.115
Monterrey Net Migration	16.955	14.344	8.855	4.910	3.861

Source: Partida (2003)

3. ANALYSIS

In cases where migration data are unavailable or incomplete, we can indirectly estimate age-specific directional migration flows using several techniques. One such method involves using current net migration data along with historical or theoretical age-specific outmigration schedules. These schedules describe the proportion of directional migration for each age group relative to the sum of directional migration for all age groups. Because age patterns of directional outmigration tend to be quite regular, we can use these schedules, derived from previous census data or from multiparameteric models, to make assumptions about the population for which we have incomplete or inadequate data. This technique is described in further detail below.

3.1 Preliminaries

A few relevant definitions and explanations are provided below.

The *net migration rate*, $N_i^t(x)$, is the net migration total for region *i* for age group *x* during time period (*t, t+5*) divided by the total population of *i* in age group *x* at time period *t.* This definition is unusual in that the denominator is *i*'s population at the start, rather than the middle, of the interval. Because the numerator refers to the population that has survived the time interval to be counted, here we call it the *net conditional survivorship proportion.*

The *conditional survivorship proportion*, $S_{ij}^t(x)$, describes *directional* migration propensities. $S_{ij}^t(x)$ is a fraction where the numerator is the number of people of age *x* living in region *i* at the year *t* who were living in region *j* by the year *t+5* (again, this figure is conditional upon survivorship). The denominator is $K_i^t(x)$, or the total population of age *x* living in region *i* at the start of the interval. $S_{ij}^t(x)$ thus represents the fraction of *i's* population that migrated to *j* during the (*t, t+5*) interval and survived to be counted.

While the total number of net migrants is equal to the difference between the number of inmigrants and outmigrants, this relationship does not hold for the proportions, i.e., $N_{ij} \neq S_{ij} - S_{ji}$. This is because N_{ij} also depends on the relative population size of regions *i* and *j*: if *i* is 10 times the size of *j*, and the conditional survivorship proportions are equal ($S_{ij} = S_{ji}$), clearly more people will migrate from *i* to *j* than vice versa, so that $N_{ij} \neq 0$. We can adjust for such different population stocks using a population ratio between the two regions, as follows:

$$N_i^t(x) = \frac{K_j^t(x)}{K_i^t(x)} S_{ji}^t(x) - S_{ij}^t(x) \tag{1}$$

where $K_i^t(x)$ is the population of age *x* in region *i* at the beginning of the interval *t*. According to this equation, if $S_{ij} = S_{ji}$, then $N_i(x)$ is positive if $K_j(x)/K_i(x) > 1$, 0 if $K_j(x)/K_i(x) = 1$, and negative if $K_j(x)/K_i(x) < 1$.

The proportional distribution $r_{ij}^t(x,+)$ is calculated as follows:

$$r_{ij}^t(x,+) = \frac{S_{ij}^t(x)}{S_{ij}^t(+)} \qquad \text{where } S_{ij}^t(+) = \sum_x S_{ij}^t(x) \tag{2}$$

3.2 The Linear Relationship

A dramatic rise in migration often begins around age 16 and peaks during the early twenties as individuals move to get married, attend school, or obtain jobs. A second, smaller peak occasionally occurs around retirement, when people migrate away from the family home. A final rise occurs around age 75 as individuals migrate away from the retirement community into assisted-living communities or nursing homes.

Because of these regularities, we can adopt previously derived age-specific outmigration schedules (e.g., ones from an earlier census) or indirectly estimate such schedules with a multiparameteric model and use the results to estimate directional flows in situations where directional migration data are incomplete. In order to do this, we must also have the current net migration data that we would like to break up into in- and outmigration flows. If we do not have such net migration data, we can estimate them using the standard residual method, which assumes that the population change in region *i* that is *not* due to natural increase or decrease is attributable to net migration (Bogue et al. 1982). In the case of Mexico, however, net migration data are available. From Equation 1, we know that

$$N_i^t(x) = \frac{K_j^t(x)}{K_i^t(x)} S_{ji}^t(x) - S_{ij}^t(x).$$

If we have current population data $- K_i^t(x)$ and $K_j^t(x) -$ and current net migration data, $N_i^t(x)$, for the regions we are analyzing, we can begin to solve this equation. But since we currently have one equation and two unknowns, $S_{ji}^t(x)$ and $S_{ij}^t(x)$., we need a second equation. If we have a migration schedule, one option is to use age-to-age (ATA) ratios to create a second equation. For instance, say we are interested in knowing the migration proportions for movement between Mexico's southern region and the rest of the country. The schedule should tell us the difference in migration propensities between 15- and 20-year olds, and then this difference can be used to construct a second equation. Our first equation is as follows:

$$N_S^{1995}(15) = \frac{K_{RMex}^{1995}(15)}{K_S^{1995}(15)} S_{RMex-S}^{1995}(15) - S_{S-RMex}^{1995}(15), \tag{3}$$

which gives us the following, when we substitute in values we have from census data:

$$-0.02881 = \frac{7{,}568{,}974}{1{,}462{,}308} S_{RMex-S}^{1995}(15) - S_{S-RMex}^{1995}(15).$$

And we also know that

$$N_S^{1995}(20) = \frac{K_{RMex}^{1995}(20)}{K_S^{1995}(20)} S_{RMex-S}^{1995}(20) - S_{S-RMex}^{1995}(20). \tag{4}$$

In order for the second equation to be useful to us, we must transform the $S_{RMex-S}^{1995}(20)$ and $S_{S-RMex}^{1995}(20)$ values into $S_{RMex-S}^{1995}(15)$ and $S_{S-RMex}^{1995}(15)$. If we knew the ATA ratios $r_{S-RMex}^{1995}(20,15)$ and $r_{RMex-S}^{1995}(20,15)$, we could then replace the $S_{RMex-S}^{1995}(20)$ and $S_{S-RMex}^{1995}(20)$ values above with $S_{RMex-S}^{1995}(15)$ x $r_{RMex-S}^{1995}(20,15)$ and $S_{S-RMex}^{1995}(15)$ x $r_{S-RMex}^{1995}(20,15)$, respectively. Since we do not know these values for 1995, we will use the 1985 ATAs as an approximate estimate. We obtain the ratios using the following formula:

$$r_{ij}^t(x+5,x) = \frac{S_{ij}^t(x+5)}{S_{ij}^t(x)}, \text{ so } r_{S-RMex}^{1985}(20,15) = \frac{S_{S-RMex}^{1985}(20)}{S_{S-RMex}^{1985}(15)} = \frac{0.039426}{0.053233} = 0.740638.$$

And for the reverse flow we obtain

$$r_{RMex-S}^{1985}(20,15) = 1.320899.$$

We can then begin to solve equation (4) as follows:

$$N_S^{1995}(20) = \frac{K_{RMex}^{1995}(20)}{K_S^{1995}(20)} S_{RMex-S}^{1995}(15) \text{ x } 1.320899 - S_{S-RMex}^{1995}(15) \text{ x } 0.740638.$$

Substituting the values we know, we obtain:

$$-0.00718 = \frac{6{,}903{,}236}{1{,}196{,}265} S_{RMex-S}^{1995}(15) \text{ x } 1.320899 - S_{S-RMex}^{1995}(15) \text{ x } 0.740638. \tag{5}$$

Solving (3) and (5), we obtain $\hat{S}_{S-RMex}^{1995}(15) = 0.048156$ and $\hat{S}_{RMex-S}^{1995}(15) = 0.003737$. The observed values, for comparison purposes, are 0.05163 and 0.004409.

But the ATA method can produce unrealistic and even negative estimates of survivorship proportions if the ATA ratios differ considerably over time. A better solution may be to use the ratio between one age group and the sum of all age groups, so that instead of two equations and two unknowns, we obtain least-squares estimates of the migration levels that, along with the age profiles, produce the "best" estimated conditional survivorship proportions. This method, developed by Castro (1985), modifies equation (1) as follows:

$$N_i^t(x) = \frac{K_j^t(x)}{K_i^t(x)}\left[S_{ji}^t(+)\right]r_{ji}^t(x,+) - \left[S_{ij}^t(+)\right]r_{ij}^t(x,+) \qquad (6)$$

where $S_{ij}^t(+) = \sum_{x=1}^{n} S_{ij}^t(x)$ and the age profile is defined by the proportional distribution

$r_{ij}^t(x,+) = \frac{S_{ij}^t(x)}{S_{ij}^t(+)}$, as noted in the section on preliminaries. Thus $S_{ij}^t(x) = S_{ij}^t(+)r_{ij}^t(x,+)$,

which is the relationship that was substituted into Equation (1) to obtain Equation (6).

Given the values of $K_i^t(x)$, $K_j^t(x)$, and $N_i^t(x)$ for a particular census, and adopting the *preceding* census's values of $r_{ij}^t(x,+)$, we have as many equations as age groups (in our example, 12 of them). We cannot solve 12 equations for two unknowns, since it is unlikely they will give consistent values of $S_{ij}^t(+)$ and $S_{ji}^t(+)$. However, we can perform a regression analysis with two independent variables and no intercept term (i.e., a model fitted through the origin) to derive the least-squares estimates of $S_{ij}^t(+)$ and $S_{ji}^t(+)$. The regression equation is as follows:

$$N_i^t(x) = \beta_1 X_{1i} + \beta_2 X_{2i}, \qquad (7)$$

where

$N_i^t(x)$	=	the age-specific net migration in region *i* at time *t*
β_1	=	our first parameter, which is our estimate of $S_{ji}^t(+)$
X_{1i}	=	$\frac{K_j^t(x)}{K_i^t(x)} r_{ji}^t(x,+)$

β_2 = our second parameter, which is our estimate of $S_{ij}^{t}(+)$

X_{2i} = $r_{ij}^{t}(x,+)$.

Example. Consider the previously discussed case, where we want conditional survivorship proportions for migration between Mexico's southern region and the rest of the country. We can assume we have net migration data for 1995-2000, and an age-specific migration schedule derived from 1985-1990 migration data. As our dependent variable, we use the 1995-2000 net migration values observed in the South [i.e., $N_{S}^{1995}(x)$]. For our two independent variables, X_{1i} and X_{2i}, we use $\frac{K_{RMex}^{1995}(x)}{K_{S}^{1995}(x)} r_{RMex-S}^{1995}(x,+)$ and $r_{S-RMex}^{1995}(x,+)$.

Performing the regression using data for all 12 age groups, we obtain estimates of $S_{S-RMex}^{1995}(+) = 0.324062$ and $S_{RMex-S}^{1995}(+) = 0.042219$. (For comparison, the observed values are 0.30510 and 0.04026). These estimates of $S_{ij}^{t}(+)$ can then be disaggregated into estimates for individual age groups, $S_{ij}^{t}(x)$. For instance, if x=15 in 1995, we obtain

$$\hat{S}_{S-RMex}^{1995}(15) = \beta_2 \text{ x } r_{S-RMex}^{1985}(15,+) = 0.324062 \text{ x } 0.16256 = 0.05268$$

and

$$\hat{S}_{RMex-S}^{1995}(15) = \beta_1 \text{ x } r_{RMex-S}^{1985}(15,+) = 0.042219 \text{ x } 0.10390 = 0.00439.$$

Note that these values are more accurate estimates of the observed values of $S_{S-RMex}^{1995}(15) = 0.05163$ and $S_{RMex-S}^{1995}(15) = 0.004409$ than were the estimates yielded by the ATA ratio method (where $\hat{S}_{S-RMex}^{1995}(15) = 0.048156$ and $\hat{S}_{RMex-S}^{1995}(15) = 0.003737$).

For each age-specific directional migration flow, we can construct two regression equations and two estimates of $S^t_{ij}(x)$ and $S^t_{ji}(x)$. Consider the previous example. As we've done, we can calculate $\hat{S}^{1995}_{S-RMex}(+)$ and $\hat{S}^{1995}_{RMex-S}(+)$ using the following equation:

$$N^{1995}_{S}(x) = \hat{S}^{1995}_{RMex-S}(+) \ x \left[\frac{K^{1995}_{RMex}(x)}{K^{1995}_{S}(x)} r^{1985}_{RMex-S}(x,+) \right] - \hat{S}^{1995}_{S-RMex}(+) \ x \ r^{1985}_{S-RMex}(x,+) .$$

However, we can also calculate these estimates using the equation

$$N^{1995}_{RMex}(x) = \hat{S}^{1995}_{S-RMex}(+) \ x \left[\frac{K^{1995}_{S}(x)}{K^{1995}_{RMex}(x)} r^{1985}_{S-RMex}(x,+) \right] - \hat{S}^{1995}_{RMex-S}(+) \ x \ r^{1985}_{RMex-S}(x,+) , \tag{8}$$

which yields slightly different values. For instance, if we revisit our test case of x=15, the second formula yields $\hat{S}^{1995}_{S-RMex}(15) = 0.05808$ (as opposed to 0.05268 in the original formula) and $\hat{S}^{1995}_{RMex-S}(15) = 0.00488$ (as opposed to 0.00439).

3.3 Optimizing Areal Units: Biregional vs. Multiregional Systems

Curiously, the model makes more accurate predictions for biregional systems than for multiregional ones (Rogers and Liu 2005). As a result, in cases like Mexico where population data are broken up into more than two regions, we aggregate the data into two regions for our analysis. Essentially, we pretend that each of the four regions is part of a two-region system composed of itself and the rest of Mexico. We then model directional migration separately for each of these four "biregional" systems.

We now have overall estimates for $S_{ij}(x)$ for each region in Mexico, where *i* = the region in question (i.e., one of the four original regions) and *j* = the rest of Mexico. What we wish to obtain are $S_{ij}(x)$ values where *i* = the region in question and *j* = any of the other three regions. If we assume that the three migration proportions between *i* and the other regions remain constant over time, we can use past ratios to obtain current estimates of $S_{ij}(x)$ for directional migration between any two of Mexico's four regions. For instance, if we wish to estimate 1995-2000 interregional migration on the basis of 1985-1990 flows, we do this as follows:

$$\hat{S}_{ij}^{1995} = \frac{S_{ij}^{1985}}{S_{i+}^{1985}} * S_{i+}^{1995}, \tag{9}$$

where $S_{i+}^{t} = \sum S_{ij}^{t}$ for all $j \neq i$. But obviously, this method is subject to error if interregional migration patterns change from one census to the next.

3.4 Outlying and Influential Observations

We can identify outlying observations using confidence intervals, and then rerun the regression equations using only values that lie within the confidence interval. In this situation, an age-group is considered a case, and certain age groups can be deleted from the regression in order to obtain new estimates. This strategy allows us to see if deleting outlier cases improves the model. Confidence intervals can be calculated at the 99%, 95%, 90%, and 80% significance levels.

We can also identify age-group cases that have an exceptionally strong influence on the regression equations, to see if throwing out age groups that repeatedly surface as influential might improve the model. Influence statistics like the Cook's and the dfbeta statistic prove useful in this analysis, since high values in either measure indicate a high level of influence. To calculate the Cook's statistic, the dfbeta statistic, and the standard error of mean predictions (used in calculating the confidence interval), we use SPSS.

4. PREDICTION

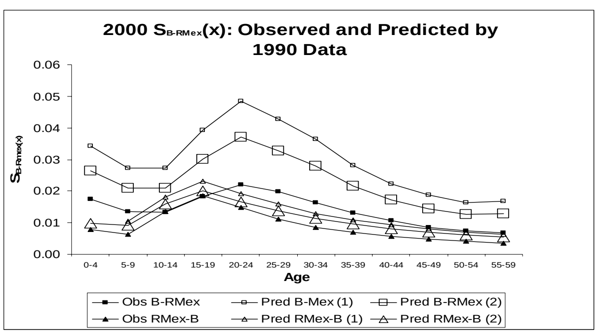
The results for the Mexican data varied considerably by region. The model performed best in the South and worst in the Border region. In both the Border and the South, the model overestimated migration proportions; in the North-Central and Central regions, it underestimated them.

4.1 Estimated Aggregate Survivorship Proportions

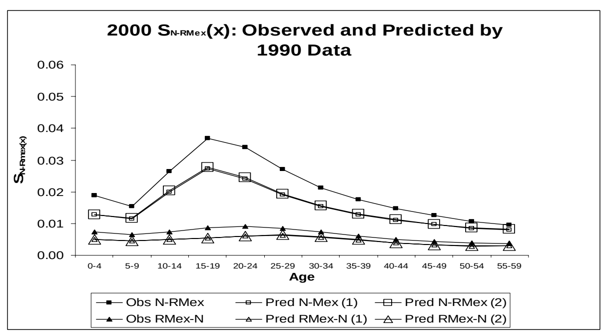
The observed and two predicted conditional survivorship proportions for each region to the rest of Mexico are illustrated in Figure 2 below, along with their corresponding observed values. Predictions (1) and (2) refer to the two regressions given by Equations 6 and 8. Values given are for the 1995-2000 period. The X axis displays the age in 1995, that is, the age at which the migrants were initially exposed to migration. The Y axis displays the associated outmigration conditional survivorship values.

Figure 2. Observed and Predicted 2000 Conditional Survivorship Proportions from Mexican Regions

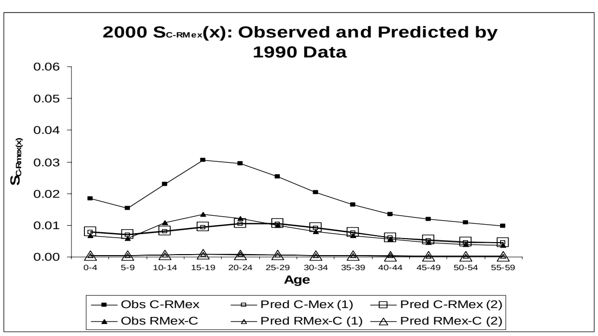
a. Outmigration from and to the Border Region



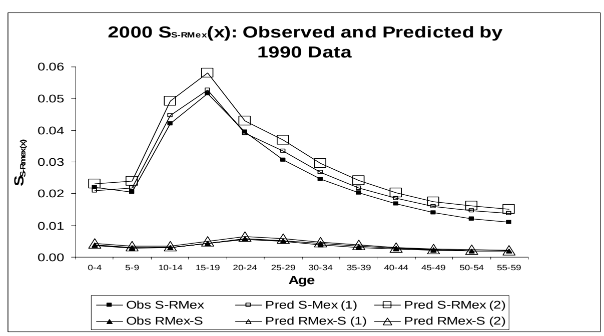
b. Outmigration from and to the North-Cental Region



c. Outmigration from and to the Central Region



d. Outmigration from and to the South Region



As we noted above, estimates for the Border differ dramatically from observed values. For instance, for 20-to-25-year-olds, the first equation's estimates are almost twice as large as the observed value. The two sets of estimates also differ greatly among themselves. Predictions for the South, in contrast, are much closer to the observed values and to each other.

Estimates underpredict the outflows for both the North-Central and, especially, the Central regions. For 15-to-19-year-olds, the observed migration proportion from the Central to the rest of Mexico region is over three times greater than either estimated value. Estimated proportions for the reverse flow, that from the rest of Mexico to Central, never reach .001 for any age group, whereas observed values climb above .01. Some of the error may be from the Central region's unstable $r_{ij}(x)$ proportions, which we discuss later in the paper.

In every case except the South, the model's predictions of migration from the region in question to the rest of Mexico were actually worse predictors of the observed 2000 rates than were the observed rates from 1990.

We also evaluated the model's predictions by examining the Mean Absolute Percent Error (MAPE) statistics for each region. MAPE statistics were calculated as follows:

$$\sum_{x=1}\left[\frac{\left|\hat{S}_{ij}^{t}(x)-S_{ij}^{t}(x)\right|}{S_{ij}^{t}(x)} \text{ x } \frac{K_{i}^{t}(x)}{K_{i}^{t}(+)}\right] \text{x } 100$$

Again, the results varied considerably by region. In general, however, MAPE statistics for conditional survivorship proportions were less encouraging than the high R^2 values might suggest. Averaging the two MAPE scores for all four regions yielded an overall MAPE of 46.07. For comparison purposes, the average MAPE score for the four U.S. regions (Northeast, Midwest, South, West) was only 25.77 (Rogers and Liu 2005). In the Mexico analysis, predictions were most accurate for the North-Central and South regions, and least accurate for the Border region, as can be seen in Table 3 (p. 19).

4.2 Estimated Destination-Specific Survivorship Proportions

In the second step, we disaggregated the estimated age-specific migration flows to the Rest-of-Mexico during 1995-2000 into three underlying destination-specific flows, using the migration flow proportions during 1985-1990. This is another potential source of error.

For every region except the South, the model's destination-specific predictions were also worse than simply reproducing the 1990 destination-specific survivorship proportions. For the South, however, the model's predictions were as good as the 1990 proportions, or better. MAPE scores for aggregate and destination-specific migration proportions are listed in Table 5 below.

Table 5. MAPE Scores for Estimation of Outmigration Conditional Survivorship Proportions from the Mexican Regions

Flow	MAPE (%)	
	Regression 1	Regression 2
Border – Rmex	112.09	62.58
Border – North-Central	123.59	71.39
Border – Central	106.98	58.66
Border – South	60.88	23.32
North-Central – Rmex	26.72	25.76
North-Central – Border	25.15	24.17
North-Central – Central	26.01	25.04
North-Central – South	44.83	44.10
Central – Rmex	59.62	58.71
Central – Border	70.66	70.00
Central – North-Central	46.36	45.16
Central – South	59.48	58.56
South – Rmex	7.07	16.05
South – Border	42.82	36.96
South – North-Central	8.44	8.69
South – Central	32.21	45.78

Variance in the ratios of destination-specific flows accounts for some of the error in the disaggregated survivorship proportions. (For details on how such ratios varied from 1985 to1990 and from 1995 to 2000, see the section on assumptions, below.) For example, while the model predicts migration from the South to the Border, North-Central, and Central regions more accurately than simply using the 1985 survivorship proportions, these destination-specific estimates are still less accurate than the aggregate estimate for migration from the South to the Rest of Mexico.

Paradoxically, however, errors in disaggregating the flows can also improve overall estimates. This is because errors made in predicting aggregate flow can be offset by errors made in disaggregating the migration proportions into destination-specific units. As Liu (2005) points out, such beneficial offsetting effects can happen in two cases: where an overpredicted aggregate flow combines with an underpredicted destination-specific migration ratio, as in the case for migration from the Border to the South, or when an underpredicted aggregate flow combines with an overestimated destination-specific migration ratio, as is the case from the Central to the North-Central region.

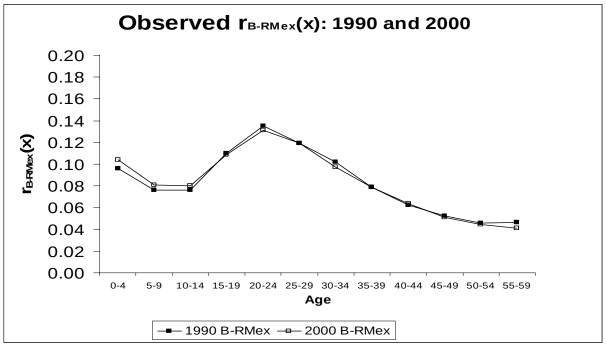
4.3 Assumptions

The model made several key assumptions. Evidence for the validity of these assumptions is examined below.

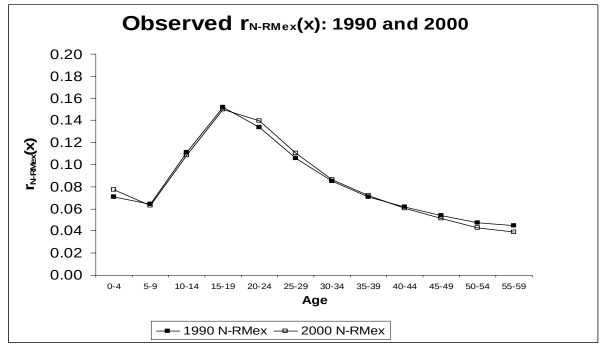
One of the model's assumptions is that $r_{ij}^{t-10}(x,+) \approx r_{ij}^{t}(x,+)$. Figure 3 shows how reliable this assumption was for the Mexican populations in 1990 and 2000.

Figure 3. Observed Age-to-Age-Summed (ATAS) Migration Ratios for Flows from One Particular Mexican Region to the Rest of Mexico: 1985-1990 and 1995-2000

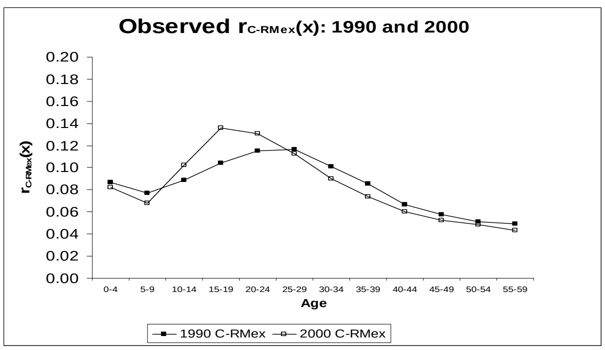
a. Outmigration from the Border Region



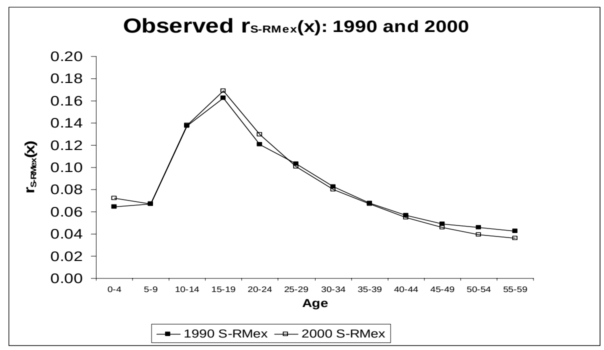
b. Outmigration from the North-Central Region



c. Outmigration from the Central Region



d. Outmigration from the South Region



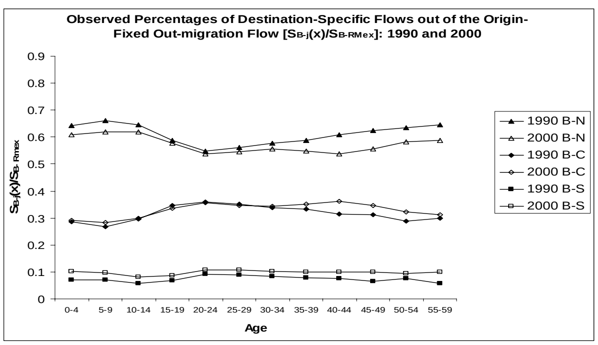
Clearly, the assumption that $r_{ij}^{t-10}(x,+) \approx r_{ij}^{t}(x,+)$ is not supported for the Central region. This is undoubtedly one reason that the model performed poorly in this area. In the other regions, however, the assumption proved valid. This explains why estimates for most of the regions exhibited the correct age *pattern*, even if the estimation *levels* were inaccurate.

The model also assumes that we can construct a relatively robust regression model to estimate $S_{ij}^{t}(+)$ and $S_{ji}^{t}(+)$. This assumption proved valid, as adjusted R^2 values proved quite high, ranging from 0.963 to 0.990.

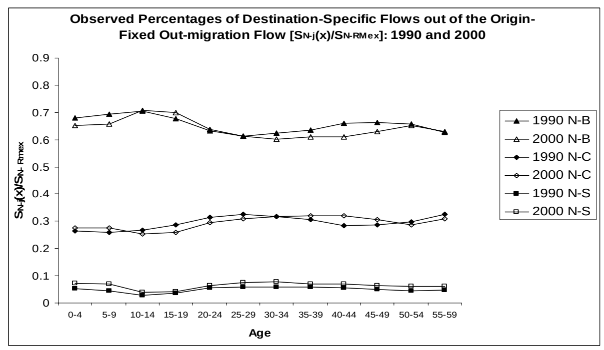
When we disaggregated the survivorship proportions back into individual regions, we made a final assumption: that migration proportions between each region *i* and the other three regions remained relatively constant from one census to the next. This assumption proved valid for the Border and North-Central regions. It proved less so for the Central region, which saw outmigration to the Border increase dramatically relative to outmigration to the North-Central region. The assumption also did not hold well for the South, which saw outmigration to the Central region fall relative to outmigration to the Border and North-Central regions. Figure 4 illustrates these trends in greater detail.

Figure 4. Destination-Specific Flows as Percentages of Total Outmigration (Observed) for Each Region of Origin, 1990 and 2000

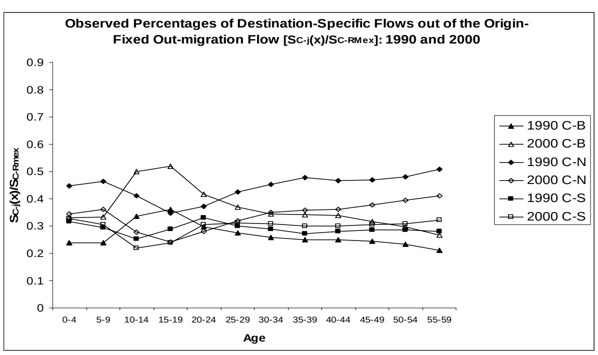
a. Outmigration from the Border Region



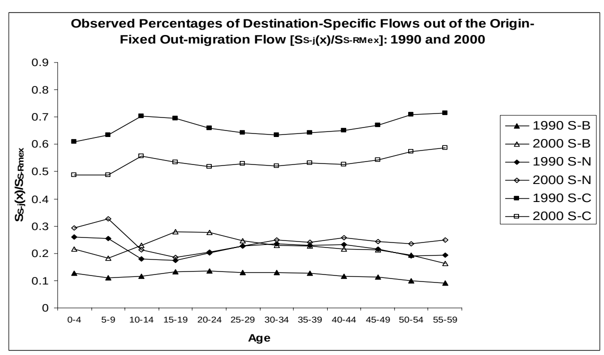
b. Outmigration from the North-Central Region



c. Outmigration from the Central Region



d. Outmigration from the South Region



5. DISAGGREGATION BY GENDER

It seems likely that the above estimates could be improved if more information about a population were available–for example, if migrants were classified by gender. Because homogeneous subpopulations exhibit similar behaviors, the estimation of their migration characteristics should be more accurate, and aggregating those separately estimated migration characteristics should in turn produce a more accurate overall estimate. Our discussion, below, of estimation with disaggregation by gender draws on thesis work by Liu (2005). But, as will become evident, the results were somewhat inconclusive.

With a subscript *g* representing the gender of the population, Equation 4 can be modified as:

$$ {}_{g}N_{i}^{t}(x) = \frac{{}_{g}K_{j}^{t}(x)}{{}_{g}K_{i}^{t}(x)}\left[{}_{g}S_{ji}^{t}(+)\right]{}_{g}r_{ji}^{t}(x,+) - \left[{}_{g}S_{ij}^{t}(+)\right]{}_{g}r_{ij}^{t}(x,+)\,. \qquad (10) $$

The subscript *g* can denote either males or females, with the age-specific outmigration conditional survivorship proportions being estimated separately for each. They can then be weighted by the ratios of their corresponding populations, and aggregated at the age-group level:

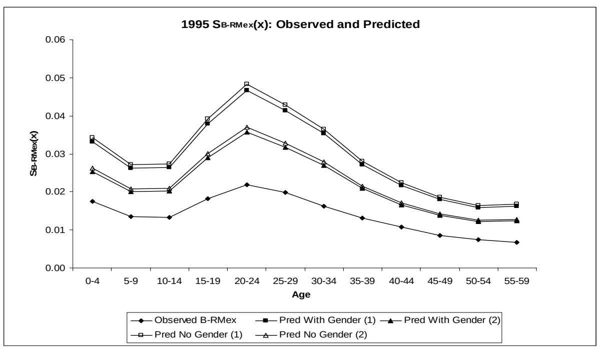
$$ S_{ij}^{t}(x) = \sum_{g}\left[\left({}_{g}K_{i}^{t}(x)\Big/\sum_{g}{}_{g}K_{i}^{t}(x)\right){}_{g}S_{ij}^{t}(x)\right]. \qquad (11) $$

Again, the subscript *i* can refer to either of the two regions in a biregional system; this means two sets of estimated age-specific directional migration conditional survivorship proportions can be obtained for both male and female subpopulations. As these two formulas often yielded considerably different estimates in the U.S. analysis, we calculated both sets of models for the Mexico case study to get “best” estimates. Again, we proceed using a biregional system, and we then decompose our aggregate estimate of total outmigration into three destination-specific flows, obtaining in the end mixed results.

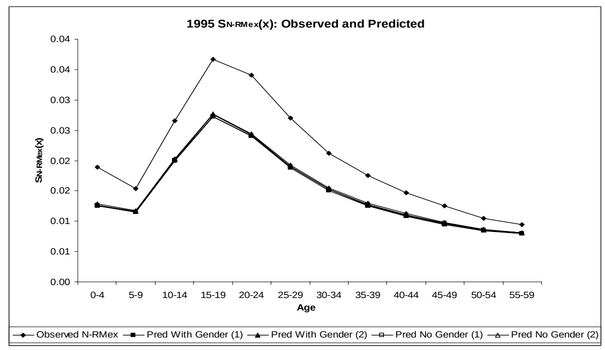
The gender-specific estimates were better than the aggregate estimates for the Border region, but worse for the Central region. For the remaining two regions, there was little distinction between the gender-specific and aggregate estimates. Figure 5 shows the observed values for Sij(x) along with the estimates predicted by both the gender-specific and non-gender-specific models.

Figure 5. Conditional Survivorship Proportions: Predictions With and Without Gender

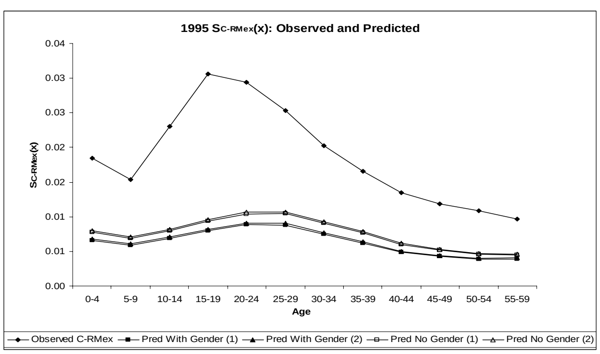
a. Outmigration from the Border Region



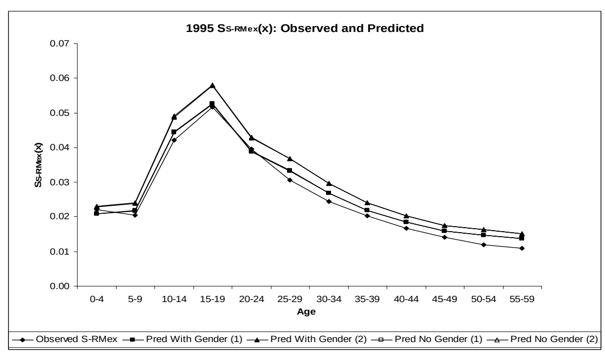
b. Outmigration from the North-Central Region



c. Outmigration from the Central Region



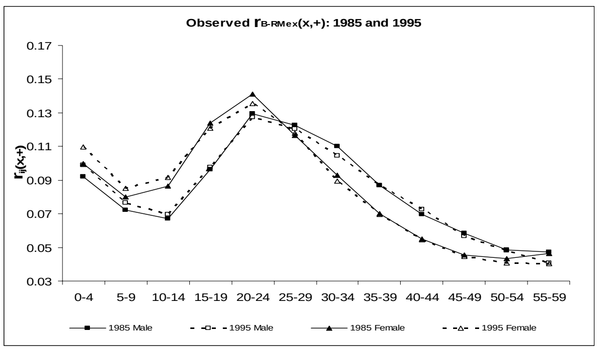
d. Outmigration from the South Region



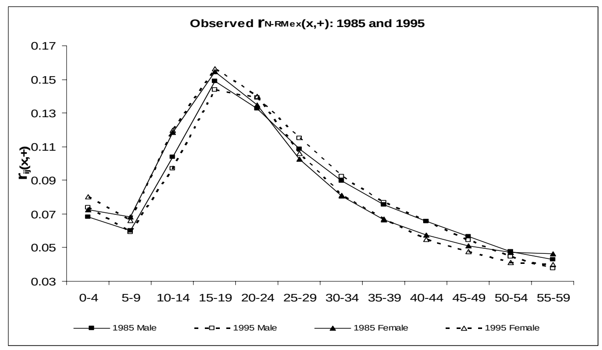
As in the original model, the gender-specific model assumes that age profiles remain stable over time. The age profiles of the four origin-specific flows are shown in Figure 6 below. The patterns during the two migration periods (1985-1990 and 1995-2000) are very stable, the only difference being that the age patterns now are gender-specific, with females having higher migration propensities before and around the same peak migration age, and lower migration propensities after the peak migration age, than males. Such differences support our decision for performing a gender-specific analysis.

Figure 6. Observed Gender-Specific Age-to-Age-Summed (ATAS) Migration Ratios for Flows from One Particular Mexican Region to the Rest of Mexico: 1985-1990 and 1995-2000

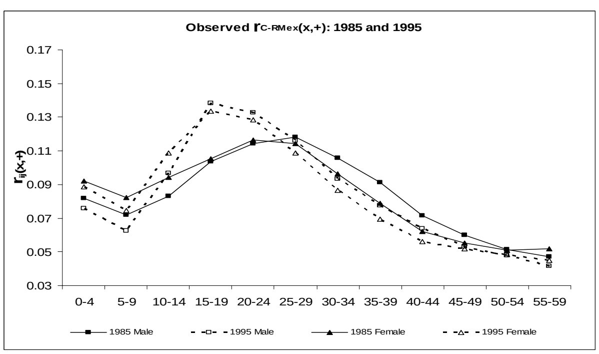
a. Outmigration from the Border Region



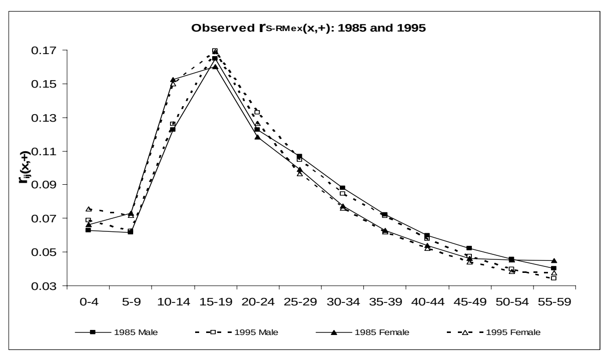
b. Outmigration from the North-Central Region



c. Outmigration from the Central Region



d. Outmigration from the South Region



When we disaggregate the origin-specific flows into their three respective destinations, we once again find no obvious differences between estimates that use gender information and those that ignore it (Liu 2005). This is to be expected, since estimates of aggregate flows are quite similar and these aggregate flows are broken down into more geographically specific flows using the same technique as before.

The MAPE error indices listed in Table 6 illustrate the effects of the disaggregation by gender. The error levels of the estimates of the outmigration flows from the Border region are between 8% and 18% lower for those "best" fits that corresponded to the "best" fits obtained without using gender-specific disaggregation. Errors for the Central region, in contrast, are higher than those for the aggregate analysis. The error levels for the North-Central and South regions are just about the same as the error levels of those estimates when no gender information is used. This similarity is apparent in both regression equations.

Table 6. MAPE Scores for Estimation of Outmigration Conditional Survivorship Proportions from the Mexican Regions Using Gender-Specific Data

Flow	MAPE (%)	
	Regression 1	Regression 2
Border – Rmex	105.08	56.94
Border – North-Central	116.20	65.46
Border – Central	100.13	53.15
Border – South	55.56	19.05
North-Central – Rmex	27.35	26.51
North-Central – Border	25.80	24.95
North-Central – Central	26.61	25.77
North-Central – South	45.30	44.67
Central – Rmex	66.11	65.09
Central – Border	75.39	74.64
Central – North-Central	54.97	53.61
Central – South	65.95	64.93
South – Rmex	6.84	15.59
South – Border	43.03	37.21
South – North-Central	8.67	8.73
South – Central	31.72	45.19

Unfortunately, the overall accuracy level of the estimation process using gender data is still not high. However, as in the case without using gender data, several of the destination-specific flows—Border-South, North-Central–Border, North-Central–Central and South–North-Central—are estimated quite accurately.

6. CONCLUSTION

Our efforts, in this paper, to replicate with Mexican data own earlier indirect estimation procedure using U.S. data, have had only limited success. The net migration method apparently is very sensitive to violations of the assumptions underlying that method. Further research is needed to identify profitable refinements that would improve the accuracy of the estimates obtained. For example, we have explored the potential improvements showing from a disaggregation by birthplace, but the results indicate that no improvement follows from such a disaggregation. Perhaps the net migration method should be combined with the two-census survival method examined in Rogers and Raymer (2005). Since the latter uses no prior information of the migration to constrain its estimation, it would seem that adding net migration controls could improve matters. Further work is proceeding to test this hypothesis.

APPENDIX A: Enlarging the Number of Age Groups

After we had completed our main analysis, more extensive Mexican census data became available that listed 19 age groups rather than 12. Whereas our original data set ended at the category of 65 and over, the newer data set categorized residents up to 95 and over. We reran the original analysis to determine whether the increased specificity of the second data set would improve our estimates. In three of the four regions, MAPE scores for the two data sets were relatively similar. For the Central region, however, the 19-group data set had a considerably stronger performance than the 12-group set. Full results appear in Table A1 below.

Table A1. MAPE Scores for Data Sets With 12 and 19 Age Groups

FLOW	12 Age Groups		19 Age Groups	
	Equation 1	Equation 2	Equation 1	Equation 2
Border – RMex	112.09	62.58	123.80	75.27
Border – North	123.59	71.39	135.85	84.70
Border – Cent.	106.98	58.66	118.45	71.08
Border – South	60.88	23.32	69.03	32.44
North – RMex	26.72	25.76	22.68	20.04
North – Border	25.15	24.17	21.08	18.38
North – Cent.	26.01	25.04	21.95	19.37
North – South	44.83	44.10	41.86	39.88
Cent. – RMex	59.62	58.71	30.31	31.26
Cent. – Border	70.66	70.00	48.54	49.25
Cent. – North	46.36	45.16	9.08	10.08
Cent. – South	59.48	58.56	30.77	31.72
South – RMex	7.07	16.05	6.74	13.70
South – Border	42.82	36.96	43.74	38.16
South – North	8.44	8.69	9.73	8.11
South – Central	32.21	45.78	29.07	41.87

APPENDIX B: Indirect Estimation of Migration Flows from Age-Specific Net Migration Proportions: Implementation using Excel Spreadsheets

by
Meg Tilton

Introduction

This appendix explains how to use Excel to indirectly estimate interregional flows using age-specific net migration proportions. The method described here estimates such flows using a regression procedure with two inputs: current age-specific *net* migration proportions and age-specific directional migration proportions from an earlier census or a multi-parameter model. Readers are advised to read through the text first, before attempting to carry out the calculations.

Constructing the Spreadsheet

Typically, census data provides total numbers of migrants by age group, rather than proportions such as $S_{ij}^{t}(x)$ and $N_{i}^{t}(x)$. In order to calculate these proportions it is useful to organize the data into an Excel table, using the rows to delineate age groups and the columns to list variables that will be required in the estimation process. Figure B1 illustrates this set-up.

Figure B1. Setting up Survivorship Values in Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	**Migration between the Border Region and the Rest of Mexico: 1990 and 2000**													
2														
3														
4				Observed							Observed Inputs		Predicted Inpu	
5								$K_j(x)$ /			$K_j(x)$ /		$K_j(x)$ /	
6		Region		Region of Destination			$K_j(x)$ /	$K_i(x)$ *			$K_i(x)$ *		$K_i(x)$ *	
7	Period	of Origin	Age	Border	Rmex	Total (K)	$K_i(x)$	$S_{ji}(x)$	$S_{ij}(x)$	$N_i(x)$	$r_{ji}(x)$	$r_{ij}(x)$	$r_{ji}(x)$	$r_{ij}($
8														
9	1985-1990	Border	0-4	1497893	26762	1,524,655	5.91439	0.04575	0.01755	0.02820	0.43778	0.09582		
10			5-9	1546803	21824	1,568,627	5.61346	0.04029	0.01391	0.02638	0.38551	0.07595		
11			10-14	1556774	22105	1,578,879	5.11196	0.06371	0.01400	0.04971	0.60962	0.07643		
12			15-19	1309755	26860	1,336,615	4.84108	0.07665	0.02010	0.05656	0.73343	0.10971		
13			20-24	1076278	27295	1,103,573	4.78150	0.06259	0.02473	0.03786	0.59891	0.13502		
14			25-29	915254	20482	935,736	4.73543	0.05148	0.02189	0.02959	0.49254	0.11949		
15			30-34	741839	14102	755,941	5.03746	0.04458	0.01865	0.02593	0.42657	0.10184		
16			35-39	585780	8557	594,337	4.86844	0.03636	0.01440	0.02197	0.34793	0.07860		
17			40-44	500736	5790	506,526	4.85342	0.03087	0.01143	0.01944	0.29536	0.06240		
18			45-49	417662	4025	421,687	4.66444	0.02587	0.00954	0.01632	0.24748	0.05211		
19			50-54	322380	2736	325,116	4.81460	0.02312	0.00842	0.01470	0.22119	0.04594		
20			55-59	268324	2314	270,638	4.93941	0.02131	0.00855	0.01276	0.20392	0.04668		
21			Total	10,739,478	182,852	10,922,330			0.18318			1.00000		
22														
23		Rest of	0-4	69759	8947651	9,017,410	0.16908	0.00297	0.00774	-0.00477	0.01620	0.07402		
24		Mexico	5-9	63202	8742227	8,805,429	0.17814	0.00248	0.00718	-0.00470	0.01353	0.06868		
25			10-14	100596	7970576	8,071,172	0.19562	0.00274	0.01246	-0.00972	0.01495	0.11925		
26			15-19	102456	6368202	6,470,658	0.20657	0.00415	0.01583	-0.01168	0.02266	0.15150		
27			20-24	69078	5207661	5,276,739	0.20914	0.00517	0.01309	-0.00792	0.02824	0.12526		
28			25-29	48169	4382944	4,431,113	0.21117	0.00462	0.01087	-0.00625	0.02523	0.10401		
29			30-34	33702	3774323	3,808,025	0.19851	0.00370	0.00885	-0.00515	0.02022	0.08468		
30			35-39	21612	2871880	2,893,492	0.20540	0.00296	0.00747	-0.00451	0.01614	0.07147		
31			40-44	15636	2442745	2,458,381	0.20604	0.00236	0.00636	-0.00401	0.01286	0.06086		
32			45-49	10907	1956026	1,966,933	0.21439	0.00205	0.00555	-0.00350	0.01117	0.05306		
33			50-54	7516	1557789	1,565,305	0.20770	0.00175	0.00480	-0.00305	0.00954	0.04594		
34			55-59	5768	1331025	1,336,793	0.20245	0.00173	0.00431	-0.00258	0.00945	0.04128		
35			Total	548,401	55,553,049	56,101,450			0.10451			1.00000		
36														

In this example, i = the Border and j = the rest of Mexico. The census data has provided us with values for columns D and E. From these original data, we calculate the remaining columns according to the Table B1 below. Note that it is easier to calculate the columns in the order listed in the table rather than in alphabetical order (e.g., you should calculate column I before column H). Also note that the formulas given in the final column of the table pertain specifically to the figure above; you will obviously need to change the row numbers to fit your own table.

Table B1. Data and Variables Used in Net Migration Method for Estimating Directional Migration Flows

Column	Description	Sign	Calculation Process	Excel Formula (for row 9; formulas are relative unless noted with a $)
F	Total population	$K_i(x)$	Add numbers of migrants to the population who stayed	=SUM(D9:E9)
G	Population Ratio for regions *i* and *j*	$\frac{K_j(x)}{K_i(x)}$	Divide the population of *j* by the population of *i*	=F23/F9
I	Conditional survivorship proportion	$S_{ij}(x)$	Divide the number of migrants from *i* to *j* by *i*'s total population	=E9/F9
H	Population-adjusted survivorship proportion	$\frac{K_j(x)}{K_i(x)} S_{ji}(x)$	Multiply the *j-i* population ratio by *j*'s survivorship proportion	=G9*I23
Cell I21*	Sum of age-specific survivorship proportions	$\sum_x S_{ij}(x)$	Sum all the age-specific survivorship proportions	=SUM(I9:I20)
J	Net conditional survivorship proportion	$N_i(x)$	Subtract outmigration from inmigration, adjusting for population as follows: $N_i(x)=\frac{K_j(x)}{K_i(x)} * S_{ji}(x) - S_{ij}(x)$	=H9-I9
K	Proportional distribution [better name?]	$r_{ij}(x)$	Divide $S_{ij}(x)$ by $\sum_x S_{ij}(x)$	=I9/I$21
L	Population-adjusted proportional distribution	$\frac{K_j(x)}{K_i(x)} r_{ji}(x)$	Multiply *j*'s proportional distribution by the *j-i* population ratio	=G9xL23
M	Population-adjusted proportional distribution (Predicted Input 1)	$\frac{K_j(x)}{K_i(x)} r_{ji}(x)$	Multiplying *j*'s proportional distribution (the one calculated for the first data set*) by the current *j-i* population ratio	= G37xL23
N	Proportional distribution (Predicted input 2)	$r_{ij}(x)$	Using the proportional distribution calculated for the first data set	=L9

* This obviously does not have to be cell I21; it should be the first free cell in your chart that is below all the age-specific survivorship proportions.

Typically, we are working with data from two different censuses. The earlier set provides our model migration schedule, while the second set (which at least in theory is more limited) provides us with the net migration data that we wish to break down into destination-specific flows. In this situation, it is useful to construct the Excel spreadsheet with the earlier data on top and the later data on the bottom. The A-L columns are identical in both cases. For the second data set, however, some additional calculations are necessary. First, we add two columns, Columns M and N.

What we really need, however, is not the population ratio for the end of the second time interval, but the one at the interval's start. We would also like to recalculate the M-column values accordingly. In order to backcast the ratio to the beginning of the interval, we average the 1990 and 2000 population figures, both of which we have. In our example, we do this by adding several rows beneath our original two data sets, as follows in Figure B2.

Figure B2. Calculating Population Ratios

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	**Migration between the Border Region and the Rest of Mexico: 1990 and 2000**														
2															
3															
4							Observed				Observed Inputs		Predicted Inputs		
5								$K_j(x)$ /			$K_j(x)$ /		$K_j(x)$ /		
6		Region		Region of Destination			$K_j(x)$ /	$K_i(x)$ *			$K_i(x)$ *		$K_i(x)$ *		Pre
7	Period	of Origin	Age	Border	Rmex	Total (K)	$K_i(x)$	$S_{ji}(x)$	$S_{ij}(x)$	$N_i(x)$	$r_{ji}(x)$	$r_{ij}(x)$	$r_{ji}(x)$	$r_{ij}(x)$	S_{ij}
64															
65															
66	1995		2 1/2			1,694,006	5.407998						0.40029	0.09582	
67	Border		7 1/2			1,633,860	5.452482						0.37445	0.07595	
68			12 1/2			1,646,920	4.956889						0.59112	0.07643	
69			17 1/2			1,527,109	4.513213						0.68375	0.10971	
70			22 1/2			1,348,687	4.368117						0.54713	0.13502	
71			27 1/2			1,162,409	4.356462						0.45312	0.11949	
72			32 1/2			972,754	4.593562						0.38898	0.10184	
73			37 1/2			777,473	4.570201						0.32661	0.07860	
74			42 1/2			614,667	4.712968						0.28681	0.06240	
75			47 1/2			511,342	4.608954						0.24454	0.05211	
76			52 1/2			392,655	4.656631						0.21394	0.04594	
77			57 1/2			331,731	4.726615						0.19514	0.04668	
78															
79															
80	1995		2 1/2			9,161,179	0.184911						0.01772	0.07402	
81	Rest of		7 1/2			8,908,592	0.183403						0.01393	0.06868	
82	Mexico		12 1/2			8,163,600	0.201739						0.01542	0.11925	
83			17 1/2			6,892,169	0.221572						0.02431	0.15150	
84			22 1/2			5,891,220	0.228932						0.03091	0.12526	
85			27 1/2			5,063,991	0.229544						0.02743	0.10401	
86			32 1/2			4,468,406	0.217696						0.02217	0.08468	
87			37 1/2			3,553,208	0.218809						0.0172	0.07147	
88			42 1/2			2,896,906	0.212181						0.01324	0.06086	
89			47 1/2			2,356,752	0.216969						0.01131	0.05306	
90			52 1/2			1,828,447	0.214748						0.00987	0.04594	
91			57 1/2			1,567,963	0.211568						0.00988	0.04128	
92															
93															
94															
95															

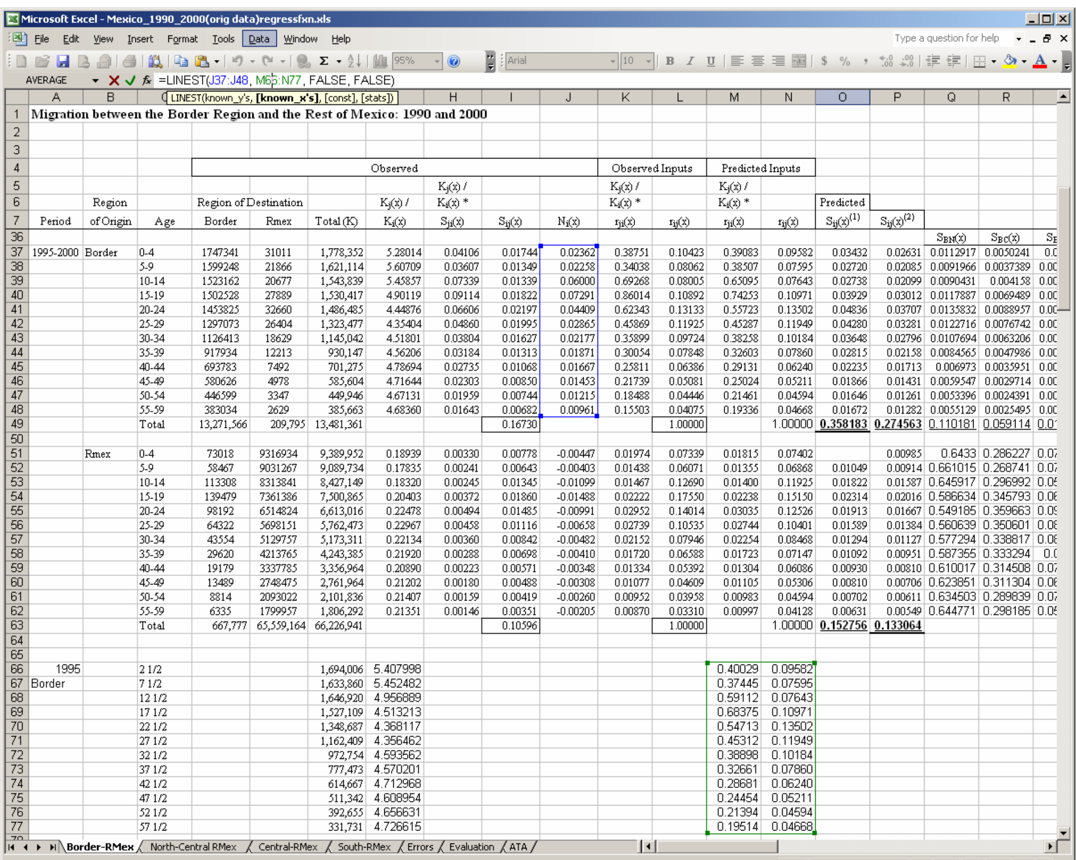
Here Column C represents the median age in each age group (e.g., 2½ for children under five). In column Column F we put our estimated $K_{1995}(x)$, which is equal to $(K_{1990}(x) + K_{2000}(x))/2$. In Figure B2, for instance, we would calculate column F66 by adding together cells

D9 and D23 from Figure B1, adding the corresponding values from the later data set, and then dividing by two. (While we are pretending not to know D9 and D23, we would still know their sum, which is the total Border population at the time of the 1990 census.)

The values for columns G are obtained by dividing *j*'s population by the population of *i*. For example, we calculate cell G66 with the formula = F80/F66. Values for columns M and N are calculated according to the table above, only now we use the population figures from the beginning of the time interval (e.g., the values in rows 66-91).

We now have the data to run our regression. We do so setting the constant to 0, so that the regression line goes through the origin. Figure B3 shows the regression equation in the function bar, along with the regression inputs: the highlighted Column J values are the dependent variable, while the highlighted Column M and N values are the independent variables.

Figure B3. Running the Regression



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Migration between the Border Region and the Rest of Mexico: 1990 and 2000																	
2																		
3																		
4				Observed							Observed Inputs		Predicted Inputs					
5								Kj(x) /			Kj(x) /		Kj(x) /					
6		Region		Region of Destination			Kj(x) /	Ki(x) *			Ki(x) *		Ki(x) *		Predicted			
7	Period	of Origin	Age	Border	Rmex	Total (K)	Ki(x)	Sji(x)	Sij(x)	Nij(x)	rji(x)	rij(x)	rji(x)	rij(x)	$S_{ij}(x)^{(1)}$	$S_{ij}(x)^{(2)}$		
36																	SBN(x)	SBC(x)
37	1995-2000	Border	0-4	1747341	31011	1,778,352	5.28014	0.04106	0.01744	0.02362	0.38751	0.10423	0.39083	0.09582	0.03432	0.02631	0.0112917	0.0050241
38			5-9	1599248	21866	1,621,114	5.60709	0.03607	0.01349	0.02258	0.34038	0.08062	0.38507	0.07595	0.02720	0.02085	0.0091966	0.0037389
39			10-14	1523162	20677	1,543,839	5.45857	0.07339	0.01339	0.06000	0.69268	0.08005	0.65095	0.07643	0.02738	0.02099	0.0090431	0.004158
40			15-19	1502528	27889	1,530,417	4.90119	0.09114	0.01822	0.07291	0.86014	0.10892	0.74253	0.10971	0.03929	0.03012	0.0117887	0.0069489
41			20-24	1453825	32660	1,486,485	4.44876	0.06606	0.02197	0.04409	0.62343	0.13133	0.55723	0.13502	0.04836	0.03707	0.0135832	0.0088957
42			25-29	1297073	26404	1,323,477	4.35404	0.04860	0.01995	0.02865	0.45869	0.11925	0.45287	0.11949	0.04280	0.03281	0.0122716	0.0076742
43			30-34	1126413	18629	1,145,042	4.51801	0.03804	0.01627	0.02177	0.35899	0.09724	0.38258	0.10184	0.03648	0.02796	0.0107694	0.0063206
44			35-39	917934	12213	930,147	4.56206	0.03184	0.01313	0.01871	0.30054	0.07848	0.32603	0.07860	0.02815	0.02158	0.0084565	0.0047986
45			40-44	693783	7492	701,275	4.78694	0.02735	0.01068	0.01667	0.25811	0.06386	0.29131	0.06240	0.02235	0.01713	0.006973	0.0035951
46			45-49	580626	4978	585,604	4.71644	0.02303	0.00850	0.01453	0.21739	0.05081	0.25024	0.05211	0.01866	0.01431	0.0059547	0.0029714
47			50-54	446599	3347	449,946	4.67131	0.01959	0.00744	0.01215	0.18488	0.04446	0.21461	0.04594	0.01646	0.01261	0.0053396	0.0024391
48			55-59	383034	2629	385,663	4.68360	0.01643	0.00682	0.00961	0.15503	0.04075	0.19336	0.04668	0.01672	0.01282	0.0055129	0.0025495
49			Total	13,271,566	209,795	13,481,361			0.16730			1.00000		1.00000	**0.358183**	**0.274563**	0.110181	0.059114
50																		
51		Rmex	0-4	73018	9316934	9,389,952	0.18939	0.00330	0.00778	-0.00447	0.01974	0.07339	0.01815	0.07402		0.00985	0.6433	0.286227
52			5-9	58467	9031267	9,089,734	0.17835	0.00241	0.00643	-0.00403	0.01438	0.06071	0.01355	0.06868	0.01049	0.00914	0.661015	0.268741
53			10-14	113308	8313841	8,427,149	0.18320	0.00245	0.01345	-0.01099	0.01467	0.12690	0.01400	0.11925	0.01822	0.01587	0.645917	0.296992
54			15-19	139479	7361386	7,500,865	0.20403	0.00372	0.01860	-0.01488	0.02222	0.17550	0.02238	0.15150	0.02314	0.02016	0.586634	0.345793
55			20-24	98192	6514824	6,613,016	0.22478	0.00494	0.01485	-0.00991	0.02952	0.14014	0.03035	0.12526	0.01913	0.01667	0.549185	0.359663
56			25-29	64322	5698151	5,762,473	0.22967	0.00458	0.01116	-0.00658	0.02739	0.10535	0.02744	0.10401	0.01589	0.01384	0.560639	0.350601
57			30-34	43554	5129757	5,173,311	0.22134	0.00360	0.00842	-0.00482	0.02152	0.07946	0.02254	0.08468	0.01294	0.01127	0.577294	0.338817
58			35-39	29620	4213765	4,243,385	0.21920	0.00288	0.00698	-0.00410	0.01720	0.06588	0.01723	0.07147	0.01092	0.00951	0.587355	0.333294
59			40-44	19179	3337785	3,356,964	0.20890	0.00223	0.00571	-0.00348	0.01334	0.05392	0.01304	0.06086	0.00930	0.00810	0.610017	0.314508
60			45-49	13489	2748475	2,761,964	0.21202	0.00180	0.00488	-0.00308	0.01077	0.04609	0.01105	0.05306	0.00810	0.00706	0.623851	0.311304
61			50-54	8814	2093022	2,101,836	0.21407	0.00159	0.00419	-0.00260	0.00952	0.03958	0.00983	0.04594	0.00702	0.00611	0.634503	0.289839
62			55-59	6335	1799957	1,806,292	0.21351	0.00146	0.00351	-0.00205	0.00870	0.03310	0.00997	0.04128	0.00631	0.00549	0.644771	0.298185
63			Total	667,777	65,559,164	66,226,941			0.10596			1.00000		1.00000	**0.152756**	**0.133064**		
64																		
65																		
66	1995		2 1/2			1,694,006	5.407998						0.40029	0.09582				
67	Border		7 1/2			1,633,860	5.452482						0.37445	0.07595				
68			12 1/2			1,646,920	4.956889						0.59112	0.07643				
69			17 1/2			1,527,109	4.513213						0.68375	0.10971				
70			22 1/2			1,348,687	4.368117						0.54713	0.13502				
71			27 1/2			1,162,409	4.356462						0.45312	0.11949				
72			32 1/2			972,754	4.593562						0.38898	0.10184				
73			37 1/2			777,473	4.570201						0.32661	0.07860				
74			42 1/2			614,667	4.712968						0.28681	0.06240				
75			47 1/2			511,342	4.608954						0.24454	0.05211				
76			52 1/2			392,655	4.656631						0.21394	0.04594				
77			57 1/2			331,731	4.726615						0.19514	0.04668				

We then run the second regression equation using the reverse-flow data. That is, instead of estimating $S_{ij}^t(+)$ and $S_{ji}^t(+)$ with the equation

$$N_i^t(x) = \frac{K_j^t(x)}{K_i^t(x)}\left[S_{ji}^t(+)\right]r_{ji}^t(x,+) - \left[S_{ij}^t(+)\right]r_{ij}^t(x,+),$$

we do so with the equation

$$N_j^t(x) = \frac{K_i^t(x)}{K_j^t(x)}\left[S_{ij}^t(+)\right]r_{ij}^t(x,+) - \left[S_{ji}^t(+)\right]r_{ji}^t(x,+).$$

In Figure A3, the new dependent variables would be J51:J62, while the new independent variables would be M80:N91 (off the screen in Figure B3). We now create two more columns for the second data set. In our example, these are columns O and P, which hold the $\hat{S}_{ij}(x)$ values predicted by the first and second regressions, respectively.

After Excel runs each regression, it returns two coefficients, m1 and m2. We are primarily interested in estimating $S_{ij}(+)$, which is represented by m2 in the first regression and m1 in the second. We insert these parameters into the Excel spreadsheet (see cells O49 and P49 in Figure A3). We use the absolute values of the coefficients, since flows are positive. To obtain estimates for the age-specific survivorship proportions $S_{ij}(x)$, we multiply $S_{ij}(+)$ (in cells O49 and P49) by the proportional distribution $r_{ij}(x)$. We put the results in columns O and P. Note that we must use an absolute reference to $S_{ij}(+)$ when calculating the $S_{ij}(x)$ values (e.g., O37=O$49*N37).

We now have estimates for $S_{ij}(x)$ for our desired time period. However, we may have chosen to aggregate data because the model performs better in bi-regional rather than multi-regional contexts. If so, we should now disaggregate the data in order to obtain flow estimates that are more geographically precise. In the case of Mexico, for instance, we have been calculating flows between the Border region and the rest of Mexico, but what we are really interested in are the flows between the Border region and each of the remaining three regions of Mexico–North-Central, Central, and South. If we have aggregated the data, then our original migration schedule for the earlier period should tell us what proportion of total Border outmigration was represented by outmigration to each of these three regions. We can then use these ratios to estimate destination-specific flows for the later time interval. For instance, if we want to calculate migration between the Border and the North-Central region, we can do so using the following equation, easily calculated in Excel:

$$S_{B-N}^{t}(x) = \sum_{k} S_{B-k}^{t}(x) \text{ x } \frac{S_{B-N}^{t-10}}{\sum_{k} S_{B-k}^{t-10}},$$

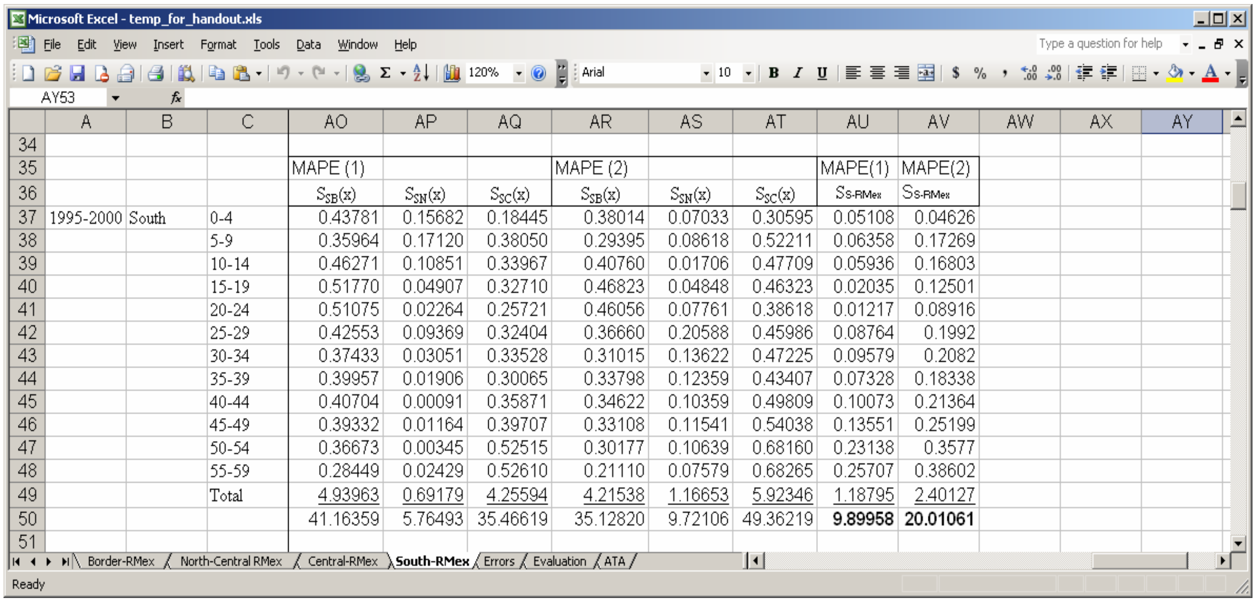
where *t* is the time interval for which we're estimating data, *t-10* is the earlier data that provides our model migration schedule, and $S_{B-N}^{t}(x)$ is the conditional survivorship proportion for migration from the Border to the North for age group *x*.

Finally, we need to assess the model's predictions. If we have full data available for the year being predicted, we can calculate the Mean Absolute Percentage Error (MAPE) as follows:

$$\sum_{ij} \left[\frac{\left| \hat{S}_{ij}^{t}(x) - S_{ij}^{t}(x) \right|}{S_{ij}^{t}(x)} \right].$$

First, we calculate the MAPE for each age-flow value. Then we sum the columns and divide by the number of age groups to get an average MAPE. Finally, we multiply by 100 to get a percentage error. Figure B4 illustrates this set-up for the test case of Mexico's South region. Row 49 gives the MAPE statistic sums for all age groups, while row 50 gives MAPE errors calculated as a percentage. Note that we can calculate separate MAPEs for each destination-specific flow from our fixed origin. In the case below, we can calculate MAPE scores for flows from the South to the Border, North-Central, and Central regions, as well as an overall MAPE for migration from the South to the rest of Mexico.

Figure B4. Calculating MAPE Statistics



	A	B	C	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY
34														
35				MAPE (1)			MAPE (2)			MAPE(1)	MAPE(2)			
36				$S_{SB}(x)$	$S_{SN}(x)$	$S_{SC}(x)$	$S_{SB}(x)$	$S_{SN}(x)$	$S_{SC}(x)$	$S_{S\text{-}RMex}$	$S_{S\text{-}RMex}$			
37	1995-2000	South	0-4	0.43781	0.15682	0.18445	0.38014	0.07033	0.30595	0.05108	0.04626			
38			5-9	0.35964	0.17120	0.38050	0.29395	0.08618	0.52211	0.06358	0.17269			
39			10-14	0.46271	0.10851	0.33967	0.40760	0.01706	0.47709	0.05936	0.16803			
40			15-19	0.51770	0.04907	0.32710	0.46823	0.04848	0.46323	0.02035	0.12501			
41			20-24	0.51075	0.02264	0.25721	0.46056	0.07761	0.38618	0.01217	0.08916			
42			25-29	0.42553	0.09369	0.32404	0.36660	0.20588	0.45986	0.08764	0.1992			
43			30-34	0.37433	0.03051	0.33528	0.31015	0.13622	0.47225	0.09579	0.2082			
44			35-39	0.39957	0.01906	0.30065	0.33798	0.12359	0.43407	0.07328	0.18338			
45			40-44	0.40704	0.00091	0.35871	0.34622	0.10359	0.49809	0.10073	0.21364			
46			45-49	0.39332	0.01164	0.39707	0.33108	0.11541	0.54038	0.13551	0.25199			
47			50-54	0.36673	0.00345	0.52515	0.30177	0.10639	0.68160	0.23138	0.3577			
48			55-59	0.28449	0.02429	0.52610	0.21110	0.07579	0.68265	0.25707	0.38602			
49			Total	4.93963	0.69179	4.25594	4.21538	1.16653	5.92346	1.18795	2.40127			
50				41.16359	5.76493	35.46619	35.12820	9.72106	49.36219	**9.89958**	**20.01061**			
51														

REFERENCES

Bogue, D.J., K. Hinze, and M. White. 1982. *Techniques of Estimating Net Migration*. Chicago: Community and Family Study Center, University of Chicago.

Castro, L.J. 1985. "Analysis of Age-Specific Gross and Net Migration Schedules." Presented at the Population Association of America, Boston, Massachusetts.

INEGI (1986), *X Censo General de Población y Vivienda 1980 Resumen General* (1980 Population Census. General Summary). Mexico City: Instituto Nacional de Estadística, Geografía e Informática.

INEGI (1992), *XI Censo General de Población y Vivienda 1990* (1990 Population Census). Aguascalientes: Instituto Nacional de Estadística, Geografía e Informática (microdata base).

INEGI (2001), *XII Censo General de Población y Vivienda 2000* (2000 Population Census). Aguascalientes: Instituto Nacional de Estadística, Geografía e Informática (microdata base).

Liu, Junwei. 2005. "Estimating Directional Migration Flows Indirectly Using Age-Specific Net Migration Data: A Case Study of Mexico." Master's Thesis, University of Colorado.

Partida, Virgilio. 1993. *Estimación de los niveles de migración en el censo de México de 1980* (Estimation of migration levels in the 1980 census of Mexico). Mexico City: Centro de Estudios en Población y Salud, Secretaría de Salud.

Plane, D.A. 1981. "Estimation of Place-To-Place Migration Flows from Net Migration Totals: A Minimum Information Approach," *International Regional Science Review*, 6 (1), 33-51.

Rogers, A. and Junwei Liu. 2005. "Estimating Directional Migration Flows From Age-Specific Net Migration Data." Forthcoming in *Review of Urban and Regional Development Studies*.

Rogers, A. and Raymer J. 2005. "Origin dependence, secondary migration, and the indirect estimation of migration flows from population stocks." *Journal of Population Research*, 22(1):1-19.