

A New Approach in Measuring Local Migration and Population

Zongli Tang

Department of Sociology
Auburn University Montgomery
Montgomery, AL 36124, USA
ztang@mail.aum.edu

Abstract

This paper presents a new model for measuring local migration and population and report results of a promising pilot application to Massachusetts. This model operationalizes Ravenstein's classic "push-pull" paradigm, which posits that local migration is determined by the area's relative attractiveness or a compound function of distinct factors that push migrants out of the area or pull them in. The attraction factors and changes are measured using varied data sources, including decennial census migration flow data and data on group quarters and school enrollments. This model yields timely population estimates with accuracy superior to the corresponding estimates based on the Census Bureau's methodology. Such results warrant further applications to test and refine this promising approach.

Key Words: Social policy, local population estimates, methodology

Résumé

Cet article introduit un nouveau modèle pour mesurer la migration locale de la population et pour présenter les résultats d'un projet pilote prometteur au Massachusetts. Ce modèle opérationnalise le paradigme classique de *push* et *pull* de Ravenstein, qui postule que la migration locale est déterminée par l'attrait relatif d'une région ou par une composante de plusieurs facteurs distincts qui pousse les migrant hors de leur région ou les attire vers une autre. Les changements et les facteurs d'attrait sont mesurés à l'aide de plusieurs sources de données, tel que les données sur la mobilité des migrations du recensement décennal et les données sur les logements de groupe et les inscriptions scolaires. Ce modèle offre des estimations propices qui font preuve d'un niveau d'exactitude supérieur aux estimations correspondantes basées sur la méthodologie du *Census Bureau*. De tels résultats démontrent l'importance d'applications futures pour tester et raffiner cette approche encourageante.

Mots-clés : Politique sociale, estimations de population locale, méthode d'estimation

Introduction

Measuring annual population and population growth following a decennial census is called postcensal population estimation. In efforts to improve measurements of population estimates for states and local areas, a variety of methods have been created, including Component Methods (of I and II) developed by the US Census Bureau (The US Census Bureau, 1947 and 1960), Regression Ratio Correlation Methods (Schmitt & Crosetti 1954; Schmitt & Crosetti 1956), Housing Unit Methods (Starsinic & Zitte 1968), Vital Rates/Censal Ratio Methods (Bogue 1950), Composite Methods (Bogue & Duncan 1959), and Survey Methods (Ericksen 1973; Rives 1982). These methods are well documented by a small number of books and also by the Current Population Reports series published by the Census Bureau (Committee on National Statistics 1980; Lee & Goldsmith 1982; Murdock & Ellis 1991; Rives, et al. 1995; The Census Bureau 1996, Smith & Cody 1999; Bryan 2003). Our research is the latest in a series of efforts to produce accurate estimates.

The most widely adopted means in producing estimates has been the component method. This method estimates current population by adding vital data (births and deaths) and migration counts to the base year population. While vital statistics such as these are generally available at various levels, however finding adequate data and appropriate measurements for estimating migration, especially domestic migration, remains a challenge to demographers.

Currently, the US Census Bureau and many other agencies utilize tax return data (IRS) to estimate domestic migration for the population aged 0-64 and Medicare enrollments for the population aged 65 and older. This is the so-called Administrative Records Method (AR). However, as pointed out by Galdi (1978) and Smith (1999), in using this method, geographic issues including incorrect and outdated mailing addresses, inaccurate reflections on boundary changes, and coding changes, frequently complicate the IRS data, especially at the sub-county level. To overcome these issues, the Bureau replaced the AR method with the Housing Unit Method (HU) in estimating sub-county (or city/town) population beginning with the 1996 round of estimates. In doing so, a methodological inconsistency emerged, for the AR method is basically a component method while the HU method is not.

In using the HU method, two types of measurements are required: group quarters data and households data. The former involves the population living in mental hospitals, nursing homes, jails, college dormitories, and military barracks; the latter consists of population in single-family homes. The need to assemble so many different types of measurements from separate agencies poses practical difficulties, and the collected measurements might not meet necessary standards of accuracy and completeness. Moreover, since there is no easy way to gather household information, in practice, the U. S. Census Bureau collects only house building permit data, treating the data as the measurement of annual changes in household population. However, house building permit data are not good enough to measure the complex population changes, which are a joint effect of many variables, including migration, death, and birth. In fact, no component of population growth is well measured using this method.

Because of such shortcomings with the HU method, the U.S. Census Bureau, following the top-bottom procedure, first produces estimates at state and county levels using the AR method, then estimates at the sub-county level using the HU method, and finally adjusts sub-county estimates by county counts. As mentioned, the AR method depends heavily on the IRS data. The biggest issue with IRS is under-coverage in addition to geographic inaccuracy. For instance, in the IRS method, college students who are claimed by their parents as dependents in tax reports, or who fill out IRS forms for the first time are not classified as migrants even if they have changed their addresses to college towns. Immigrants, minorities and individuals of low socioeconomic status are under-covered as many of them do not report their income to the IRS at all.

The issue of under-coverage can be a serious source of bias in migration estimation for regions such as Massachusetts, where a large number of students move in from other states and countries yearly. In the last two decades, population estimates constructed by the U. S. Census Bureau displayed a

dramatic underestimation for that state. In preparing the 1991-2000 population estimates for Massachusetts, it was essential to explore a new approach in order to reduce underestimation at the state level while measuring population at the other two levels (i.e., county and sub-county) in an efficient and accurate way.

Theoretical Framework and Data

General Framework

This model retains the component method's logic but adds a refinement for estimating migration - domestic migration in particular. The theoretical framework for this approach derives from the classic pull-push paradigm originated by Ravenstein (1889), whereby human migration behavior in an area is governed by local attraction factors. These factors are a compound function of a series of push and pull variables influenced by socioeconomic conditions relative to the outside areas.

Ravenstein's abstract theory is not practically applicable without concrete quantitative measurements of attractions. The focus here is placed on devising such measurements and validating their usefulness. The following formulas describe the process:

$$\begin{aligned} A_t &= \text{pull} - \text{push} \\ &= I_t - O_t \\ \text{or } &= i_t - o_t \\ &= m_t \end{aligned} \tag{1}$$

A_t is the variable of attractions, equaling the difference between pull and push factors. In-migrants (I_t) or in-migration rate (i_t) are used to measure pull factors, and out-migrants (O_t) or out-migration rate (o_t) to measure push factors. The difference between in-migration rate i_t and out-migration rate o_t is the net migration rate m_t , measuring the extent of the relative attraction in an area at a given time. A positive m_t indicates a positive attraction or the domination of pull over push factors. A negative m_t demonstrates a negative attraction or the denomination of push over pull factors. The higher the m_t or the net migration rate, the more attractive the region and vice versa. Pull and push factors can be operationalized in various ways. In the above formula, demographic measurements are employed in operationalization. In other applications, economic indexes and social indicators could be considered possibilities.

Attraction factors change or remain continuous over time depending on distinct socioeconomic determinants in that area. Certain determinants such as geographic location, climate, natural environment, education, and transportation are relatively constant; others, especially economic conditions, typically fluctuate. Therefore, two types of determinants or attraction factors are identified in this model: constant and variable. Formula 1 then evolves into:

$$\begin{aligned} A_t &= A_b + C_t \\ &= m_b + C_t \end{aligned} \quad (2)$$

Here, A_t , the attraction variable in an area, consists of A_b , the constant or basic attraction, and C_t , the variable attraction. A_b is determined by a series of elemental or relatively constant conditions in that community, and in this model it will be operationalized as m_b , the net migration rate at the base time or the base period (i.e., 1985-90 in this estimation) on assumptions that the constant conditions in that community in the new period would be basically the same as those in the last period and that these conditions would be a reflection of migration flow. C_t is determined by the following formula:

$$C_t = C_{1t} + C_{2t} + C_{3t} + \dots + C_{nt} \quad (3)$$

estimated by numerous timely changes in local socioeconomic environments especially economic conditions. It can be seen that the more the changes are recognized, the more precisely the variable attraction C_t is measured. These changes, C_{1t} through C_{nt} , will be operationalized as a set of net migration change rates between the previous and the current periods. This is discussed in the next subsection.

Framework for the Population Aged 0-64

In estimation, Massachusetts population is divided into two age groups: 64 and younger, and 65 and older. The new method is applied primarily to the age group 64 and younger. The constant attraction A_b and the variable attraction C_t are estimated separately for each of the three levels: state, county, and sub-county or Minor Civil Divisions (MCD) (i.e., city/town in Massachusetts).

As discussed previously, A_b was operationalized as the base time net migration rate m_b , and, in the age group 0-64, it is estimated on the basis of data from the long form questionnaires in the 1990 Census, commonly referred to as the "place of residence five years ago" tables and files. This, in the age group 0-64, is estimated on the basis of the data from the long form questionnaires in the 1990 census, commonly referred to as the "place of residence five years ago"

tables and files. A special tabulation, "Special Tape File 28," which shows the total number of persons living in that place by residence status five years ago, is available for 1990 Massachusetts residents at state, county, and sub-county or MCD levels. The Census Bureau prepares such tabulation under contract for any state.¹ For most states, data are available on a county-to-county basis. MCD-to-MCD migration flow data are available only for the New England region. The gross county-to-county migration flow data in the 2000 census has been recently released by the Bureau.²

From the data contained in the tabulation, one is able to calculate the five-year average net migration rate as m_b for the state, counties, and MCDs. Refer to Appendix I for calculation formulas.

From Formula 3, the variable attraction C_t is further operationalized as the following:

$$\begin{aligned} C_t &= C_{1t} + C_{2t} + C_{3t} \\ &= (C_t^s - C_b^s) + (C_t^p - C_b^p) + (C_t^q - C_b^q) \end{aligned} \quad (4)$$

C_{1t} is defined here as the annual net migration change rate for the population aged 6-17, equaling the difference between C_t^s , the annual net migration rate for the population aged 6-17 at time t , and C_b^s , the five-year average net migration rate for the same group at the base time (i.e., 1985-90 for this estimation). These rates are calculated on school enrollments of grades 1 through 12 on the assumption that the enrolled students are representatives of the population aged 6-17 and that children will change schools with their parents. Enrollment data are collected annually by state education agencies as well as the National Center for Education Statistics (NCES). In Massachusetts, the data are gathered by the Massachusetts Department of Education (MDE), available by grade (from kindergarten to grade 12) and school district of student residence (i.e., city/town in Massachusetts). The data files are referred to as the annual "School Attending Children (SAC)" or January reports. In Component Methods I and II, the net migration rate for the population aged 0-64 is estimated entirely on the basis of elementary school enrollments. High school enrollments are ruled out owing to the high dropout rate in this group.

Since 1986, the MDE has collected the annual data of dropout students in high schools as well as the dropouts who return to schools by October 1 of the following year. In the country, all local education agencies were required to report dropouts to the National Center of Education Statistics (NCES) beginning with the 1992-93 school year. In the 2000-01 school year, 45 states reported dropout data to the NCES using the CCD (The Common Core of Data) forms.

The availability of dropout information allows one to fill in the gaps of high school statistics and extend the employment of the SAC data from elementary schools to all schools (i.e., grades 1 through 12), giving this model an advantage over the quality achieved in Component Methods I and II.³

C_{2t} is defined here as the annual net migration change rate for the population aged 25-44, equaling the difference between C_t^p , the annual net migration rate for the population aged 25-44 at time t , and C_b^p , the five-year average net migration rate for the same population at the base time. It is assumed that this age group is the parental counterpart of the student population and that non-family persons have the migration pattern similar to family persons. Net migrants for this group are derived through multiplying net student age migrants by 1.52, a ratio obtained through dividing the national ever-married population aged 25-44 (or the sum of currently married, separated, widowed, and divorced persons in that age group) by the student age population (aged 6 to 17) in the 1990 census. The ratio is calculated on national rather than local basis because of the importance of interstate migration on estimates. In other cases, ratios estimated on basis of local populations could be considered alternatives.

C_{3t} is defined here as the annual net migration change rate for group quarters, equaling C_t^q , the annual net migration rate for group quarters at time t , minus C_b^q , the five-year average net migration rate for the same population at the base time. Group quarter populations are considered to be only those college students in school dormitories proceeding from their importance on Massachusetts estimates, especially on college town estimates. Another consideration involves the issue of overlapping since certain group quarters such as people in mental hospitals, military bases, and nursing houses have been or will be counted in age group population.

Group quarter data are collected annually by state data agencies. Usually, these agencies are the members of the Federal-State Cooperative Program for Local Population Estimates (FSCPE). In Massachusetts, the information is assembled by the state data center, an agency of the University of Massachusetts at Amherst. The collected data are reported to the Bureau and used in its annual estimates. Calculation formulas for C_t are contained in Appendix II.

Framework for the Population Aged 65 plus

The Medicare data, perceived as the best in estimating the elderly population for the quality of coverage, and the method developed by the Bureau (The US Census Bureau, 2002), are adopted in the estimation of the old age migration at state and county levels. The Medicare data are not available for cities/towns, and therefore, the new method will be applied to this level.

At this level, A_b is estimated by the same approach as we did on the population aged 0-64 (see Appendix III for the formula). But, only one measurement of the variable attraction C_t is identified this time. Because of limitations to data, and also considering that the old age population is relatively stable in mobility, it is assumed that C_t in a specific city or town would be affected by annual socioeconomic changes in that county, which could be operationalized as the annual net migration change rate, equaling the difference between the current migration rate and the base time rate in that county. Migrants estimated in this way will be controlled for county counts (see Appendix IV for formulas).

Results

Applying the net domestic migration rate, obtained in the previous section, to the current year expected population (or the sum of survived population plus net immigrants in the current year), generates domestic migration estimates (see Appendix V for the formula). Adding domestic migrants to the expected population yields the complete estimates for that year (see Appendix VI for the formula). Estimates in this case began with the 1990 census. Following the top-bottom procedure, estimates are first estimated for the state, then down to counties, and finally cities/towns by age, sex, and race in Massachusetts.⁴ The annual vital data including births and deaths were gathered by the Department of Public Health in Massachusetts, and used by the Bureau in their estimation as well. The figures of international migration, the same as used by the Bureau, were provided by the INS.⁵

Estimates must be evaluated to see whether any progress has been made. An estimate is considered accurate if it is close to the value of the parameter. The 2000 census counts will be used as the parameter in the evaluation, even though the census itself is subject to various errors. Deviations between estimates and the census counts are assumed to be due to errors in estimates. Meanwhile, a similar comparison will be carried out between the Bureau's estimates and the 2000 census counts to determine if the new approach is superior to the Bureau's. Since the Bureau did not release the 2000 estimates publicly, comparisons will be conducted among the 2000 census counts (April 1, 2000), this study's 2000

and 1999 estimates (April 1, 2000 and July 1, 1999), and the Bureau's 1999 estimates (July 1, 1999).⁶ If the deviation between the 2000 census counts and the 1999 estimates is smaller than that between the census counts and the Bureau's 1999 estimates, an improvement is assumed to be achieved.

The progress made at the state level is quite evident. Table 1 shows that the 2000 estimates (titled as New in this table as well as following tables) are only 3,441 less than the census counts, and that the 1999 estimates are 44,754 less as compared with the Bureau's 173,925 less than the census counts. The percent deviation (or error) of the 1999 estimates is -0.7, two points lower than the Bureau's. Figure 1 describes population trends in Massachusetts as shown by the two decennial censuses and annual estimates of the new method and the Bureau during the decade of 1990-2000. It can be seen that the underestimation has been considerably reduced.

Table 1
Comparisons of Population Models for Massachusetts at the State Level
Massachusetts: 1999 and 2000

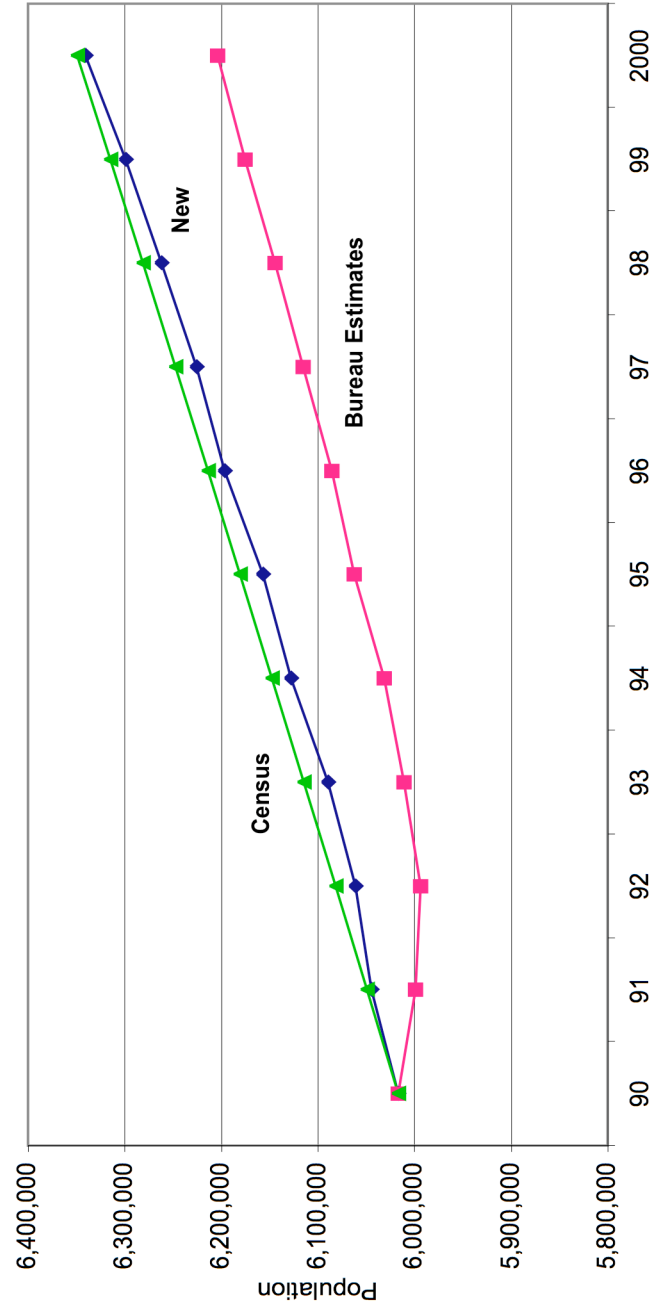
Models	Counts	Deviation	Percent Deviation	MAPE by County %	MAPE by MCD %
2000 Census	6349097				
2000 Estimates (New)	6345656	-3441	-0.06	1.1	5.1
1999 Estimates (Bureau)	6175172	-173925	-2.74	3.6	5.3
1999 Estimates (New)	6304343	-44754	-0.70	1.6	5.2

Table 2. Comparisons of Population Models by County for Massachusetts: 1999 and 2000

County	Census			Estimates			Deviation			Percent Deviation			MAPE by MCD		
	2000	New		2000	Bureau		2000	New		2000	New		2000	New	
		2000	1999		1999	1999		1999	1999		1999	1999		1999	1999
Barnstable	222230	221493	218641	212519	-737	-3589	-9711	-0.003	-0.002	-0.044	0.049	0.053	0.064	0.071	0.083
Berkshire	134953	136140	136277	132218	1187	1324	-2735	0.009	0.01	-0.02	0.049	0.048	0.04	0.134	0.101
Bristol	534678	530363	526416	520258	-4315	-8262	-14420	-0.008	-0.015	-0.027	0.049	0.048	0.04	0.134	0.101
Dukes	14987	14033	13871	14051	-954	-1116	-936	-0.064	-0.074	-0.062	0.049	0.051	0.041	0.065	0.066
Essex	723419	709053	706119	704407	-14366	-17300	-19012	-0.02	-0.024	-0.026	0.049	0.051	0.041	0.065	0.066
Franklin	71535	71675	71629	70806	143	94	-729	0.002	0.001	-0.01	0.065	0.066	0.064	0.044	0.073
Hampden	456228	458305	453078	438279	2077	-3150	-17949	0.005	-0.007	-0.039	0.06	0.063	0.055	0.045	0.036
Hampshire	152251	152463	152041	150892	212	-210	-1359	0.001	-0.001	-0.009	0.06	0.063	0.055	0.045	0.036
Middlesex	1465396	1482685	1474708	1426606	17289	9312	-38790	0.012	0.006	-0.026	0.045	0.047	0.036	0.045	0.036
Nantucket	9520	9573	9281	8206	53	-239	-1314	0.006	-0.025	-0.138	0.006	0.025	0.138	0.036	0.025
Norfolk	650308	658467	653452	643580	8159	3144	-6728	0.013	0.005	-0.01	0.036	0.033	0.025	0.041	0.043
Plymouth	472822	469811	466521	473026	-3011	-6301	204	-0.006	-0.001	0	0.041	0.043	0.03	0.049	0.112
Suffolk	689807	689243	683520	641695	-564	-6287	-48112	-0.001	-0.009	-0.07	0.049	0.046	0.112	0.042	0.058
Worcester	750963	742301	738791	738629	-8662	-12172	-12334	-0.012	-0.016	-0.016	0.042	0.044	0.058	0.042	0.058
State Total	6349097	6345605	6304343	6175172	-3489	-44752	-173925	-0.001	-0.007	-0.027	5.1	5.2	5.3	5.1	5.3

Note: 1999 Estimates (Bureau) and the 2000 population data are provided by the U.S. Census Bureau.

**Figure 1. Population Estimates Comparison
Massachusetts: 1990 - 2000**



Progress is made at the county level too. The mean absolute percent error (MAPE) for counties is 1.6 percent for the 1999 estimates, three points lower than the Bureau's in the same year. Table 2 presents further comparisons at the county level. Of 14 counties in Massachusetts, 12 are improved in terms of the MAPE.

Although the Bureau did not publish its 2000 estimates, a recently released research paper written by Harper, Coleman, and Devine (2003) reveals that the MAPE of the Bureau's 2000 estimates for Massachusetts at the MCD level is 12.4 percent. In contrast, the MAPE of our estimates as shown in Table 1 is only 5.1 percent.

In general, estimates are more accurate for large populations and less accurate for small populations as observed in Table 3. Despite the similar pattern found with both the Bureau's and this study's estimates and these estimates exhibit a lower degree of error except for the population size of 2500 to 9999. For large cities with a population greater than 100,000, the MAPE in the 1999 estimates shows 0.6 percent for the new method, 3.4 points lower than the Bureau's.

Table 3
MAPE by Population Size by City/Town, Massachusetts: 1999 and 2000

Population Size	Counts	Estimates		
		New	Bureau	
		2000	1999	1999
0 - 499	11	12.08	13.03	12.99
500 - 999	19	9.36	9.34	11.54
1,000 - 2,499	45	5.66	6.24	6.33
2,500 - 4,999	44	5.54	6.09	5.11
5,000 - 9,999	71	5.00	5.29	4.52
10,000 - 24,999	95	4.06	4.04	4.41
25,000 - 49,999	45	3.50	3.33	3.50
50,000 - 99,999	17	3.78	3.50	3.61
100,000 +	4	0.76	0.60	4.00

Note: 1999 Estimates (Bureau) and the 2000 population data are provided by the U.S. Census Bureau.

Bias among subgroups, explained as a specific situation in which estimates tend to be too high (upward bias) or too low (downward bias) for certain areas, can be measured as the absolute percent deviation or error (APE). High APE scores indicate a greater amount of bias, and low APE scores suggest a lesser degree of bias. Table 4 illustrates the distribution of the APE scores across regions. The new method once again presents a lower degree of bias than the Bureau.

Discussion

This pilot application in Massachusetts has noticeably improved the estimates especially for the state total and large populations. To a certain degree, this success in estimates could be attributed to progress made on data coverage. The theoretical framework in the model makes it possible.

Table 4
Distribution of Absolute Percent Deviation or Error (APE)
Massachusetts: 1999 and 2000

Range of MAPD %	New		Bureau
	2000	1999	1999
40 - 49	0	0	2
30 - 39	2	1	2
20 - 29	5	6	8
10 - 19	45	43	37
5 - 9	98	101	86
1 - 4	169	172	187
0 - 1	32	28	29

Enrollments are the primary data source of this estimation. Minorities, the poor, and immigrants, who are more likely to settle down in large cities and also more likely to be left out of IRS records, would not be overlooked by the school enrollment system. It accounts for a major contribution to improvements made in large populations. Moreover, school data collectors do not treat students who no longer live within their districts as enrolled there, and they are more capable of tracking address changes than the IRS. This gives enrollment data another advantage over the IRS for the reduced errors relevant to geography.

Nevertheless, one cannot say that the issue of coverage does not exist any more with enrollments. Both Component Methods I and II are based on elementary enrollments. The assumption that school children represent the whole population in migration is essential to the two methods. This assumption however is subject to question because a bias in terms of horizontal representativeness emerges when the inference drawn from enrollments is applied to the whole population.

To improve the representativeness as addressed, it is necessary to seek new migration measurements that are unconnected to enrollments. That is why group quarters was introduced to delegate the college age population who behaved differently from other groups in migration. Meanwhile, the constant attraction variable in the model was operationalized as the base time migration rate, which actually denoted a migration pattern of people who were counted in neither enrollments nor group quarters. For these people, who are generally aged between 50 and 64, would not move so frequently as the younger people, it was hypothesized that their migration behavior was determined by the elemental socioeconomic conditions in that area. As assumed, these conditions hardly change over time. Having adopted these measures, one is able to embrace all groups with dissimilar migration patterns into the estimation. Consequently, the improved data coverage generated a better accuracy in estimation.

Yet, progress made in small populations is not so apparent as we did on large populations. Unlike enrollments, dropouts are recorded by the school of attendance rather than by the district of residence. In calculating net dropouts, one needs to aggregate the data on school districts (or cities/towns) on the assumption that students attend schools located in their districts of residence. In reality, however, a certain degree of inconsistency between residence and attendance is unavoidable. Furthermore, it is not unusual that several small towns share one high school. Under these circumstances, one must split the dropouts among these towns based on population percentages. This would generate another type of inconsistency. Problems like these have very little impact on estimates for the state and counties, but would certainly bias estimates for cities/towns, especially small cities/towns.

While enrollment data are available for both public and private schools in Massachusetts, dropouts are recorded only by public schools. Even if private schools make up a slight proportion in Massachusetts and hold lower drop rate than public schools, the deficiency would surely result in an underestimation of in-migrants, and accordingly undercounts in population estimates of cities/towns where private schools are situated.⁷

Conclusion

Although this model made improvements in a single state during a single decade, it does demonstrate its potential and warrants further test applications elsewhere. All the data employed in this model, including vital statistics, immigrants, enrollments, group quarters, and dropouts, are obtainable in all states. The US Census Bureau has recently released the 2000 long form migration data (county-to-county for most states), which are required for estimating the constant attraction factor in the model.

Starting in 2005, the US Census Bureau plans to implement the American Community Survey (ACS) in all counties. The Bureau expects that once the survey is in full operation, the ACS will be able to provide annual migration information for areas and population groups of 65,000 or more beginning in 2006 summer. The Bureau intends to use the ACS to replace the long form in the decennial census. The ACS, according to the Bureau, will sample about 2.5 percent of the population, while the long form is a survey of about 17 percent of the population. Even with the larger sample size, the long form in the 2000 census was unable to establish a true migration trend for many areas especially small areas.⁸ Therefore, it is doubtful that the ACS, given such a small sample size, is capable of offering migration measurements as accurate as the long form. The striving for methodological innovations in the field is by no means over. In this struggle, our model gives an alternative direction and provides a flexible platform, especially for areas and developing countries where a timely random sample survey is impossible to carry out.

Acknowledgement:

The author thanks Professor Eric Sterling for his useful comments. An earlier version of this article was presented to the annual conference of the Population Association of America, April 2002.

Zongli Tang

End Notes:

1. MCD-to-MCD data are available only for the New England area. For most states, data are available at only two levels including state and county, which are called as county-to-county data.
2. See <http://www.census.gov/population/www/cen2000/ctytoctyflow.html>.
3. Annual net migration of the school age population is first determined by comparing the number of enrollments from kindergarten to grade 11 in one year with the number enrolled in grades 1 to 12 in the following year, then added student deaths and deducted student international migration.
4. Stepwise derivations of migration and population estimates are provided upon request.
5. The estimates for cities/towns in Massachusetts will be provided upon request.
6. The published version of the 2000 estimates was revised on the basis of the 2000 census by the Bureau.
7. Private schools represented about ten percent of the total students in Massachusetts in 2001.
8. This is not discussed in detail in this paper due to the limited space.

References:

- Batutis, Michael J. & Prevost, Ronald C. (1991), "Computers, Database Structure and population Estimates Methodology: Directions for the Coming Decades." Discussion Paper at the annual meeting of the Population Association of America, Washington, D. C.
- Bogue, D.J. (1950), "A Technique for Making Extensive Population Estimates," *Journal of the American Statistical Association* 45: 149-63.
- Bogue, Donald J. & Duncan, Beverly B. (1959), "A Composite Method for Estimating Postcensal Population of Small Areas by Age, Sex, and Color," *Vital Statistics - Special Report* 47.

- Bryan, Tom (2003). Population Estimation, a forthcoming chapter in the revised book ***Methods and Materials of Demography***.
- Bruce D. Spencer & Lee, Che-Fu (1980), *Postcensal Population Estimation Methods of The Census Bureau*, in Panel on Small Area Estimates of Population and Income, ***Estimating Population and Income of Small Areas***. Washington, D.C.: National Academy Press.
- Galdi, D. (1978), ***Memorandum for the Record: IRS Geo-Coding Methodology***. November 29, 1978. Bureau of the Census, US Department of Commerce.
- Crosetti, A. H. & Schmitt, R.C. (1956), "A Method of Estimating the Intercensal Population of Counties." ***Journal of the American Statistics Association*** 51: 587-90.
- Ericksen, Eugene P. (1978), "A Method for Combining Sample Survey Data and Symtomatic Indicators to Obtain Population Estimates for Local Areas," ***Demography*** 10: 137-60.
- Hammer, R.B., P.R.Voss & Blakely R.M. (1999), "Spatially Arranged Growth Forces and Small Area Population Estimates Methodology." Discussion Paper at The Conference on the Methodology of Small Area Estimates, Washington, D.C., 1999.
- Harper, G & C.D.Coleman (2002), "Evaluation of 1990 Based Subcounty Population Estimates and Components." Discussion Paper in the annual meting of PAA, Atlanta.
- Kitigawa, Evelyn M. & Spencer, Bruce D. (1981), "On Estimating Population and Income for Local Areas," in U.S. Department of Commerce, ***Statistical Reporter*** No. 81-8.
- Lee, Everett S. & Goldsmith, Harrold F., eds., (1982), ***Population Estimates: Methods for Small Area Analysis***. Beverly Hills: SAGE Publications.
- Lowe, Teresa, Donald Pittenger & Walker, John.(1977), "Making the Housing Unit Method Work: A Progress Report." Discussion Paper at the annual meeting of the Population Association of America.

Zongli Tang

- McHenry, J.P. (1999), "A Local Alternative to National Demographic Data Vender Population and Household Estimate." Paper presented at the Conference on the Methodology of Small Area Estimates, Washington, D.C.
- Murdock, S.H. & Ellis, D.R. (1991), *Applied Demography*. Boulder: Westview Press.
- Myers, Dowell (1990), Introduction: "The Emerging Concept of Housing Demography," in Dowell Myers (ed.), *Housing Demography*, Madison: University of Wisconsin Press. Pp. 3-31,
- Myers, Dowell. (1987). "Extended Forecasts of Housing Demand in Metropolitan Areas: The Coming Downturn," *Appraisal Journal* 55: 266-78.
- National Research Council. (1980), *Estimating Population and Income for Small Areas*. Washington, DC: National Academy Press.
- Ravenstein, E. (1889). "The Laws of Migration," *Journal of the Royal Statistical Society* 52: 241-301.
- Rives, Norfleet W. Jr. (1982), "Assessment of a Survey Approach," in Everett S. Lee & Goldsmith, Harold F. (eds.), *Population Estimates: Methods for Small Area Analysis*. Beverly Hills: SAGE Publications. Pp.79-96.
- Rives, et, al., (1995), *Basic Methods for Preparing Small-Area Population Estimates*, University of Wisconsin-Madison, Applied Population Laboratory.
- Schmitt, Robert C. (1954), "A Method of Estimating the Population of Cities." *American Journal of Public Health* 44(11): 1426-27.
- Schmitt, R.C. & Crosetti, A.H.. (1954), "Accuracy of the Ratio-Correlation Method for Estimating Postcensal Population," *Land Economics* 30: 279-81.
- Shryock, Henry S. & Siegel, Jacob S. & associates (1971), *The Methods and Materials of Demography*. U.S. Bureau of the Census.

A New Approach in Measuring Local Migration and Population

- Smith, Stanley K. & Cody, Scott (1999), "Evaluating the Housing Unit Method: A Case Study of 1990 Population Estimates in Florida." Discussion Paper at the Population Estimates Methods Conference, Washington, DC.
- Starsinic, Donald E & Zitter, Meyer (1968), "Accuracy of the Housing Unit Method in Preparing Population Estimates for Cities," *Demography* 5: 475-84.
- U.S. Bureau of the Census (2002), "Methodology for Estimates of State and County Total Population". Paper released on the Census Web Page
- U.S. Bureau of the Census (1990), "State and Local Agencies Preparing population and Housing Estimates." *Current Population Reports Series P-25*, No. 1063.
- U.S. Bureau of the Census (1947), "Suggested Procedures for Estimating the Current Population of Counties." *Population Special Reports Series P-47*, No. 4.
- Zitter, M. & Shryock, H.S. (1964), "Accuracy of Method of Preparing Postcensal Population Estimates for Local Areas," *Demography* 1: 227-41.

Appendix:

I. A_b , the constant attraction factor for the population aged 0-64 at state, county, and city/town levels.

$$A_b = m_b = (M_{85-90} / 5) / ({}_sP_{90})$$

M_{85-90} : net domestic migrants aged 0-64 for a given area between 1985 and 1990 (from the special tabulation in the 1990 census).

${}_sP_{90}$: population aged 0-64 in the 1990 census for a given area, domestic migration not included.

II. C_t , the variable attraction factor for the population aged 0-64 at state, county, and city/town levels.

$$C_t^s - C_b^s = (M_t^s) / (P_t) - (M_{85-90}^s / 5) / (P_{90})$$

C_t^s : annual migration rate for the population aged 6-17 at time t

C_b^s : five-year average net migration rate for the population aged 6-17 at the base time, 1985-90

M_t^s : expected net domestic migrants for school students (grades 1 through 12) for a given area at time t, calculated by grade-progression method from SAC data.¹

P_t : expected population aged 0-64 at t time for a given area.

$$P_t = P_{t-1} + B_{t-1,t} - D_{t-1,t} + IM_{t-1,t}$$

M_{85-90}^s : net domestic migrants of population aged 6-17 for a given area between 1985 and 1990, calculated from SAC data, calculated by grade-progression method.

P_{90} : population aged 0-64 in the 1990 census for a given area, domestic migration not included

$$C_t^p - C_b^p = (M_t^p) / (P_t) - (M_{85-90}^p / 5) / (P_{90})$$

C_t^p : annual net migration rate for the population aged 25-44 at time t

C_b^p : five-year average net migration rate for the population aged 25-44 at the base time, 1985-90

¹ Annual net migration of the school age population is first determined by comparing the number of enrollments from kindergarten to grade 11 in one year with the number enrolled in grades 1 to 12 in the following year, then added student deaths and deducted student international migration.

M_t^p : expected net domestic migrants of population aged 25-44 for a given area at time t on the assumption that the population of age 25-44 is the parental cohort of the student population.

$$M_t^p = M_t^s * 1.52$$

M_{85-90}^p : expected net domestic migrants of for population aged 24-44 for a given area between 1985 and 1990 on the assumption that the population of age 25-44 is the parental cohort of the student population.

$$M_{85-90}^p = M_{85-90}^s * 1.52$$

P_t and P_{90} : defined above

$$C_t^q - C_b^q = (M_t^q) / (P_t) - (M_{85-90}^q / 5) / (P_{90})$$

C_t^q : annual net migration for group quarters at time t

C_b^q : five-year average net migration rate for the group quarters at the base time, 1985-1990

M_t^q : expected net migrants of group quarters for a given area at time t, calculated as the difference between the number of current year group quarters and the number of the previous year group quarters.

M_{85-90}^q : net migrants of group quarters for a given area between 1985 and 1990, calculated by the above mentioned method

P_t and P_{90} : defined above.

P_t and P_{90} : defined above

III. **A_b , the constant attraction factor for the population aged 65+ at the city/town level**

$$A_b = m_b = (M_{85-90} / 5) / ({}_sP_{90})$$

M_{85-90} : net domestic migrants aged 65+ for a given MCD between 1985 and 1990 (from the special tabulation in the 1990 census).

${}_sP_{90}$: population aged 65+ in the 1990 census for a given MCD, domestic migration not included.

IV. Ct, the variable attraction factor for the population aged 65+ at the city/town level.

$$C_t = \frac{M_t^c - [\sum (P_t^m * A_b)]}{\sum P_t^m}$$

M_t^c : the net migrants of age 65+ for a given county, calculated on the Medicare data.

P_t^m : the survived old population plus the net immigrants for a given city/town.

A_b : the attraction factor or the base time net migration rate (m_b) of the old population for a given city/town, calculated above.

V. Migration Estimation.

$$DM_t = m_t * (P_0 + B - D + IM)$$

DM_t : net domestic migrants at time t or current year for a given area

m_t : net domestic migration rate at time t or current year for a given area.

P_0 : population at previous time (year) for a given area.

B : births at time t or current year for a given area.

D : deaths at time t or current year for a given area.

IM : net international migrants at time t or current year for a given area.

VI. Population Estimation.

$$P_t = P_0 + B - D + DM + IM$$

P_t : population estimates at time t or current year for a given area

DM : net domestic migrants at time t or current year for a given area

P_0 , B , D , and IM : defined above.