A Method for Indirectly Estimating Historical U.S. Migration Patterns: 1850-1880

Jani Little

Andrei Rogers

Draft

(Not to be quoted without permission from the authors)

May 1, 2012

Abstract

How can standard census variables provide inter-regional migration flows in the absence of a direct migration question? Three different methods are developed and compared. The first is a log-linear model and the second is a regression model. Both employ commonly available census variables are predictors. The preferred method is the third method, which uses a mixture of log-linear modeling, the method of offsets, and basic regression modeling. This method requires fewer assumptions than other methods available at present. Applied to a period in U.S. history that faced great demographic upheaval, the method is validated by the migration data recently released in the North Atlantic Population Project (NAPP) linked dataset, which tracked a representative sample of the U.S. 1880 population and found their places of residence ten years earlier, in 1870. The paper demonstrates that inter-regional migration flows (total and for males and females) can be accurately estimated from the life-time migration data of the children (aged 0-9) during the same decade. Accurate estimates of migrant flow streams depend, however, on having accurate estimates of the total number of migrants out of a region. These can be estimated using the non-migration propensities of the children and the population size of the region at the beginning of decade as predictor variables.

Table of Contents

1. Introduction

When adequate historical population data are not available, demographers generally have to estimate such data indirectly by combining related information (drawn from several sources, time periods, and geographical areas) with models that impose regularities exhibited by past patterns in the particular study area or elsewhere. These efforts generally seek to produce estimates of a variable on the basis of information that may truly be indirectly related to its value.

A recent text on the indirect estimation of gross (directional) migration streams by Rogers, Little, and Raymer (2010) has outlined a set of methods that draws on models and statistical procedures to approximate both the age and the spatial interaction patterns of inter-regional migration. For this paper, we extend some of these procedures described in that text to indirectly estimate historical U.S. migration flows during the 19th century---specifically, the years 1850 to 1880.

The federal census in 1850, for the first time, reported a citizen's place of birth, accompanying it with his or her place of current residence to establish a measure of "lifetime" migration. By focusing on the first age group, 0-9 year olds, one can obtain a crude proxy of childhood migration and associate it, statistically, with the corresponding migration propensities at all age groups (Rogers and Jordan, 2004).

In the next section we describe the mid-19th century historical context of our migration analysis as reported by the decennial censuses during those years and derived from the Integrated Public Use Microdata Surveys (IPUMS) made available by the University of Minnesota's Population Center (MPC). The sizes and age pyramids of regional populations by sex, and the apparent levels and spatial patterns of childhood migration of that period are presented and discussed.

Section 3 of the paper focuses on the observed migration patterns for the 1870-1880 decade, drawing on the special tabulation of the census data provided by the North Atlantic Population Project (NAPP), which identifies changes of residence between 1870 and 1880. In Section 4, a thorough analysis of these migration data sets up statistical relationships in the form of regression models, which are then borrowed to develop indirectly estimated migration patterns for other 19th century decades (i.e., 1850-1860 and 1860-1870). These results are presented and discussed in Section 5.

The final section of the paper discusses the findings and outlines conclusions and possible directions for further research.

2. <u>Population Trends 1850-1880</u>

2.1 The Historical Context: 19^{th} Century America

Migration flows from East to West and from rural to urban areas characterized nineteenth century America. Unprecedented levels of immigration and continual internal migration shaped the regional growth patterns of nineteenth century America. Before the Civil War, 1854 experienced the largest number of new immigrants when 427,833 people arrived; another peak occurred in 1882, when 788,992 arrived; (Schlereth, 1991, p.8). Over a fourth of those who came left, some returning home to their families every year.

A huge and continuing out migration flow of the native-born population took place from the East to the West after the Louisiana Purchase in 1803. This native born outflow from East to West was balanced by the very large inflow of foreign-born persons into the East until the 1890s, when the West began to attract more total immigrants than did the East (Schlereth, 1991, p.13).

The Louisiana Purchase extended the nation’s boundaries across the Mississippi River westward, and the discovery of migration routes through the Rocky Mountains extended the frontier to the West. The last to be settled was the Great Plains Frontier (1865-1890), the large expanse between the Mississippi and the Rocky Mountains (Flanders, 1998, p.77). By the 1890 Census, the United States frontier no longer moved westward, as settled areas appeared throughout the West, and in the previously empty spaces of the country.

Discovery of gold in California’s Sacramento Valley in 1848 attracted some 20,000 male migrants. Ten years later more miners migrated West in search of gold deposits in Pikes Peak, Colorado and the Comstock Lode in Nevada (Smith, ed. 1892, p.58). California’s gold Rush began a process of growth that rapidly grew the state demographically and economically. Then “a patchwork of states followed; Nevada in 1864; Nebraska in 1867; Colorado in 1876; the Dakotas in 1889; and Idaho and Wyoming in 1890,” (Smith, ed., 1992, p.62).

The importance of railroads in the growth and economic development cannot be overemphasized. They linked distant cities to each other, and with the completion of the transnational railroad in 1869, connected both ends of the country, greatly facilitating the settlement of all parts of the nation. Furthermore, rivalries resulted in the birth and death of cities. By 1880, the outliers of a nation system of cities and a national modern railway network were largely in place on the map of America. Manufacturing centers stretching from Boston to Baltimore to Chicago and St. Louis defined the nation's manufacturing heartland core, surrounded by a large periphery. That core transformed a preindustrial into an industrial society, and pushed the percent of the population that lived in urban areas from about 10 percent in 1840 to over 25 percent in 1880.

2.2 Apparent Migration Patterns

2.2.1 Changes in Population Size by Sex

Table 2.1 sets out the regional population totals by sex. The changes over the three decades are startling. At the same time the nation's male population increased almost three fold, the male population of the West increased more than eight times, and the corresponding females population grew by sixteen times. With the exception of the Northeast, all regions increased their share of the national population. The Northeast's share declined by a third. Finally, the West had the highest male/female ratio by a substantial margin, although the sex ratio declined monotonically over the thirty years from a high of 3.02 in 1850 to a low of 1.53 in 1880. The corresponding ratios for the other regions never exceeded 1.08.

2.2.2 Changes in Population Age Compositions by Sex

The age composition (or pyramid) of a population generally reflects the historical patterns of a fertility and mortality that gave rise to it. But the changing age pyramids of the population of the West Region over the decades; 1850-1880, exhibited in Figure 2.1, reveal the striking importance of in-migration in 1850 and 1860, and the gradual transition towards a more "normal" regional age pyramid by 1880.

The age composition of the population in 1850 is that of a typical age composition of in-migrants (i.e., the outmigrants originating in the other three regions). The age compositions reflect the common age profile of outmigrants. Young adults exhibit the highest regional outmigration proportions, while young teenagers show the lowest. Infant migrants move with their parents, therefore, their proportional shares of the total mirror those of their parents, high for young adults and much lower for older adults. But as natural increase grows as a contributor to the West’s population growth, its impact on that region’s population age pyramid transforms it gradually to resemble the age pyramids of the other three regions.

2.2.3 Place of Residence, by Place of Birth, by Sex

Table 2.2 sets out the regional changes in the proportions of U.S.-born males living in regions different from their region of birth. The measures are indicators of ‘lifetime” migration, and as such confound the redistributional impacts of repeat migration (including return migration). They, therefore, are imperfect indicators of geographical mobility. They are, however, course indicators of origin-specific attraction for one destination over others.

Considering the proportions of those who were counted as living in their region of birth, one finds that the Northeast lost the most “local-borns,” at the same time that the West gained the most. The latter region apparently kept almost all of its local-borns, and the Midwest was a close second in this category.

Because the impacts of multiple migrations over the lifetime and the migration preferences of the foreign born are hidden in these measures, their value is reduced to only a cloudy picture of decade-specific migration patterns. Now, in contrast, we turn to the childhood migration patterns, which we deem to be the most valuable measures of inter-regional preferences available in census data.

2.2.4 Childhood Migration Patterns

Demographers use model schedules to capture and describe regularities in the age patterns of demographic rates. In analyzing a large number of age-specific schedules of migration, Rogers and Castro (1981) discovered that the mathematical expression that they called the multi-exponential function provides a remarkably good fit to a wide variety of empirical inter-regional migration schedules. Since then a number of demographers and geographers have adopted it in various studies of migration all over the world.

The observed regularities in age patterns exhibited by countless migration schedules suggest that data on the propensities associated with infant or childhood migration may be linked statistically to the propensities in each of the subsequent age groups and, therefore, also those of all age groups aggregated together.

A single census population distribution that is age, residence, as well as birthplace specific, offers an indication of migration in its very first age group. Children who, for example, are 0-4 years old at the time of the census and living in region j, having been born in region, must have migrated during the birth date to census date (i.e., during the immediately preceding five-year time interval). At their young age it is unlikely that they experienced more than one migration. Such estimates of child migration suggest crude measures of migration levels and of spatial patterns. Regression equations obtained from previous migration data may be used to expand these child migration levels and patterns into the full range of age-specific levels and patterns (Rogers and Jordan, 2004).

Table 2.3 sets out the four-region child migration counts, and corresponding propensities for staying and for migrating, reported by the censuses of 1860, 1870, and 1880. The off-diagonal numbers are translated into the corresponding coefficients of variation (CVs), one for each row of every panel in the table, and Figure 2.2 illustrates their changes over time. This index describes the relative degrees of geographical concentration (or spatial focus) associated with each set of outmigration flows from every origin (Rogers and Sweeney, 1998). The larger the CV coefficient, the greater is the degree of spatial focus associated with the outflows from one region to all of the others. The CV is defined as the standard deviation to mean ratio of a distribution of numbers, ordered from high to low. The standard deviation measures the amount of variation; its division by the mean makes the index a relative one (Rogers and Sweeney, 1999 p. 234).

According to Figure 2.2, the historical pattern of the CV index, from 1850 to 2000, for our particular four-region disaggregation of the U.S. national population, is one of almost linear decline over the 20th century for all the four regions. This decline is evident in both the 0-4 and 0-9 measures of child migration. But our principal interest is the period 1850-1880, and here the patterns are mixed. The most established and most urbanized region, the Northeast exhibits a post-1860 decline. The West, on the other hand, shows sharply increasing CVs from 1850-1900, followed by a modest 20th century decline. After the flat pattern during 1860-1870, the South mirrors the decline exhibited by the Northeast. Finally, from 1880-1900, the Midwest shows a sharply increasing spatial focus of its outflows, but for the other decades it generally mirrors the pattern of the South.

Our findings suggest that the evolution of the spatial focus of each region's outmigration flows generally followed that region's development. The more established regions of the Northeast and the South began to see a general decline of spatial outflow focus around the mid-19th century, whereas the Midwest and West began theirs much later, following the considerable increases of spatial focus exhibited during the second half of the 19th century.

The patterns of regional growth and redistribution presented in our tables and figures evidence those reported in the historical literature cited in Section 2.1. These data now will be used to infer the likely age-specific, direction, migration flows, for which no published data are available. Our indirect estimates of these missing data draw on widely observed regularities that are modeled and imposed, along with a few additional covariates obtained from the decennial censuses.

The multiplicative components given in Table 2.4 show that 287634, 277547, and 290263 are the reported numbers of U.S.-born interregional migrants, aged 0-9, and living in the and U.S. at the time of the censuses in 1860, 1870 and 1880, respectively. The origin components are reported in the "Total" column, and these represent the shares of all migrants from each region. Comparing Panel A with Panels B and C reveals that the Northeast made a declining contribution over the period from 57 percent to 37 percent while the Midwest became more prominent as a contributor of migrants. The Midwest contributed 15% in 1860, 21% in 1870 and 33% in 1880. The destination components are found in the row totals in each panel, and these give the shares of all migrants in each region: in 1870, 75% were in the Midwest and 5% in the West. By 1880, 50% were in the Midwest and 16% in the West. The interaction components represent the ratios of observed migrants to expected migrants, where, for example, the expected flow between the Northeast and the South between 1850 and 1860 would be T*O1*D3=287634*.57*.14=22953.19.

In the 1850-1860 decade, the Midwest sent roughly three times more than expected to the other regions, and the West sent two times more than expected to the Northeast and the South.

For both regions, the pattern settled down by the 1870-1880 decade. On the other hand, the numbers of children, born in the Northeast and the South, that migrated to the West were consistently less than expected in every decade.

Table 2.5 uses the ratios of migration flows across decades as indicators of changes in migration structure from one decade to another. The comparison between the 1850-1860 decade and the 1860-1870 decade is in Panel A. Surprisingly, the most striking results are the flows out of the West. Those to the Northeast were 1.24 times higher in 1860-1870 decade than in the 1850-1860 decade. West flows to the Midwest were 1.59 times higher in the 1860-70 decade than the previous decade. In contrast, flows out of the West to the South were only 1/10th as large in the 1860's as in the 1850's. By the 1870's the flows from the West to the Northeast had decreased and were only two-thirds as large in the 1870's as they had been in the 1860's.

3. Observed Migration, 1870-1880

3.1 The NAPP Linked Data by Sex

As part of the NAPP, a representative sample of the 1880 population (aged 10 years and older) was selected, and then each of these persons was associated with a household 10 years earlier, in the 1870 census. The locations of these two residences, in 1870 and then ten years later in 1880, provided the basis for the migration data. If the residences were located in different regions at the time of the two censuses, the person was counted as a "migrant." If the locations were within the same region in the two censuses, the person was classified as a "non-migrant." Table 3.1 displays the weighted counts based on the NAPP linked data. The non-migrants are reported in the cells on the diagonal and the migrants are reported in the off-diagonal cells.

3.2 The NAPP Migration Propensities by Sex

A migration propensity is the proportion of persons staying, or migrating, from an origin region to a destination region. Migration propensities are also known as conditional survivorship rates (denoted Sij) because they are based only on persons who were counted by the first Census in region I and survived to be counted by the next Census in region j. The propensity for staying in the Northeast, for example, is estimated as the number of persons who were counted, by the 1870 Census, as residing in the Northeast and then again by the 1880 Census (8,594,315 from Table 3.1) divided by the total number of people who were counted as residing in the Northwest in 1870 (9,174,559 from Table 3.1). The migration propensity for intercensal migration from the Northeast to the Northeast is then 8,594,315/9,174,559=.93676. This represents the proportion of Northeast residents, who made no residential migration between Censuses, or who changed residences, but were residing in the Northeast Region at the time of the 1880 Census.

Another example of a migration propensity, calculated for persons originating in the South in 1870 (9,031,488), who subsequently moved to the Midwest, and resided there at the time of the 1880 Census (232,185), is 232,185/9,031,488=.02571. In the same way, the migration (and non-migration) propensities were calculated for all sixteen origin-destination combinations, for the total population and for the male and female populations separately. These are reported in Table 3.2.

What is apparent from this table is that the migration propensities of the males are quite different from the females. In general, females are more likely than males to be non-migrants. In the Northeast, the female propensity to stay was .943 versus .932. This sex difference was present for all regions. However, in the West, the difference was more dramatic, .956 for the females versus .879 for the males.

Of the migrants, the females consistently showed a stronger preference than the males for the Northeast, and the males were more likely than the females to prefer the West. It appears that almost 10 percent (.09) of the males in the West in 1870 migrated to the Midwest by 1880.

3.3 A Comparison of the Child and Non-child Migration Propensities by Sex

The migration propensities have an advantage of comparability over the actual flow counts. When the migration propensities, derived from the NAPP data, for those people 10 years of age or older at the time of the 1880 census (referred to here as the non-child population) are compared to the child migration propensities set out in Table 2.13, it becomes clear that there is a good deal of correspondence between the two sets of propensities.

On the other hand, there are elements of the matrices of migration propensities that differ substantially between the child and non-child patterns. These are highlighted in Table 3.3 with the use of multipliers. When these are lower than 1.0, it indicates the non-child propensity for migration is lower by a factor equal to the multiplier. When they are above one, it indicates the non-child propensity is higher by a factor equal to the multiplier. It is clear from multipliers in the cells representing no migration, i.e. Northeast to Northeast, Midwest to Midwest, etc, that children (aged 0-9) are more likely to be non-migrants than the older population, whether male or female.

Another observation based on Table 3.3 results suggests that the non-child population is much more likely to migrate to the West than the child population. This is especially true in the case of the Northeast where the non-child propensities for migration are higher than the child propensities by factors of 70 and more. It is apparent as well that the child migration propensities correspond more closer to the females propensities than to the males, and this appears to be consistent for all origin-destination combinations.

The correspondence between the child and the non-child migration propensities are summarized in Table 3.3. In the first column, the R^2s are close to 1.0, and, at the same time, the Mean Average Percentage Errors (MAPE) approach 50 percent. Since the R^2 is highly influenced by the high propensities for non-migration relative to low propensities for migration, the R^2 is not appropriate in this situation. On the other hand, in column two where the migration propensities (without the non-migration propensities) are compared (N=12), the R^2 gives an accurate assessment of the degree of correspondence. These suggest a moderately close correspondence, but, based on the MAPEs, large deviations between migration propensities of the children and the non-children. The association between the child and female propensities is stronger than between the child and male propensities. The correspondence between propensities for non-migration (N=4) is quite close based on the MAPEs, especially between the non-child females and the children (MAPE=2.20).

4. Methods for Estimating Migration, 1870-1880

The challenge is to estimate intercensal migration for three decades in history (1850-1860, 1860-1870, and 1870-1880) when there has not been, until recently, any inter-regional migration data available. From the NAPP data, estimated migration flows for the 1870-1880 decade are shown in Table 3.1. Rogers, Little, and Raymer (2010) offer three possible avenues to indirectly estimate migration flow matrices. The first involves using log-linear models with offsets to estimate the unknown intercensal migration matrices, which would be similar in form to those displayed in Table 3.1. The second method is a regression model approach, and it uses identified covariates, derived from standard census questions, to estimate the inter-regional migration propensities. The third approach, which is preferred, uses a mixture of the first two. We will argue that it offers advantages of simplicity and accuracy over the log-linear and the regression modeling approaches.

4.1 The Log-linear Model with Offset Approach

Migration choices are not made randomly or with equal preference for one region or another. The migration spatial structure captures the preferences that a migrant from region i has for region j, controlling for the differences in the numbers of out-migrants and in-migrants. The migration structure is preserved, when the destination preferences are preserved, even though the regional totals may change disproportionately.

A migration structure implies specific interaction effects, which can be reproduced with a saturated log-linear model. For example, the total number of persons migrating between the Northeast and the South, reported in Table 3.1, could be expressed with this multiplicative component form of the log-linear model:

$$n_{13} = (T)(O_1)(D_3)(OD_{13})$$

$$=90{,}481$$

Interpreting the multiplicative components is a straightforward task. T is the total number of persons counted in 1870 and in 1880 (T=28,289,761). The origin component (O_1) represents the share of migrants from region 1 (O_1=9,174,559/28,289,761) and the destination component (D_3) is the share of all migrants to region 3 (D_3=8,907,152/28,289,761). The interaction component (OD_{13}) represents the ratio of observed migrants to expected migrants.

This is equal to $n_{13}/[T*O_1*D_3]$. All of the elements of the migration matrix could be represented in a similar manner, and the total set of multiplicative components would completely represent the migration structure for the total population, 1870-1880. This is the saturated model in log-linear form:

$$\ln(n_{ij}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_{ij}^{OD}$$

The log-linear model with offset approaches apply auxiliary data from a known migration matrix to estimate an unknown migration pattern. The log-linear model with offset is as follows, where n_{ij}^* is the auxiliary data and $\hat{n}_{ij}$ is the estimated data:

$$\ln\left(\hat{n}_{ij}\right) = \lambda + \lambda_i^O + \lambda_j^D + \ln\left(n_{ij}^*\right)$$

Using this model, the estimated spatial migration structure for 1870-1880, for each of the sexes, given by the NAPP linked data and reported in Table 3.1, could be borrowed and applied to the decades with unknown migration patterns: 1860-1870 and 1850-1860. This approach has been used very successfully by Rogers and others when applied in the last half of the 20th century.   It assumes a constant migration structure from one decade to another, and, in this context, the method would have to assume that the 1870-1880 migration structure represents the pattern one and two decades prior, 1860-1870 and 1850-1860, which seems like an unrealistic assumption, given the changes that occurred during that time in history.

The log-linear model with offset is also capable of imposing the migration structure of one population (with a known migration pattern) onto another population (with an unknown migration pattern). This second option for application of the log-linear model with offset approach is based on the lifetime migration patterns of the children, aged 0-9 at the time a census. These data (shown in Table 2.3) offer the best and most direct evidence of the migration behavior of the remaining population, aged 10 and older. The place of birth as well as the place of residence data are available in most censuses, and because, for example, the child migration has occurred within the last year (for those under 1 year of age) or in the last 5 years (for those under 5 years of age), the evidence of the child migrating since birth, presumably, more accurately conveys recent migration behavior than evidence from a different decade.

Estimating the migration structure of the older population from the migration pattern of the children, aged 0-9 at the time of the census, offers improvements over the first log-linear model with offset option that applied the 1870-1880 migration regime to the other decades. First of all, the child migrations occur in the same time frame as the migrations of the older population, so there is no assumption about the constancy of a migration regime over time. Second, the child's region of birth and region of later residence are available in most censuses, which means the auxiliary data exist in historical as well as current contexts.

The general scheme of this method is displayed in Table 4.1. Panel A displays the missing and unknown migration flows, and Panel B displays the auxiliary data that are used in the method. It is important to note that the marginal totals are necessary and must be available to implement the method. If they are unknown, they would need to be estimated.

Table 4.2 presents the estimated flows for all persons based on the log-linear model with offset method, aged 10 and above in 1880, as well as for males and females separately. Notice that all cells in each row and column add up to the marginal totals. And these are the marginal totals given by the observed data (See Table 3.1). The method is evaluated by the R^2 and MAPE statistics, reported in Table 4.3. These results indicate the correspondence between the observed migration flows presented in Table 3.1 and the estimated data in Table 4.2. In the first column, the statistics are reported for non-migrants and migrants combined. In the second column the statistics are reported for just the off-diagonal elements, i.e. the migrants only, and in the third column the statistics are reported for the diagonal elements, i.e. the non-migrants only. It is clear that the method optimizes the fit of the largest elements, the diagonal elements, and this contributes to a high degree of error in the estimates of the migrants. Males and females show quite different migration patterns and, based on the R^2s, the child migration pattern is more closely aligned with the female migration pattern than with the males. The second problem is that, for any log-linear model with offset method, the marginals (row and column totals) in the estimated migration matrix are assumed to be known. If they are not known, they must be estimated, and this adds an unquantifiable amount of uncertainty to the method.

The method of offsets can improve the estimates of the migrants by setting the non-migrant and effectively eliminating the non-migrants from the estimation process. This procedure is demonstrated in Table 4.4 and the resulting estimates are displayed in Table 4.5. The improvement gained by estimating the migrants separately is documented in Table 4.6. These results can be contrasted with the second column of Table 4.3. The R^2s are now close to 1.0 and the MAPEs are substantially reduced for the total population as well as for the males and females.

4.2 The Regression Model Approach

The regression model approach does not have the constraints as the log-linear model with offset approaches. It makes use of the convenient and available child (aged 0-9) migration pattern to predict the migration pattern of the older population. The method was first used by Rogers and Jordan (2004) where they converted the migration flows to propensities (Sijs) and estimated the total propensities for inter-regional migration in the five years prior to 2000 based on the migration propensities of the infants (aged 0-4).

The regression method applied here extends the work of Rogers and Jordan (2004) by identifying additional covariates of migration that are also available in census data. Separate regression models are estimated for the non-migrant propensities, S_{ii} , and for the migrant propensities, S_{ij}.

The dependent variables are the observed conditional survivorships reported in Table 3.2 for all persons (aged 10 and older) and disaggregated by sex. The evidence presented in Table 3.4 demonstrates that the correlations between these propensities and those of the children (aged 0-9) are strong. However, the errors are also substantial. Additional predictor variables are included in the regression approach to reduce this error. These variables were selected based on exploratory models with some guidance from the literature (Rogers and Jordan, 2004), and they are reported in Table 4.7.

Models were used to estimate the propensity for staying in the region (Table 4.8A). The size of the population in the region was positively associated with males staying in the region indicating that a more populated area is more acceptable to males and is not as likely to encourage movement away. The effect is not as strong for females, but, nevertheless, a region's population size significantly and positively affects the female propensity for staying in the region. On the other hand, the non-migrant propensities of the 0-9 year olds is more important in predicting the non-migrant propensities of females than of males.

The migrant models (Table 4.8B) are more complicated because there is more variation in the propensities for moving from i to j. The proportion of males born in i, who are living in j at the end of the period (1880), is a good predictor of total and male migration propensities. For females, the sex ratio (males to females) at the destination j and well as the change in sex ratio in j (1880 minus 1870) are more important for the females.

Two of the predictor variables are measured from the second census, i.e. 1880. The sex ratios were measured at both time periods, 1870 and 1880. Therefore, data from both census years are necessary to predict the intercensal migration patterns.

As expected, the child migration patterns are consistently important in predicting the patterns of the rest of the population. On the other hand, the sex ratio is important for the females, but not in a positive way. Destination regions that are male dominant are less attractive to females and, in addition, destination regions where the gains in the male population exceed the pace of female population are also less attractive to females.

The model fits for the migrants can be visually seen in Figures 4.1 and 4.2. Table 4.9 presents the complete goodness of fit analysis. The R^2 s are substantially higher than those reported for the log-linear models with offset in Table 4.3, and the MAPEs are generally about half the size of the MAPEs associated with the log-linear models.

4.3 The Mixed Method Approach

A comparative evaluation of the models suggests that the method of offsets, which imposes the migration structure of the children on the remaining population, is not an effective method when estimating the non-migrants and migrants simultaneously. It is, however, quite accurate in estimating the migrants separately when the non-migrants are set equal to zero and effectively removed from the migration structure. Compared with the regression method, the method of offsets for migrants only is the more accurate method for predicting the total migrants and the males. However, for the females the two methods are similar (R2=1.00 for the method of offsets and .97 for the regression method) and (MAPE=20.92 for the method of offsets and 19.89 for the regression method). While the regression method is effective at estimating migrants and non-migrants, it is clearly more effective than the method of offsets at estimating the non-migrants.

The mixed method approach is a two-step procedure. Initially, the regression method is used to estimate the numbers of non-migrants, i.e. the numbers of persons who will be found in the same region ten years later. Then the method offsets is used to estimate the migration flows in such a way that the marginal totals (all migrants out of a region plus the non-migrants who remain in the region) are equal to the numbers of persons residing in the region at the beginning of the decade who survived to be counted by the census ten years later. The goodness of fit statistics for the mixed method are summarized in Table 4.10

4.4 Comparing the Methods

The errors associated with each i to j flow, under each method, are presented in Table 4.11. Based on the total numbers in error, the mixed method approach is superior to the other methods if used to estimate the migration structure of the total population or the migration structure of the two sexes separately.

Table 2.1 Population Distributions by Region and Sex: 1850-1880

A. Males

Counts	1850	1860	1870	1880
Northeast	4301769	5223645	5934312	7113279
Midwest	2764115	4647009	6706911	8943675
South	3041206	3701179	6111822	8231054
West	130782	422506	594012	1061265
Total	10237872	13994339	19347057	25349273

Proportions	1850	1860	1870	1880
Northeast	0.42	0.37	0.31	0.28
Midwest	0.27	0.33	0.35	0.35
South	0.30	0.26	0.32	0.32
West	0.01	0.03	0.03	0.04
Total	1.00	1.00	1.00	1.00

B. Females

Counts	1850	1860	1870	1880
Northeast	4263915	5285315	6118285	7298823
Midwest	2548688	4292323	6302608	8302806
South	2890038	3580453	6270763	8191163
West	43253	200157	368764	693201
Total	9745894	13358248	19060420	24485993

Proportions	1850	1860	1870	1880
Northeast	0.44	0.40	0.32	0.30
Midwest	0.26	0.32	0.33	0.34
South	0.30	0.27	0.33	0.33
West	0.00	0.01	0.02	0.03
Total	1.00	1.00	1.00	1.00

C. Sex Ratios

	1850	1860	1870	1880
Northeast	1.01	0.99	0.97	0.97
Midwest	1.08	1.08	1.06	1.08
South	1.05	1.03	0.97	1.00
West	3.02	2.11	1.61	1.53
Total	1.05	1.05	1.02	1.04

Figure 2.1 Changes in Population Age Composition by Sex, West Region

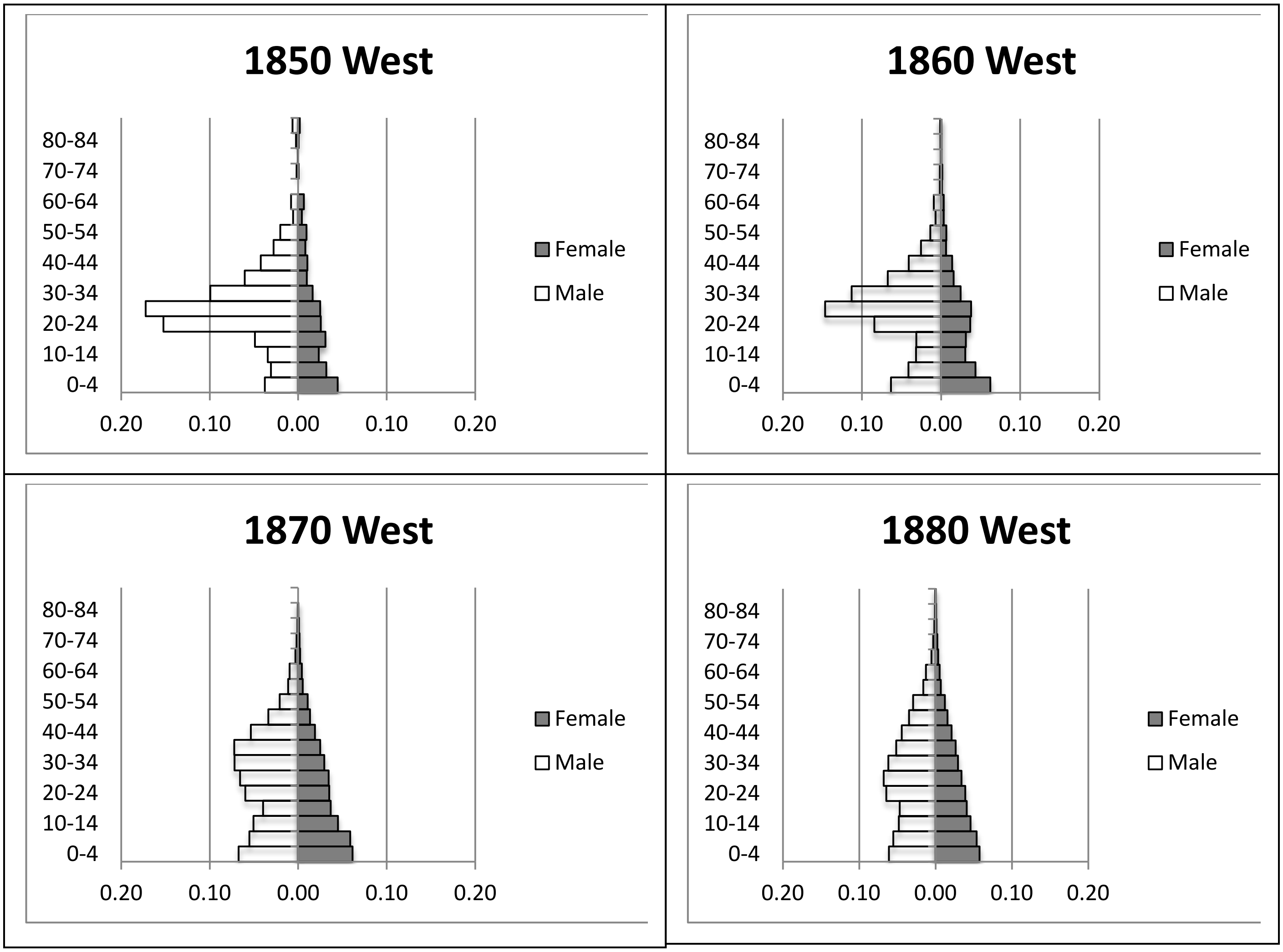

Table 2.2 Changes in Proportions of Males Born in Region (i) and Living in Region (j)

Region of Birth (i)	Region of Residence (j)	Proportion Born in Region i, Living in Region j 1850	1860	1870	1880
Northeast	Northeast	0.85	0.81	0.81	0.83
	Midwest	0.13	0.16	0.16	0.14
	South	0.02	0.02	0.02	0.02
	West	0.01	0.02	0.02	0.02
Midwest	Northeast	0.01	0.01	0.01	0.01
	Midwest	0.96	0.95	0.96	0.94
	South	0.02	0.02	0.02	0.02
	West	0.01	0.02	0.02	0.02
South	Northeast	0.01	0.01	0.01	0.01
	Midwest	0.12	0.11	0.08	0.06
	South	0.86	0.87	0.91	0.92
	West	0.01	0.01	0.01	0.01
West	Northeast	0.03	0.00	0.01	0.01
	Midwest	0.00	0.01	0.02	0.02
	South	0.01	0.00	0.00	0.01
	West	0.96	0.99	0.96	0.97

Table 2.3 Migration Patterns of Children Aged 0-9 at Time of Census

	Panel A. 1860 Region of Residence (Aged 0-9)				
Region of Birth	Northwest	Midwest	South	West	Total
Northeast	1398332	32695	7700	2099	1440826
Midwest	6799	1432730	11600	1700	1452829
South	5200	21196	1146625	500	1173521
West	100	300	200	72588	73188
Total	1410431	1486921	1166125	76887	4140364
Propensities for Migrating	Northwest	Midwest	South	West	Total
Northeast	0.971	0.023	0.005	0.001	1.000
Midwest	0.005	0.986	0.008	0.001	1.000
South	0.004	0.018	0.977	0.000	1.000
West	0.001	0.004	0.003	0.992	1.000
	Panel B. 1870 Region of Residence (Aged 0-9)				
Region of Birth	Northwest	Midwest	South	West	Total
Northeast	1415775	34667	6709	1703	1458854
Midwest	5808	1867098	14131	2005	1889042
South	4307	23138	1917967	1800	1947212
West	604	1808	100	116224	118736
Total	1426494	1926711	1938907	121732	5413844
Propensities for Migrating	Northwest	Midwest	South	West	Total
Northeast	0.970	0.024	0.005	0.001	1.000
Midwest	0.003	0.988	0.007	0.001	1.000
South	0.002	0.012	0.985	0.001	1.000
West	0.005	0.015	0.001	0.979	1.000
	Panel C. 1880 Region of Residence (Aged 0-9)				
Region of Birth	Northwest	Midwest	South	West	Total
Northeast	1066238	27686	5962	101	1099987
Midwest	3234	800474	6871	1717	812296
South	4835	22777	2653096	1663	2682371
West	664	2203	544	191530	194941
Total	1074971	853140	2666473	195011	4789595
Propensities for Migrating	Northwest	Midwest	South	West	Total
Northeast	0.969	0.025	0.005	0.000	1.000
Midwest	0.004	0.985	0.008	0.002	1.000
South	0.002	0.008	0.989	0.001	1.000
West	0.003	0.011	0.003	0.983	1.000

Table 2.4 The Multiplicative Components of Child Migrants

	Panel A. 1860 Region of Residence (Aged 0-9)				
Region of Birth	Northeast	Midwest	South	West	Total
Northeast	0.00	1.19	0.72	0.71	0.57
Midwest	3.59	0.00	3.92	2.94	0.15
South	1.67	1.14	0.00	0.58	0.28
West	2.18	0.63	2.64	0.00	0.00
Total	0.08	0.72	0.14	0.05	287634
	Panel B. 1870 Region of Residence (Aged 0-9)				
Region of Birth	Northeast	Midwest	South	West	Total
Northeast	0.00	1.28	0.71	0.68	0.46
Midwest	2.68	0.00	3.11	2.41	0.21
South	1.21	1.27	0.00	0.57	0.30
West	2.71	1.00	0.27	0.00	0.02
Total	0.11	0.65	0.17	0.07	277547
	Panel C. 1880 Region of Residence (Aged 0-9)				
Region of Birth	Northeast	Midwest	South	West	Total
Northeast	0.00	1.49	0.68	0.63	0.37
Midwest	1.61	0.00	2.19	1.98	0.33
South	1.53	1.54	0.00	0.39	0.27
West	1.80	1.30	0.67	0.00	0.03
Total	0.11	0.50	0.23	0.16	290263

Table 2.5 Ratios of Observed Migration Flows Across Decades

	Panel A. 1870/1860 Region of Residence (Aged 0-9)				
Region of Birth	Northeast	Midwest	South	West	Total
Northeast	0.00	1.08	0.99	0.96	0.81
Midwest	0.75	0.00	0.79	0.82	1.45
South	0.73	1.11	0.00	0.99	1.09
West	1.24	1.59	0.10	0.00	6.15
Total	1.36	0.89	1.25	1.24	0.96
	Panel B. 1880/1870 Region of Residence (Aged 0-9)				
Region of Birth	Northeast	Midwest	South	West	Total
Northeast	0.00	1.16	0.95	0.93	0.80
Midwest	0.60	0.00	0.70	0.82	1.57
South	1.26	1.21	0.00	0.69	0.90
West	0.67	1.30	2.49	0.00	1.13
Total	0.98	0.77	1.34	2.38	1.05

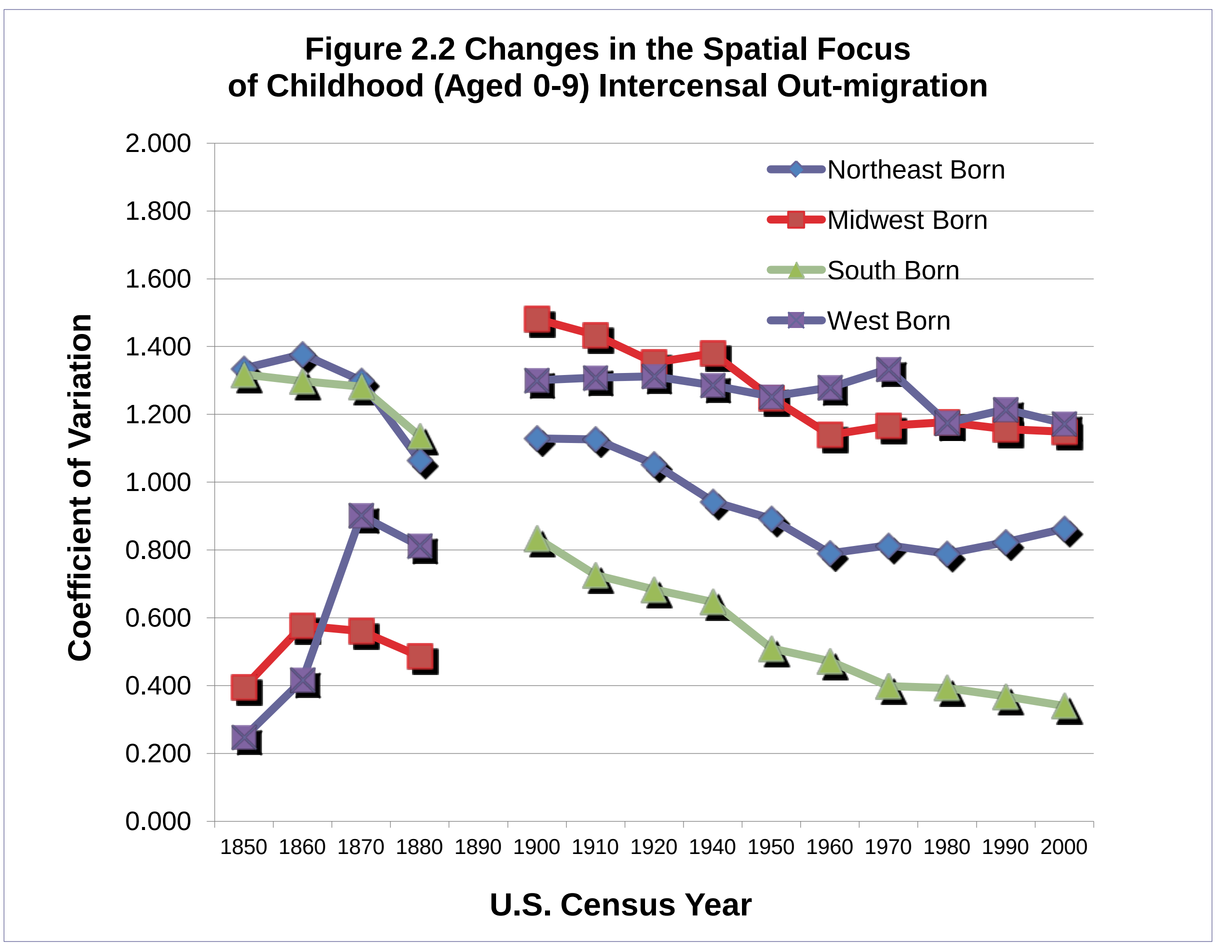

Table 3.1 Migration Flows, 1870-1880, Aged 10+ by Sex

Origin, 1870	Destination, 1880				
A. Total	Northeast	Midwest	South	West	Total
Northeast	8594315	438509	70481	71254	9174559
Midwest	167031	8997716	161110	131088	9456945
South	90933	232185	8672032	36338	9031488
West	14201	40887	3529	568152	626769
Total	8866480	9709297	8907152	806832	28289761
B. Males	Northeast	Midwest	South	West	Total
Northeast	5027250	270171	49039	45925	5392385
Midwest	98482	5381235	112529	96838	5689084
South	51189	151935	5222536	29055	5454715
West	8870	36675	3269	355221	404035
Total	5185791	5840016	5387373	527039	16940219
C. Females	Northeast	Midwest	South	West	Total
Northeast	3567065	168338	21442	25329	3782174
Midwest	68549	3616481	48581	34250	3767861
South	39744	80250	3449496	7283	3576773
West	5331	4212	260	212931	222734
Total	3680689	3869281	3519779	279793	11349542

Table 3.2 Migration Propensities, 1870-1880, Aged 10+ (by Sex) and Aged 0-9

Origin	Destination	Total	Male	Female	Child (Aged 0-9)
Northeast	Northeast	0.93676	0.93229	0.94313	0.96671
	Midwest	0.04780	0.05010	0.04451	0.02469
	South	0.00768	0.00909	0.00567	0.00522
	West	0.00777	0.00852	0.00670	0.00338
Midwest	Northeast	0.01766	0.01731	0.01819	0.00393
	Midwest	0.95144	0.94589	0.95982	0.97812
	South	0.01704	0.01978	0.01289	0.01101
	West	0.01386	0.01702	0.00909	0.00694
South	Northeast	0.01007	0.00938	0.01111	0.00262
	Midwest	0.02571	0.02785	0.02244	0.01172
	South	0.96020	0.95744	0.96442	0.98470
	West	0.00402	0.00533	0.00204	0.00097
West	Northeast	0.02266	0.02195	0.02393	0.00442
	Midwest	0.06523	0.09077	0.01891	0.01417
	South	0.00563	0.00809	0.00117	0.00339
	West	0.90648	0.87918	0.95599	0.97802

Table 3.3 Correspondence between Child and Non-child Migration Propensities

Total		Non-migrants and Migrants (N=16)	Migrants (N=12)	Non-migrants (N=4)
	R^2	0.998557	0.6289404	0.099994
	MAPE	44.958989	58.57495	4.111107
Males				
	R^2	0.9967477	0.5082331	0.039093
	MAPE	49.370217	64.061287	5.297005
Females				
	R^2	0.9997544	0.6847586	0.965596
	MAPE	43.80936	57.678025	2.203364

Table 4.1 The Method of Offsets, Using the Child Migration Data to Estimate the Unknown Migration Data

A. Unknown Data	**1880 Region of Residence (Aged 10+)**				
1870 Region of Residence	Northwest	Midwest	South	West	Total
Northwest	??	??	??	??	9174559
Midwest	??	??	??	??	9456945
South	??	??	??	??	9031488
West	??	??	??	??	626769
Total	8866480	9709297	8907152	806832	28289761
B. Auxiliary Data	**1880 Region of Residence (Aged 0-9)**				
Region of Birth	Northwest	Midwest	South	West	Total
Northwest	1066238	27686	5962	101	1099987
Midwest	3234	800474	6871	1717	812296
South	4835	22777	2653096	1663	2682371
West	664	2203	544	191530	194941
Total	1074971	853140	2666473	195011	4789595

Table 4.2 The Method of Offsets Estimates

A. Total	**1880 Region of Residence (Aged 10+)**				
1870 Region of Residence	Northwest	Midwest	South	West	Total
Northwest	8803419	275206	35648	60285	9174559
Midwest	30424	9257546	63868	105107	9456945
South	31206	170936	8806814	22532	9031488
West	1430	5608	822	618909	626769
Total	8866480	9709297	8907152	806832	28289761
B. Males	**1880 Region of Residence (Aged 10+)**				
Region of Birth	Northwest	Midwest	South	West	Total
Northwest	5151429	174859	23475	42622	5392385
Midwest	16835	5562206	39772	70271	5689084
South	16762	99698	5323632	14623	5454715
West	764	3253	494	399523	404035
Total	5185791	5840016	5387373	527039	16940219
C. Females	**1880 Region of Residence (Aged 10+)**				
Region of Birth	Northwest	Midwest	South	West	Total
Northwest	3651372	100340	12327	18135	3782174
Midwest	13815	3695253	24179	34615	3767861
South	14803	71279	3482939	7752	3576773
West	699	2408	335	219292	222734
Total	3680689	3869281	3519779	279793	11349542

Table 4.3 Method of Offsets, Correspondence between the Observed and the Estimated Non-migrants and Migrants

Total		Non-migrants and Migrants (N=16)	Migrants (N=12)	Non-migrants (N=4)
	R^2	1.00	0.90	0.96
	MAPE	41.43	40.44	3.95
Males				
	R^2	1.00	0.75	0.98
	MAPE	44.28	43.02	5.06
Females				
	R^2	1.00	0.88	1.00
	MAPE	30.61	30.08	2.12

Table 4.4 The Method of Offsets, Using the Child Inter-regional Migration Flows to Estimate the Unknown Flows

A. Unknown Data	**1880 Region of Residence (Aged 10+)**				
1870 Region of Residence	Northwest	Midwest	South	West	Total
Northwest	0	??	??	??	580244
Midwest	??	0	??	??	459229
South	??	??	0	??	359456
West	??	??	??	0	58617
Total	272165	711581	235120	238680	1457546
B. Auxiliary Data	**1880 Region of Residence (Aged 0-9)**				
Region of Birth	Northwest	Midwest	South	West	Total
Northwest	0	79461	16789	10884	107134
Midwest	17440	0	48795	30783	97018
South	13413	60033	0	4948	78394
West	1553	4975	1189	0	7717
Total	32406	144469	66773	46615	290263

Table 4.5 The Method of Offsets, Estimates of Migrants Only

A. Total	**1880 Region of Residence (Aged 10+)**				
1870 Region of Residence	Northwest	Midwest	South	West	Total
Northwest	0	443356	69652	67236	580244
Midwest	149630	0	159637	149961	459229
South	102562	235412	0	21482	359456
West	19973	32813	5831	0	58617
Total	272165	711581	235120	238680	1457546
B. Males	**1880 Region of Residence (Aged 10+)**				
Region of Birth	Northwest	Midwest	South	West	Total
Northwest	0	272978	46150	46007	365135
Midwest	84752	0	113241	109856	307849
South	58896	157328	0	15955	232179
West	14893	28475	5445	0	48814
Total	158541	458781	164837	171818	953977
C. Females	**1880 Region of Residence (Aged 10+)**				
Region of Birth	Northwest	Midwest	South	West	Total
Northwest	0	170647	23454	21008	215109
Midwest	65190	0	46051	40139	151380
South	44411	77151	0	5715	127277
West	4023	5002	778	0	9803
Total	113624	252800	70283	66862	503569

Table 4.6 Method of Offsets, Correspondence between the Observed and the Estimated Migrants

Total		Migrants (N=12)
	R^2	0.99
	MAPE	13.40
Males		
	R^2	0.99
	MAPE	15.98
Females		
	R^2	1.00
	MAPE	20.92

Table 4.7 The Predictor Variables Used in the Regression Models

Origin(i)	**Destination(j)**	**Aged 0-9 S_{ij}, 1880**	**Prop Males Born in i, Living in j, 1880**	**Sex Ratio in j, 1870**	**Change in Sex Ratio in j, 1880-1870**
Northeast	Northeast	0.9667	082792794	0.97	0
	Midwest	0.0247	0.1367413	1.06	0.01
	South	0.0052	0.01513284	0.97	0.03
	West	0.0034	0.02019791	1.61	-0.08
Midwest	Northeast	0.0039	0.0077171	0.97	0
	Midwest	0.9781	0.9430839	1.06	0.01
	South	0.011	0.02431418	0.97	0.03
	West	0.0069	0.02488483	1.61	-0.08
South	Northeast	0.0026	0.01016024	0.97	0
	Midwest	0.0117	0.05987428	1.06	0.01
	South	0.9847	0.92338497	0.97	0.03
	West	0.001	0.00658051	1.61	-0.08
West	Northeast	0.0044	0.00829851	0.97	0
	Midwest	0.0142	0.01760956	1.06	0.01
	South	0.0034	0.00676899	0.97	0.03
	West	0.978	0.96732294	1.61	-0.08

Table 4.8 Regression Model Results, Predicting Non-migrant and Migrant Propensities, 1870-1880

A. Non-migrant Models

Predictors	Total S_{ii} (R^2=.98) coefficient		Male S_{ii} (R^2=.98) coefficient		Female S_{ii} (R^2=1.00) coefficient	
Constant	-.272		-.322		-.229	
S_{ii} Aged 0-9 1880	1.201		1.223		1.212	*
Population (thousands) in j in 1870	3.874E-06	*	5.86E-06	*	1.561E-07	

B. Migrant Models

Predictors	Total S_{ji} (R^2=.91) coefficient		Male S_{ij} (R^2=.88) coefficient		Female S_{ij} (R^2=.97) coefficient	
Constant	-.190		-.339		.083	
S_{ij} Aged 0-9 in 1880	2.438	*	2.980	*	1.506	*
Population (thousands) in j in 1870	8.99E-06	*	1.31E-05	*	1.23E-06	
Proportion Males, Born in i, Living in j in 1870	-.240	*	-.399	*	.049	
Sex Ratio in j in 1870	.097		.197		-.086	*
Sex Ratio Differences in j, 1880-1870	-.639	*	-.642		-.613	*

Table 4.9 Regression Method, Correspondence between the Observed and the Estimated Non-migrant and Migrant Propensities

Total	Non-migrants and Migrants (N=16)	Migrants (N=12)	Non-migrants (N=4)
R^2	1.00	0.91	0.98
MAPE	25.68	25.65	0.11
Males			
R^2	1.00	0.88	0.98
MAPE	28.80	28.74	0.24
Females			
R^2	1.00	0.97	1.00
MAPE	19.92	19.89	0.11

Table 4.10 Mixed Method, Correspondence between the Observed and the Estimated Non-migrants and Migrants

Total	Non-migrants and Migrants (N=16)	Migrants (N=12)	Non-migrants (N=4)
R^2	1.00	0.99	0.98
MAPE	13.42	13.40	0.11
Males			
R^2	1.00	0.98	0.98
MAPE	16.04	15.98	0.24
Females			
R^2	1.00	0.99	1.00
MAPE	20.94	20.92	0.11

Table 4.10 Goodness of Fit Evaluation of the Method of Offsets, the Regression Method, and the Mixed Method

A. Total			**Method of Offsets**		**Regression Method**		**Mixed Method**	
Orig(i)	**Dest(j)**	Observed	Estimated	|Obs-Est|	Estimated	|Obs-Est|	Estimated	|Obs-Est|
Northeast	Northeast	8594315	8803419	209104	8587304	7011	8587304	7011
	Midwest	438509	275206	163303	456566	18057	443356	4847
	South	70481	35648	34833	73778	3297	69652	829
	West	71254	60285	10969	54244	17010	67236	4018
Midwest	Northeast	167031	30424	136607	204311	37280	149630	17401
	Midwest	8997716	9257546	259830	9016234	18518	9016234	18518
	South	161110	63868	97242	314240	153130	159637	1473
	West	131088	105107	25982	207146	76058	149961	18873
South	Northeast	90933	31206	59727	122279	31346	102562	11629
	Midwest	232185	170936	61249	202311	29874	235412	3227
	South	8672032	8806814	134782	8660086	11946	8660086	11946
	West	36338	22532	13806	4144	32194	21482	14856
West	Northeast	14201	1430	12771	14940	739	19973	5772
	Midwest	40887	5608	35279	35684	5203	32813	8074
	South	3529	822	2707	1819	1710	5831	2302
	West	568152	618909	50757	568235	83	568235	83
			Total Error=	**1308945**	**Total Error=**	**443456**	**Total Error=**	**130857**

B. Males			**Method of Offsets**		**Regression Method**		**Mixed Method**	
Orig(i)	**Dest(j)**	Observed	Estimated	|Obs-Est|	Estimated	|Obs-Est|	Estimated	|Obs-Est|
Northeast	Northeast	5027250	5151429	124179	5018240	9010	5018240	9010
	Midwest	270171	174859	95312	278358	8187	272978	2807
	South	49039	23475	25564	50879	1840	46150	2889
	West	45925	42622	3303	33277	12648	46007	82
Midwest	Northeast	98482	16835	81647	125009	26527	84752	13730
	Midwest	5381235	5562206	180971	5405583	24348	5405583	24348
	South	112529	39772	72757	248761	136232	113241	712
	West	96838	70271	26567	159797	62959	109856	13018
South	Northeast	51189	16762	34427	56229	5040	58896	7707
	Midwest	151935	99698	52237	124792	27143	157328	5393
	South	5222536	5323632	101096	5206763	15773	5206763	15773
	West	29055	14623	14432	1233	27822	15955	13100
West	Northeast	8870	764	8106	10118	1248	14893	6023
	Midwest	36675	3253	33422	31059	5616	28475	8200
	South	3269	494	2775	1319	1950	5445	2176
	West	355221	399523	44302	355335	114	355335	114
			Total Error=	**901097**	**Total Error=**	**366458**	**Total Error=**	**125083**

C. Females			Method of Offsets		Regression Method		Mixed Method	
Orig(i)	Dest(j)	Observed	Estimated	|Obs-Est|	Estimated	|Obs-Est|	Estimated	|Obs-Est|
Northeast	Northeast	3567065	3651372	84307	3569964	2899	3569964	2899
	Midwest	168338	100340	67998	178949	10611	170647	2309
	South	21442	12327	9115	22105	663	23454	2012
	West	25329	18135	7194	22613	2716	21008	4321
Midwest	Northeast	68549	13815	54734	78143	9594	65190	3359
	Midwest	3616481	3695253	78772	3609091	7390	3609091	7390
	South	48581	24179	24402	54916	6335	46051	2530
	West	34250	34615	365	42397	8147	40139	5889
South	Northeast	39744	14803	24941	68723	28979	44411	4667
	Midwest	80250	71279	8971	78867	1383	77151	3099
	South	3449496	3482939	33443	3454236	4740	3454236	4740
	West	7283	7752	469	4812	2471	5715	1568
West	Northeast	5331	699	4632	4780	551	4023	1308
	Midwest	4212	2408	1804	4695	483	5002	790
	South	260	335	75	564	304	778	518
	West	212931	219292	6361	212902	29	212902	29
			Total Error=	**407581**	**Total Error=**	**87296**	**Total Error=**	**47430**

Figure 4.1 The Regression Model Fit for the Total Migrants

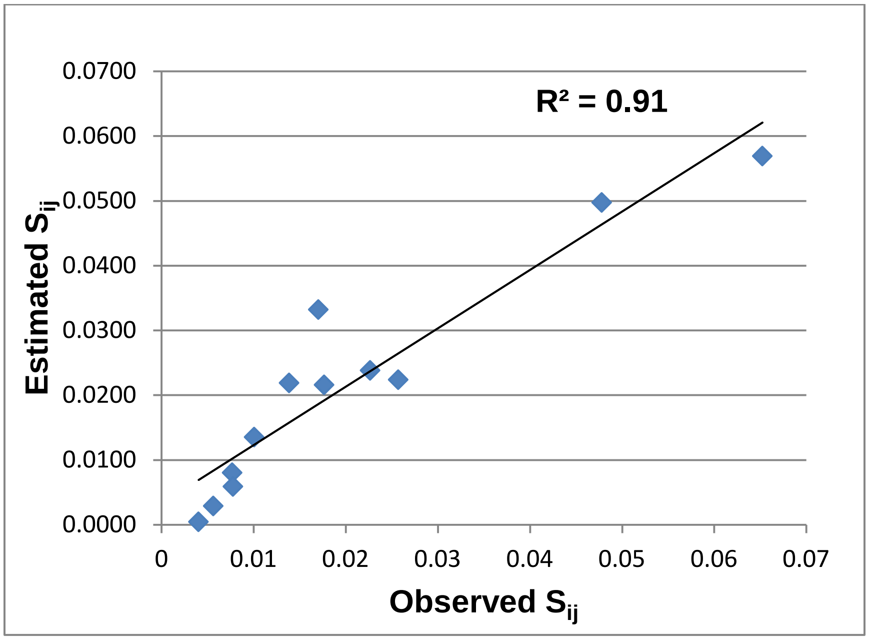

Figure 4.3 The Regression Model Fits for the Male and Female Migrants